Decision Making under Uncertainty: An Experimental Study in Market Settings

Federico Echenique Taisuke Imai Kota Saito [*]

April 30, 2021

Abstract

We implement nonparametric revealed-preference tests of subjective expected utility theory and its generalizations. We find that a majority of subjects' choices are consistent with the maximization of some utility function. They respond to price changes in the direction subjective expected utility theory predicts, but not to a degree that makes them consistent with the theory. Maxmin expected utility adds no explanatory power. The degree of deviations from the theory is uncorrelated with demographic characteristics. Our findings are essentially the same in laboratory data with a student population and in a panel survey with a general sample of the U.S. population.

JEL Classification codes: D81, D90
Keywords: revealed preferences, subjective expected utility, maxmin expected utility, ambiguity aversion

[*]Echenique: Division of the Humanities and Social Sciences, California Institute of Technology, fede@hss.caltech.edu. Imai: Department of Economics, LMU Munich, taisuke.imai@econ.lmu.de. Saito: Division of the Humanities and Social Sciences, California Institute of Technology, saito@caltech.edu.

Acknowledgement. The authors wish to thank Aurélien Baillon, Yoram Halevy, Yves Le Yaouanq, Olivia Mitchell, 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 SES-1558757 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, Imai acknowledges financial support by the Deutsche Forschungsgemeinschaft through CRC TRR 190, and Saito thanks the NSF for its support through the grant SES-1919263. The project described 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. A summary of our research was published in the TIAA Institute's "Trends and issues," and a first draft of this paper on their "Research dialogue."

1 Introduction

Subjective expected utility theory (SEU; Savage, 1954) is the standard model of decision making in the face of uncertainty, where objective probabilities about uncertain states of the world are not known to agents. The theory postulates an agent that behaves as if they have a subjective probabilistic belief over states of the world and maximizes expected utility with respect to this belief.

While SEU is the leading theory of choice under uncertainty, it is well known to face empirical challenges. In an influential paper, Ellsberg (1961) argued 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 reconciled with a subjective probability. Such avoidance of uncertain bets is termed ambiguity aversion and subsequent empirical literature has identified it in different contexts and in different subject populations (Trautmann and van de Kuilen, 2015).

The empirical literature has relied almost exclusively on the thought experiment discussed in 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). However, the artificial nature of the experiment may question the external validity of its findings. Despite its difficulty, designing choice environments that are more "natural," while providing clean identification, is an important step toward deeper empirical understandings of decision making under uncertainty.

In this study, we present an empirical investigation of SEU and its generalization, maxmin expected utility (MEU; Gilboa and Schmeidler, 1989), from a different angle, combining an experimental paradigm and measurement techniques that are inspired by recent development in revealed preference theory.

We consider a "market" environment in which an agent chooses a portfolio of Arrow-Debreu securities, given state prices and a budget. 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. Similarly, Chambers et al. (2016) provide a condition for MEU when there are two states of the world. Echenique et al. (2018) characterize an "approximate" version of SEU, allowing for errors and mistakes. These revealed-preference characterizations provide tests for SEU and MEU, as well as a measure quantifying "how much" a dataset deviates from these theories. The tests are *nonparametric* in the sense that they do not impose any specific functional forms on utility functions, such as CRRA or CARA. They do assume that agents are risk averse or risk neutral (i.e., they impose a concave von Neumann-Morgenstern utility).

We bring these nonparametric revealed-preference tests to actual choices people make in the face of uncertainty. Following 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 were asked to purchase bundles of state-contingent payoffs under varying budget constraints while not knowing the probabilities of the states of the world.

Our exploratory analysis starts with checking whether subjects are consistent with SEU, MEU, or more general utility maximization; and, if they are not consistent, how large their violations are. In order to investigate the effect of the source of uncertainty on behavior, we generate uncertainty from two different sources. The first source is the classical Ellsberg-style "urns and balls." The second one comes from simulated stock prices. To understand the robustness of our findings across different subject populations, we ran experiments in the laboratory, where we recruited undergraduate students, and on a large-scale internet panel, where we recruited subjects from a general sample of the adult U.S. population. Finally, we compare our measures of degree of deviation from SEU and the standard measure of ambiguity attitude à la Ellsberg.

1.1 Overview of Results

Our main findings are that: (1) subjects are consistent with general utility maximization and Machina and Schmeidler's (1992) probabilistic sophistication, but not SEU; (2) MEU adds no explanatory power to SEU; (3) demand responds to price changes in the direction predicted by SEU, but not enough to make the data fully consistent with SEU; (4) subjects in the laboratory and in the panel display similar patterns; and (5) the correlation between the aforementioned results and demographic characteristics are weak.

The main purpose of our study was to nonparametrically test theories of decision making 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.[1] However, the news is not good for more restrictive theories. In our experiments, the vast majority of subjects, both in the laboratory and on the panel survey, 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, there is only one subject whose choice is consistent with MEU but not SEU.

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

One might conjecture that the theories could be reconciled with the data if one allows for mistakes, but our measures of the distance from the theory do not suggest so. A more forgiving test is to check if prices are negatively correlated with quantities: 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, but not to the extent needed to make them fully consistent with SEU.

Our panel experiment allows us to connect the distance to SEU with subjects' sociodemographic characteristics. We find that the distance to SEU is weakly correlated with financial literacy, with more financially-literate subjects being closer to SEU than less literate subjects. 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 questions. The distance to SEU, or the degree of compliance with downward-sloping demand, are not related to the answers to the Ellsberg questions, but the variability of uncertainty in our market experiment is. Our between-subject experimental design 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 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., Abdellaoui et al., 2011; Baillon and Bleichrodt, 2015; Chew et al., 2017; Epstein and Halevy, 2019; Halevy, 2007).

Other studies try to estimate parameters of the models of decision making under uncertainty (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.[2] 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. Anantanasuwong et al. (2019) apply the methodology of Baillon et al. (2018b) to elicit ambiguity perceptions and attitudes from a sample of Dutch investors.

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).[3] 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., 2019; 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. (2020) 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.

[2]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).

[3]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 laboratory. 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).

2 Revealed Preferences

We introduce our notions of rationality and ways to test them nonparametrically. The discussion in this section serves to motivate our experimental design (Section 3) as well as our strategies for data analysis (Section 4).

Let S be a finite set of *states*. Let $\Delta_{++} = \{\mu \in \mathbf{R}^S_{++} : \sum_{s=1}^S \mu_s = 1\}$ denote the set of strictly positive probability measures on S. In the models we consider below, the objects of choice are state-contingent monetary payoffs, or simply *monetary acts*, which are vectors in $\mathbf{R}^S_+$.

A *dataset* is a finite collection $(p^k, x^k)_{k=1}^K$, where each $p^k \in \mathbf{R}^S_{++}$ is a vector of strictly positive prices and each $x^k \in \mathbf{R}^S_+$ is a monetary act. K indicates the number of observations. The interpretation of a dataset is that each pair (p^k, x^k) consists of a monetary act x^k chosen from the budget $B(p^k, p^k \cdot x^k) = \{x \in \mathbf{R}^S_+ : p^k \cdot x \leq p^k \cdot x^k\}$ of affordable acts. We now introduce several concepts of rationalization of the dataset, ordered from the most restrictive to the least.

Following Echenique and Saito (2015), we say that a dataset $(p^k, x^k)_{k=1}^K$ is *subjective expected utility (SEU) rational with risk aversion* if there exist $\mu \in \Delta_{++}$ and a concave and strictly increasing function $u : \mathbf{R}_+ \to \mathbf{R}$ such that, for all k,

$$y \in B(p^k, p^k \cdot x^k) \Longrightarrow \sum_{s \in S} \mu_s u(y_s) \leq \sum_{s \in S} \mu_s u(x^k_s).$$

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 possible subjective probabilities she entertains. The resulting theory is called maxmin expected utility.

Following Chambers et al. (2016), we say that a dataset $(p^k, x^k)_{k=1}^K$ is *maxmin expected utility (MEU) rational with risk aversion* if there exist a convex set $\Pi \subseteq \Delta_{++}$ and a concave and strictly increasing function $u : \mathbf{R}_+ \to \mathbf{R}$ such that, for all k,

$$y \in B(p^k, p^k \cdot x^k) \Longrightarrow \inf_{\pi \in \Pi} \sum_{s \in S} \pi_s u(y_s) \leq \inf_{\pi \in \Pi} \sum_{s \in S} \pi_s u(x^k_s).$$

Echenique and Saito (2015) and Chambers et al. (2016) develop behavioral axiomatic characterizations of risk-averse SEU and risk-averse MEU with two states, which they term the Strong Axiom of Revealed Subjective Expected Utility and Strong Axiom of Revealed Maxmin Expected Utility, respectively.[4] Discussing these axioms is beyond the scope of the paper, but roughly

[4]Chambers et al. (2016) have results for MEU with more states than two, but only under the assumption of risk neutrality.

speaking, they say that prices and quantities must be inversely related, subject to certain qualifications. We term this *downward-sloping demand* property.

We are able to check whether a given dataset is consistent with SEU or MEU by solving the linear program that is equivalent to the corresponding axiom characterizing each model.

Fact 1. *A dataset $(x^k, p^k)_{k=1}^K$ is SEU rational with risk aversion if and only if there are strictly positive numbers v_s^k, λ^k, and μ_s for $s = 1, \ldots, S$ and $k = 1, \ldots, K$ such that*

$$\mu_s v_s^k = \lambda^k p_s^k, \quad x_s^k > x_{s'}^{k'} \Longrightarrow v_s^k \leq v_{s'}^{k'}.$$

Fact 2. *Given a dataset $(x^k, p^k)_{k=1}^K$, let $K^0 = \{k : x_1^k = x_2^k\}$, $K^1 = \{k : x_1^k < x_2^k\}$ and $K^2 = \{k : x_1^k > x_2^k\}$. A dataset $(x^k, p^k)_{k=1}^K$ is MEU rational with risk aversion and two states if and only if there are strictly positive numbers $\underline{\pi}$, $\bar{\pi}$, π^k, v_s^k, and λ^k for $s = 1, 2$ and $k = 1, \ldots, K$ such that*

$$\pi^k v_s^k = \lambda^k p_s^k, \quad \bar{\pi} \geq \underline{\pi}, \quad x_s^k > x_{s'}^{k'} \Longrightarrow v_s^k \leq v_{s'}^{k'},$$

where $\pi^k = \bar{\pi}$ if $k \in K^1$, $\pi^k = \underline{\pi}$ if $k \in K^2$, and $\pi^k \in [\underline{\pi}, \bar{\pi}]$ if $k \in K^0$.

Facts 1 and 2 stem from the first-order conditions for the maximization of SEU and MEU. Thanks to these facts, testing for SEU or MEU rationality boils down to finding numbers like v_s^k, λ^k, and μ_s. See Online Appendix A for details.

When imposed on a dataset, requiring that a decision-maker maximizes expected utility *exactly*, without errors, may be too demanding. In order to capture situations where the model holds *approximately*, Echenique et al. (2018) relax the previous definition of SEU rationality by "perturbing" some elements of the model.[5]

Let $e \in \mathbf{R}_+$ be a number that controls the size of permissible perturbations. We say that a dataset $(x^k, p^k)_{k=1}^K$ is *e-price-perturbed SEU rational with risk aversion* if there exist $\mu \in \Delta_{++}$, a concave and strictly increasing function $u : \mathbf{R}_+ \to \mathbf{R}$, and $\varepsilon^k \in \mathbf{R}_+^S$ for each $k \in K$ such that, for all k,

$$y \in B(\tilde{p}^k, \tilde{p}^k \cdot x^k) \Longrightarrow \sum_{s \in S} \mu_s u(y_s) \leq \sum_{s \in S} \mu_s u(x_s^k),$$

where for all $k \in K$ and $s \in S$,

$$\tilde{p}_s^k = p_s^k \varepsilon_s^k,$$

[5] Echenique et al. (2018) introduce perturbation of utilities, prices, and beliefs and show that these three sources of perturbations are equivalent. We assume price perturbations here since this source is best suited to our empirical applications.

and for all $k, l \in K$ and $s, t \in S$,

$$\frac{\varepsilon_s^k/\varepsilon_t^k}{\varepsilon_s^l/\varepsilon_t^l} \leq 1 + e.$$

The idea behind the model is that prices are measured, or perceived, with error. We consider the multiplicative form $p_s^k \varepsilon_s^k$ for mathematical convenience. As above, we can check this notion of "approximate" rationality by setting up a linear programming problem.

Fact 3. *Given $e \in \mathbf{R}_+$, a dataset $(x^k, p^k)_{k=1}^K$ is e-price-perturbed SEU rational with risk aversion if and only if there are strictly positive numbers v_s^k, λ^k, μ_s, and ε_s^k for $s = 1, \ldots, S$ and $k = 1, \ldots, K$ such that*

$$\mu_s v_s^k = \lambda^k \varepsilon_s^k p_s^k, \quad x_s^k > x_{s'}^{k'} \Longrightarrow v_s^k \leq v_{s'}^{k'},$$

and for all $k, l \in K$ and $s, t \in S$,

$$\frac{\varepsilon_s^k/\varepsilon_t^k}{\varepsilon_s^l/\varepsilon_t^l} \leq 1 + e.$$

Note that price-perturbed SEU with $e = 0$ corresponds to the exact SEU rationality as discussed above, and any dataset becomes e-price-perturbed SEU rational if we set e large enough. We are thus interested in the *smallest* e for which the dataset becomes e-price-perturbed SEU rational. We term this number *minimal* e and denote it simply by e_*. In the sequel, minimal e will be our notion of distance between the observed dataset and SEU. Using Fact 3, we can compute e_* by setting up a constrained minimization problem as follows.

Fact 4. *Minimal e for SEU is a solution to the following problem:*

$$\min_{(\mu_s, v_s^k, \lambda^k, \varepsilon_s^k)_{k,s}} \max_{k,l \in K, s,t \in S} \frac{\varepsilon_s^k/\varepsilon_t^k}{\varepsilon_s^l/\varepsilon_t^l}$$
$$\text{s.t.} \quad \mu_s v_s^k = \lambda^k \varepsilon_s^k p_s^k, \quad x_s^k > x_{s'}^{k'} \Longrightarrow v_s^k \leq v_{s'}^{k'}.$$

Echenique et al. (2018) study perturbed versions of objective and subjective expected utility. We can extend their framework to define e-price-perturbed MEU and obtain minimal e for MEU in a similar manner. Given a dataset $(x^k, p^k)_{k=1}^K$, let us define $K^0 = \{k : x_1^k = x_2^k\}$, $K^1 = \{k : x_1^k < x_2^k\}$, and $K^2 = \{k : x_1^k > x_2^k\}$ as in Fact 2.

Fact 5. *Given $e \in \mathbf{R}_+$, a dataset $(x^k, p^k)_{k=1}^K$ is e-price-perturbed MEU rational with risk aversion and two states if and only if there are strictly positive numbers $\underline{\pi}$, $\bar{\pi}$, π^k, v_s^k, λ^k, and ε_s^k for $s = 1, 2$ and $k = 1, \ldots, K$ such that*

$$\pi^k v_s^k = \lambda^k \varepsilon_s^k p_s^k, \quad \bar{\pi} \geq \underline{\pi}, \quad x_s^k > x_{s'}^{k'} \Longrightarrow v_s^k \leq v_{s'}^{k'}, \tag{1}$$

where $\pi^k = \bar{\pi}$ *if* $k \in K^1$, $\pi^k = \underline{\pi}$ *if* $k \in K^2$, *and* $\pi^k \in [\underline{\pi}, \bar{\pi}]$ *if* $k \in K^0$, *and for all* $k, l \in K$ *and* $s, t \in S$,

$$\frac{\varepsilon_s^k / \varepsilon_t^k}{\varepsilon_s^l / \varepsilon_t^l} \leq 1 + e.$$

Minimal e for MEU is a solution to the following problem:

$$\min_{(\underline{\pi}, \bar{\pi}, \pi^k, v_s^k, \lambda^k, \varepsilon_s^k)_{k,s}} \max_{k,l \in K, s,t \in S} \frac{\varepsilon_s^k / \varepsilon_t^k}{\varepsilon_s^l / \varepsilon_t^l}$$
$$\text{s.t. constraints (1).}$$

We now turn to the most basic Bayesian model of decision under uncertainty. 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 this probability.[6] An agent is *probabilistically sophisticated* if $x \in \mathbf{R}_+^S$ is evaluated by the distribution it induces given some prior $\mu \in \Delta_{++}$. Epstein (2000) proposes the following necessary condition.

Fact 6. *If a dataset* $(x^k, p^k)_{k=1}^K$ *is probabilistically sophisticated, then there cannot exist* $k, k' \in K$ *and* $s, t \in S$ *such that*

1. $p_t^k \geq p_s^k$ *and* $p_s^{k'} \geq p_t^{k'}$, *with at least one inequality being strict, and*
2. $x_t^k > x_s^k$ *and* $x_s^{k'} > x_t^{k'}$.

Finally, we know, from Afriat (1967) and Varian (1982), that the Generalized Axiom of Revealed Preference (GARP) is a necessary and sufficient condition for a dataset to be consistent with maximization of a well-behaved utility function. We say that a bundle x^k is *directly revealed preferred to* another bundle x, denoted $x^k \succeq^R x$, if $p^k \cdot x^k \geq p^k \cdot x$, and is *strictly directly revealed preferred to* x, denoted $x^k \succ^R x$, if $p^k \cdot x^k > p^k \cdot x$.

Fact 7. *A dataset* $(x^k, p^k)_{k=1}^K$ *satisfies GARP if and only if for any sequence* $((x^{k_1}, p^{k_1}), \ldots, (x^{k_L}, p^{k_L}))$,

$$x^{k_1} \succeq^R x^{k_2}, \quad x^{k_2} \succeq^R x^{k_3}, \quad \ldots, \quad x^{k_{L-1}} \succeq^R x^{k_L} \implies \textit{not } x^{k_L} \succ^R x^{k_1}.$$

When a dataset does not satisfy GARP, we are interested in measuring how severe this violation is. Most of the existing studies applying revealed preference methods use the measure called Critical Cost Efficiency Index, inspired by Afriat's (1967) observation that the violation of GARP disappears if expenditures at each observation are deflated.[7]

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

[7] CCEI is not without problems: see Echenique et al. (2011) for a discussion and a proposed alternative.

Fact 8. *Given a dataset* $(x^k, p^k)_{k=1}^K$ *and a number* $e \in [0,1]$*, define a pair of modified revealed preference relations* $\langle \succeq^{R(e)}, \succ^{R(e)} \rangle$ *by* $x^k \succeq^{R(e)} x$ *if* $ep^k \cdot x^k \geq p^k \cdot x$ *and* $x^k \succ^{R(e)} x$ *if* $ep^k \cdot x^k > p^k \cdot x$*. We say that a dataset* $(x^k, p^k)_{k=1}^K$ *satisfies GARP(*e*) if and only if for any sequence* $((x^{k_1}, p^{k_1}), \ldots, (x^{k_L}, p^{k_L}))$*,*

$$x^{k_1} \succeq^{R(e)} x^{k_2}, \quad x^{k_2} \succeq^{R(e)} x^{k_3}, \quad \ldots, \quad x^{k_{L-1}} \succeq^{R(e)} x^{k_L} \implies \textit{not } x^{k_L} \succ^{R(e)} x^{k_1}.$$

Critical Cost Efficiency Index (CCEI) is the supremum over all the numbers e *such that* $(x^k, p^k)_{k=1}^K$ *satisfies GARP(*e*):*

$$\textit{CCEI} = \sup \left\{ e \in [0,1] : (x^k, p^k)_{k=1}^K \textit{ satisfies GARP}(e) \right\}.$$

3 Experimental Design

The goal of our experiment is to nonparametrically test models of decision making under uncertainty, measure the degree of consistency of the data with the models, and relate this degree to the standard measure of ambiguity attitude, as well as subjects' demographic characteristics. Our design mirrors the environment described in Section 2.

We conducted experiments at the Experimental Social Science Laboratory at the University of California, Irvine (hereafter *the laboratory*), and on the Understanding America Study (UAS) panel, a longitudinal survey platform (hereafter *the panel*).[8] The general structure of tasks in the laboratory and on the panel was the same, but there were several differences between the two. We shall first describe the basic tasks in Section 3.1. Then, in Section 3.2, we turn to the specific features of each implementation— such as recruiting procedures, treatment variations, and incentives. Further details and instructions appear in Online Appendices D and E.

3.1 Tasks

We first describe two 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 laboratory or the panel. Table 1 presents an overview of the laboratory and the panel experiments.

[8]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 (https://uasdata.usc.edu/survey/UAS+116).

Table 1: Structure of the experiment.

	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	—	—

Market task. In the market task, a subject chooses among portfolios of Arrow-Debreu commodities given state prices and a budget. The dataset we intend to collect in this task is of the form $(x^k, p^k)_{k=1}^K$, as introduced in Section 2. Experimental implementations of similar 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 an underlying three-state *state space* $\Omega = \{\omega_1, \omega_2, \omega_3\}$. The probabilities of these states are unknown to the subjects. For each choice problem, there are two relevant *events*, denoted by E_s, $s = 1, 2$. This three-state, two-events, design is part of the methodological innovation in our paper; its purpose will be clear below. 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 (i.e., a partition of Ω). 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 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 the 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.[9]

Our design deviates from the other studies mentioned above by introducing a novel event structure. There are three underlying states of the world $(\omega_1, \omega_2, \omega_3)$ 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

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

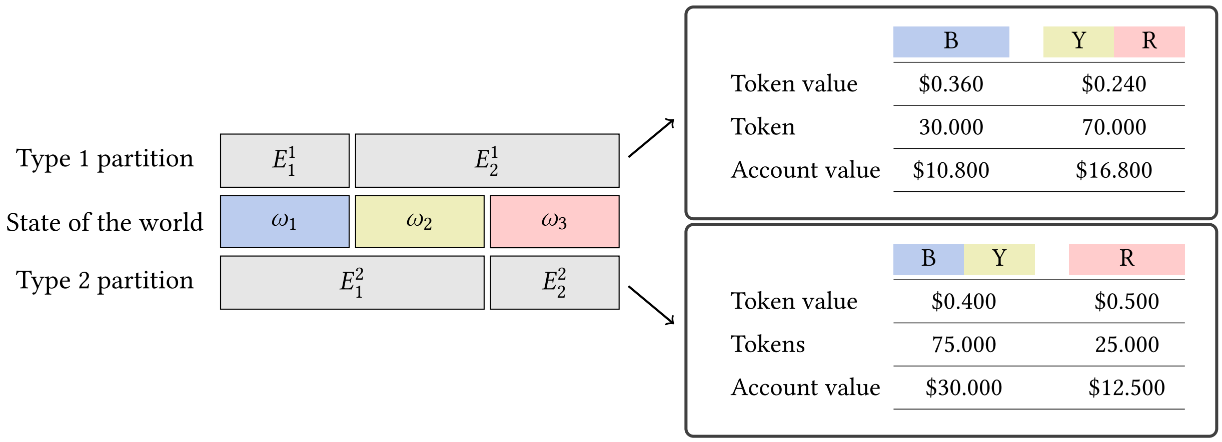

Figure 1: (Left) Event structure in two types of questions. (Right) Illustration of the allocation table for a type 1 question (top) and a type 2 question (bottom).

of questions but also across two types of questions.[10]

The design allows us to examine a very basic aspect of SEU rationality: monotonicity of probability. The monotonicity follows from the fact that SEU-rational agent should consider event $E^2_1 = \{\omega_1, \omega_2\}$ is (weakly) more likely than event $E^1_1 = \{\omega_1\}$ and, hence, the agent should allocate more tokens on event E^2_1 than on event E^1_1 if the prices and income are the same. We term this property *event monotonicity*. In the experiment, we introduced two consecutive questions that have the same budget set, but with different event structures, to test for event monotonicity. Note that these two questions are asked consecutively, meaning that a severe violation of event monotonicity can be attributed to a lack of understanding of the task or inattention, rather than limited memory.

Subjects in the experiment make decisions through a computer interface. The *allocation table* on the screen contains all the information subjects need to make their decisions in each question; see right panels in Figure 1. The allocation 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.[11]

[10] Hey and Pace's (2014) design is the closest to ours. In their experiment, uncertainty was generated by the colors of balls in a 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. We test SEU and its generalization nonparametrically.

[11] 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.

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 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, panel A.[12] 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 $100R\%$ 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, panel A).

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 2, panel B (and Table D.1 in the 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 relatively more expensive than assets in account 2 in type 2 questions. This is reflected in the steeper slopes for the budget lines presented

[12]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 elicited belief 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.

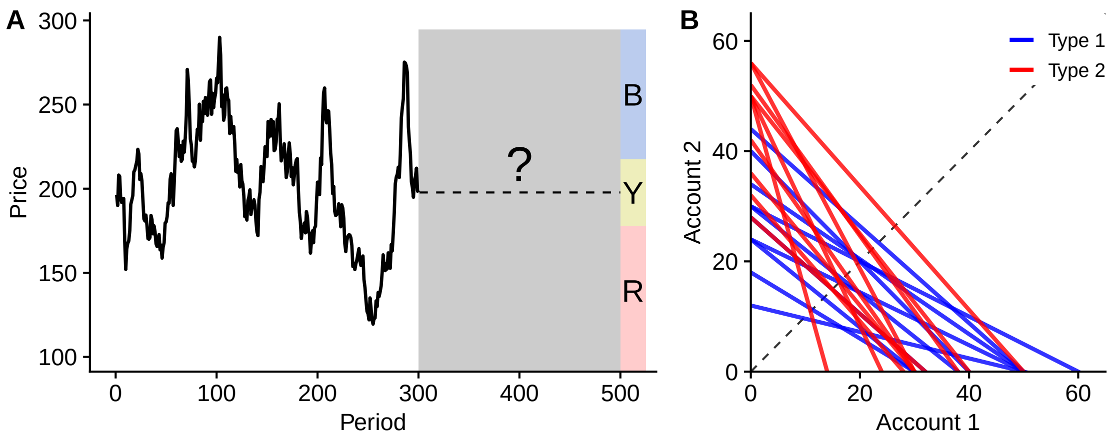

FIGURE 2: (A) Source of uncertainty in the market-stock task. (B) Set of 20 budgets.

in Figure 2, panel B.

Several remarks about our experimental design are in order. First, we allowed subjects to make fractional allocations of tokens (up to the third decimal points) between accounts.[13] Our fractional allocation design sought to mimic choices from a continuous budget line as much as possible, as in the theoretical models we try to test. Second, 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.

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) binary choice question. The purpose of including this standard task is to compare the behavior of subjects in the different designs (allocation vs. choice). Using this comparison, we can investigate how traditional evaluations of ambiguity aversion via binary choices relate to the conclusions drawn from allocation decisions in a market setting.

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

[13]The allocation table (Figure 1) also displayed account values up to the third decimal place, but subjects were informed that the amount below one cent would be rounded up.

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.[14]

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 laboratory study and $(X, b) = (100, 5)$ in the panel. In our laboratory 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 laboratory 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, before exiting the survey module, subjects answered how interesting or uninteresting the survey was and they were also asked to leave any comments if they wished. This is a standard questionnaire that the Understanding America Study (UAS) asks of all its panelist households. The demographic characteristics of the households were already recorded in the previous survey run by the UAS. We could also access datasets from previous surveys that other researchers conducted on the UAS to create additional cognitive and behavioral measures.

3.2 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 laboratory and panel experiments interacted with the exact same interface.

[14] 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.

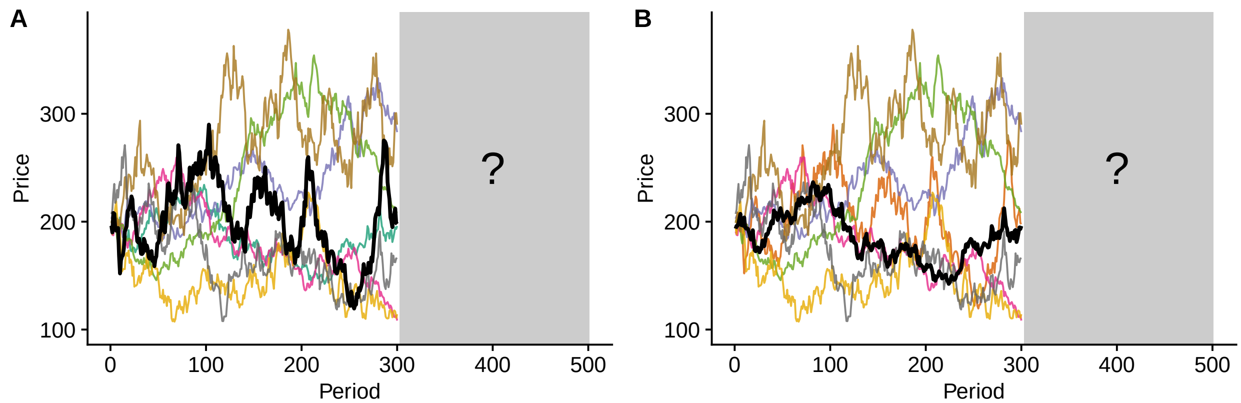

Figure 3: Context of market information. (A) Large volatility (B) Small volatility. *Notes*: One of these two figures is included in the instructions, depending on the treatment.

Recruiting and sampling. Subjects for our laboratory 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.[15] Within the UAS sample, we drew a stratified random sub-sample with the aim of obtaining a balanced 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.[16]

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 3. 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 volatility and Small volatility). The subjects were randomly assigned to either a large volatility condition (Figure 3, panel A), or a small volatility condition (panel B).[17] The instructions for the market-stock task included one of the two charts of Figure 3, depending on

[15] `https://uasdata.usc.edu/index.php`.

[16] The choice of age as the stratification variable is based on the result in Echenique et al. (2018), which shows that the degree of conformity to objective expected utility theory is negatively correlated (younger subjects are closer to the theory than elder subjects) with risk in a similar portfolio choice under known probabilities implemented on several nationally-representative panels.

[17] In the laboratory 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.

the treatment (see Online Appendix E).

Order of the tasks. Subjects in the laboratory 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 the laboratory 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 laboratory 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.

4 Results

We present results from the laboratory and panel experiments separately, but our data analysis follows the same structure. We shall first discuss the basic patterns of subjects' choices, and then proceed to the revealed-preference tests that were discussed in Section 2 above. More precisely, we apply the "exact" revealed-preference tests for general utility maximization (GARP; Fact 7), SEU (Fact 1), MEU (Fact 2), as well as the necessary condition for probabilistic sophistication (Fact 6). After observing that most of the subjects' datasets fail the tests, we quantify the severity of violations by CCEI (Fact 8) and minimal e (Facts 3 and 4).

We also discuss the relationship between the degree of consistency with the models and subjects' demographic characteristics. Finally, we look at the subjects' distance from SEU rationality and their attitude toward ambiguity measured with a simple binary choice task commonly employed in the ambiguity literature.

All statistical tests reported in this section are two-sided unless otherwise noted.

4.1 Results from the Laboratory

We conducted seven sessions at the Experimental Social Science Laboratory of the University of California, Irvine. A total of 127 subjects (62 in the small volatility treatment and 65 in the large volatility treatment; mean age = 20.16, SD = 1.58; 35% male) participated in the study.[18] Each session lasted about an hour, and subjects earned on average \$21.3 (including a \$7 showup fee; SD = 9.21).

Allocation decisions in the market tasks. Subjects faced budgets in random order, with one exception, which is related to event monotonicity discussed in Section 3.1. 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. The purpose of having these questions in fixed order 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 3). 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 may reflect subjects' ambiguity aversion. See Figure B.2 in the Online Appendix.

The empirical content of expected utility is captured in part by a negative relation between state prices and allocations as Echenique et al. (2018) discuss in depth: 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 log price ratios,

[18] 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" two extremely slow subjects to make decisions more quickly, and hence eliminated their choices from our data.

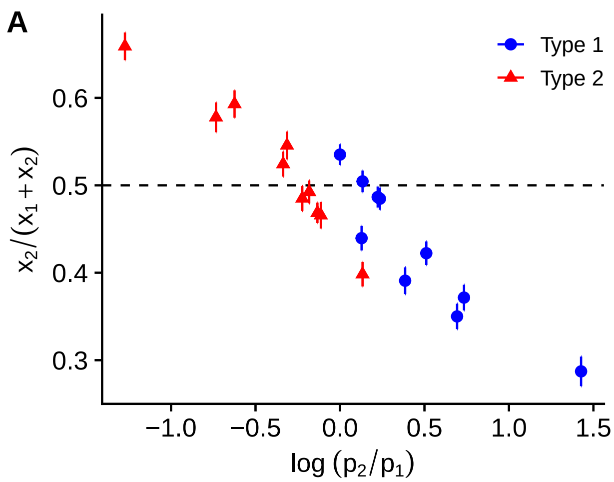

FIGURE 4: Downward-sloping demand at the aggregate level. (A) Market-stock task. (B) Market-Ellsberg task. *Notes*: Each point represents mean $x_2/(x_1+x_2)$ at each $\log(p_2/p_1)$ and bars indicate standard error of means.

$\log(p_2/p_1)$, and allocation shares, $x_2/(x_1+x_2)$, pooling choices from all subjects. Figure 4 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 (type 1 and type 2 event partitions) and in both tasks (market-stock and market-Ellsberg).

We also quantify the degree of compliance with the downward-sloping demand property at the individual level by calculating the correlation ρ^{dsd} between $\log(p_2/p_1)$ and $x_2/(x_1+x_2)$.[19] A significant majority of the subjects (92.1% in the market-stock task and 88.2% in the market-Ellsberg task) made choices that responded to prices negatively ($\rho^{\text{dsd}} < 0$; Figure B.3 in the Online Appendix).

Individual-level data exhibit heterogeneous choice patterns. Figure 5 presents the relationship between $\log(p_2/p_1)$ and $x_2/(x_1+x_2)$ for five selected subjects. As in prior studies (e.g., Ahn et al., 2014; Choi et al., 2007), there are subjects who responded to price changes smoothly (panels A and B), partially or fully "hedged" uncertainty by choosing bundles on or close to the 45-degree line (panels C and D), and chose bundles all over the space (panel E).

We proceed to ask the question: 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?

[19]We first calculate (Spearman's) correlation coefficient ρ_t for each type ($t = 1, 2$) of questions. To obtain the "average" correlation coefficient ρ^{dsd}, we first convert correlation coefficients to z-values by Fisher's transformation, take the average, and convert it back to a correlation coefficient. This procedure is summarized as $\rho^{\text{dsd}} = \tanh\left(\sum_{t=1}^{2} \tanh^{-1}(\rho_t)/2\right)$.

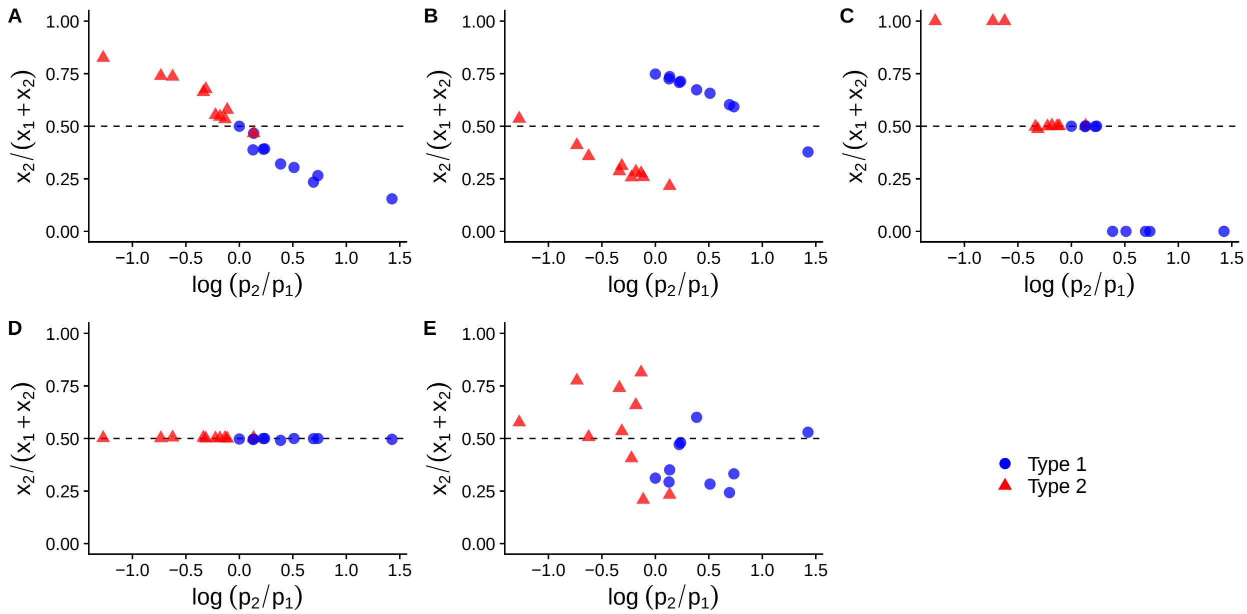

FIGURE 5: The relationship between the log-price ratio $\log(p_2/p_1)$ and the allocation share $x_2/(x_1 + x_2)$ for selected subjects.

Revealed-preference tests. We implement nonparametric, revealed-preference tests on each individual subject's choice data. These tests include: GARP, probabilistic sophistication (hereafter PS; Machina and Schmeidler, 1992), SEU (based on and extended from Echenique and Saito, 2015), and MEU (based on and extended from Chambers et al., 2016). As discussed in Section 2, we can test whether a given dataset is consistent with SEU or MEU by solving the linear program implied by the axiom that characterizes the model. We say that a dataset *passes the test* if there is a solution to the problem.

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 also satisfies event monotonicity. For SEU and MEU, we implement the test directly on the data combining the two types of problems. It is, at first glance, 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 of the methodological contribution of our paper: see Online Appendix A for details.

Table 2 presents the *pass rate* of each test. That is, the fraction of subjects (out of 127) who passed each test. We find that a majority of subjects satisfy GARP, meaning that their choices are

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

Notes: $N = 127$. 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. Pass rates for each type of questions separately are presented in Table B.1 in the Online Appendix. The power of these revealed-preference tests are discussed in Online Appendix C.

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.[20]

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% of subjects in the market-stock task and 62% of subjects in the market-Ellsberg task are not inconsistent with PS.

Testing for the exact compliance with the model may be too demanding. It is possible that small mistakes could account for a subjects' deviation from SEU or MEU. We now turn to quantifying the degree of compliance with the models, using CCEI and minimal e as described in Section 2.

Distance measures. The Critical Cost Efficiency Index (CCEI) is a measure of the degree of compliance with GARP that is widely used in the recent experimental literature (e.g., Choi et al., 2014). In our laboratory 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

[20] 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	1.5066	1.2791	0.9169	1.4949	1.2588	0.9224
Market-Ellsberg	0.9892	1.0000	0.0317	1.3094	1.0000	0.9108	1.3038	1.0000	0.9105

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 cited studies had 25 and 50 budgets respectively. Most importantly, objective probabilities were not provided in our experiment.

We use e_* (minimal e; proposed by Echenique et al., 2018) as a measure of the degree of deviation from SEU. Remember that the number e_* is a perturbation to the model that allows SEU to accommodate the observed choices. It 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.

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 6 panel A. 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 above). The choices made by about three-quarters of the subjects are positively correlated between the two tasks (Figure B.1 in the Online Appendix), and 36% of those subjects exhibit statistically significant positive correlation (one-sided, at the 5% significance level). This correlation is reflected in the degree of violation of SEU—Figure 6 panel B shows that e_* from the two tasks are highly correlated (Spearman's correlation coefficient: $r = 0.406$ for treatment Large, $r = 0.583$ for treatment Small).

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

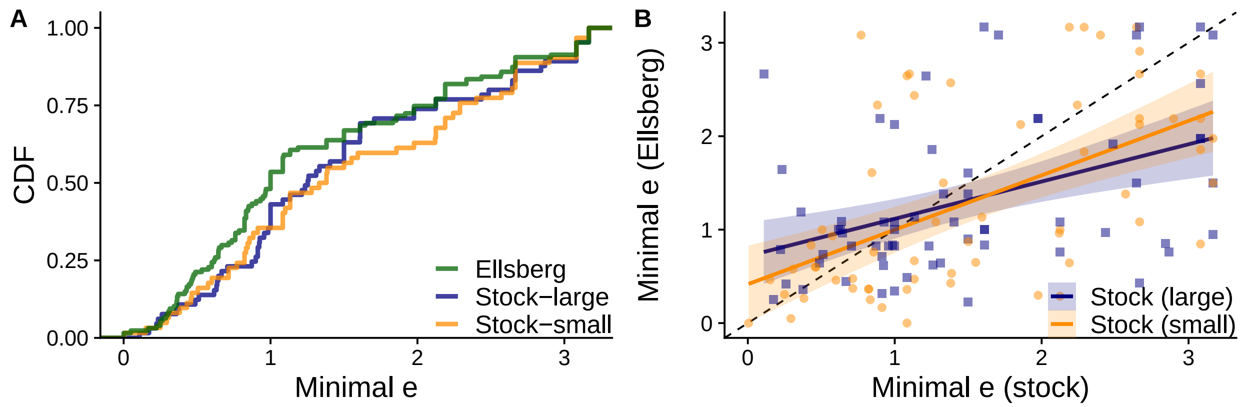

FIGURE 6: e_* from the market tasks. (A) Comparing e_* from market tasks with different sources of uncertainty (Ellsberg urn, stock price with large volatility, and stock price with small volatility). (B) Correlation between e_* from market-stock and market-Ellsberg tasks. Each dot represents a subject.

choices, as in the perturbation of the SEU model behind our calculation 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 are not determined pessimistically. Therefore, the MEU model cannot explain the subjects' choices better than SEU; the size of perturbation required for MEU is not much lower than that for SEU.

We do not observe gender differences on e_* but there is 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, however, is statistically significant only in the e_* from the market-stock task (two-sample t-tests; Market-stock: $t(57) = 1.50$, $p = 0.140$; Market-Ellsberg: $t(57) = 3.24$, $p = 0.002$). See Figure B.4 in the Online Appendix.

Ambiguity attitude. Finally, we look at the relation between behavior in the market tasks and subjects' attitudes toward ambiguity, measured using a standard Ellsberg-paradox design. As explained in Section 3.1, 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

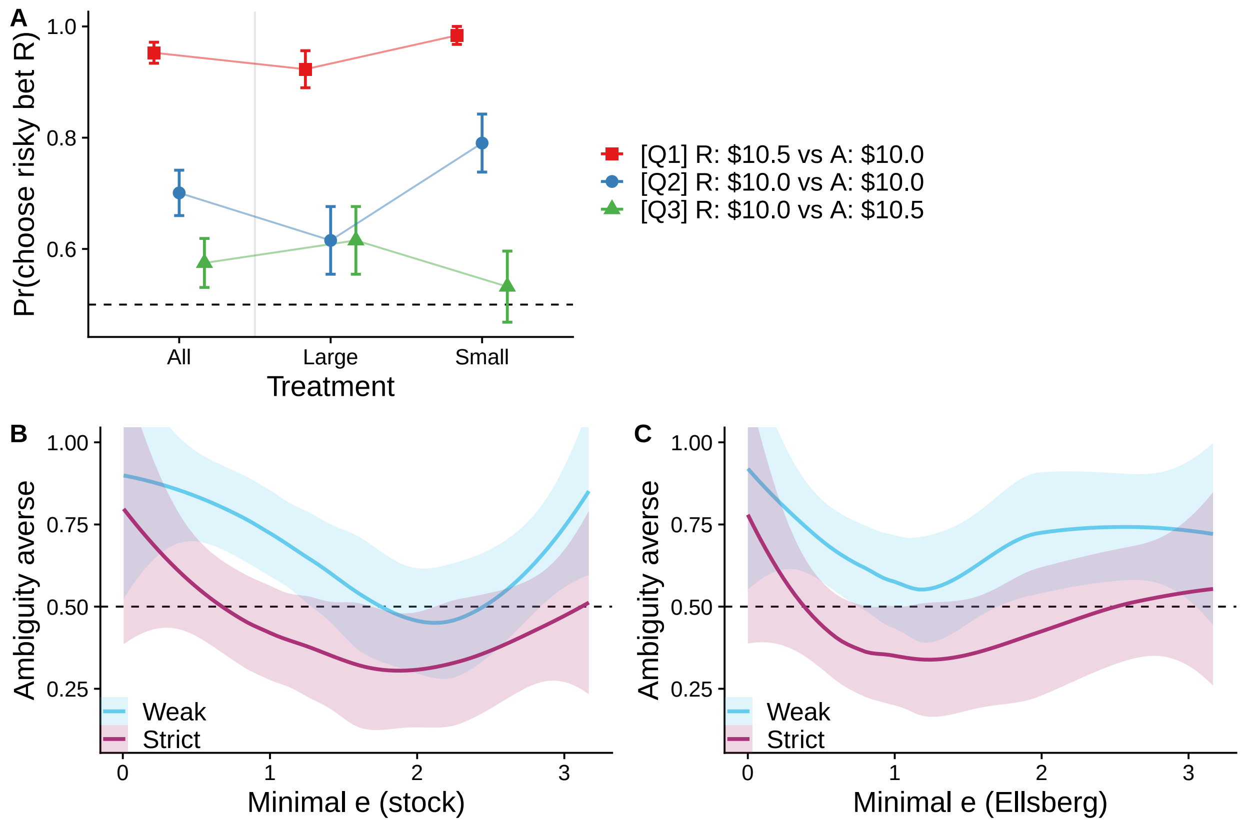

Figure 7: (A) Frequency of choosing a risky bet in each question in the standard-Ellsberg task in the laboratory data. Bars indicate standard errors of means. (BC) LOESS curves relating e_* and ambiguity attitude.

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.

4.2 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 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 consisting of subjects who used desktop or laptop computer (66%) as our "core" sample. Table 4 provides distributions of individual sociodemographic characteristics in the entire sample as well as the core sample and the excluded sample (those who did not use desktop or laptop computers). It is evident that the type of device used is correlated with some of the demographic variables (age, education level, and income level; chi-squared tests in the last column in Table 4). The sub-samples of subjects exhibited markedly different patterns of behavior as well (Online Appendix B.3). Throughout the rest of the paper, we analyze data from the core sample.[21]

The set of 20 budgets used in the market task is the 10-times scaled-up version of the one used in the laboratory (Figure 2, panel B). This keeps the relative prices the same between two studies, making the distance measure e_* comparable between data from the laboratory and the panel.

We start by checking event monotonicity, along the lines of our discussion for the laboratory experiment. 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 B.5 in the Online Appendix). There is no treatment difference. Our subjects also made choices that are, to some extent, responsive to underlying price changes: Figure 8 reports the degree of compliance with the downward-sloping demand property.

Revealed-preference tests, distance measures, and ambiguity attitude. Pass rates for GARP, SEU, and MEU presented in Table 5 are similar to those of the laboratory data presented

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

TABLE 4: Sociodemographic information.

Variable	All	Device: Desktop/laptop	Device: Tablet/mobile phone	Test
Gender				
Female	0.496	0.471	0.544	$\chi^2(1) = 3.36$
Male	0.504	0.529	0.456	$p = 0.0669$
Age group				
20-39	0.319	0.279	0.395	
40-59	0.353	0.345	0.369	$\chi^2(2) = 17.79$
60-	0.327	0.375	0.236	$p = 0.0001$
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	$\chi^2(3) = 53.7$
College or post-graduate	0.336	0.410	0.194	$p < 0.0001$
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	$\chi^2(4) = 43.97$
\$150k –	0.068	0.080	0.046	$p < 0.0001$
Occupation type				
Full-time	0.497	0.509	0.475	
Part-time	0.102	0.100	0.106	$\chi^2(2) = 0.78$
Not working	0.401	0.391	0.418	$p = 0.6759$
Marital status				
Married/live with partner	0.690	0.713	0.646	$\chi^2(1) = 3.23$
Other	0.310	0.287	0.354	$p = 0.0724$
# of observations in the sample	764	501	263	

above. 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 with GARP pushes the average CCEI score to 0.97 (Table 6). The average e_* of 1.610 is not statistically significantly different from the average 1.507 in the laboratory data (two-sample t-test, $t(626) = 1.133$, $p = 0.258$).

The pattern of choices in the standard-Ellsberg task is also similar to what we observed in the laboratory data, but the overall frequency with which the risky bet is chosen is smaller. In

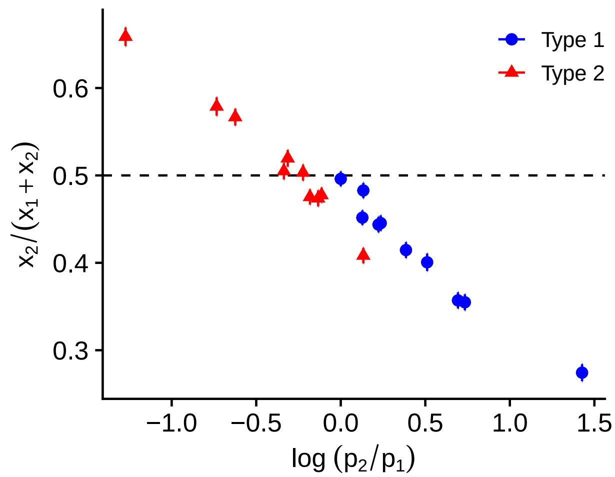

FIGURE 8: Downward-sloping demand at the aggregate level. *Notes*: Each point represents mean $x_2/(x_1+x_2)$ at each $\log(p_2/p_1)$ and bars indicate standard error of means.

TABLE 5: Pass rates.

Treatment	N	GARP	PS	SEU	MEU
Large volatility	245	0.4367	0.3959	0.0122	0.0122
Small volatility	256	0.4492	0.3945	0.0234	0.0273
Combined	501	0.4431	0.3952	0.0180	0.0200

Notes: 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. Pass rates for each type of questions separately are presented in Table B.2 in the Online Appendix. The power of these revealed-preference tests are discussed in Online Appendix C.

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 laboratory, 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 9). These numbers are lower than in the laboratory data. Now, using this classification, we look at the relationship between ambiguity aversion and e_*. Unlike Figure 7 panels B and C, using laboratory data, Figure 9 panel A exhibits a decreasing relation between the two (there is a slight indication of reflection around $e_* \approx 1.2$, 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.

Table 6: Distance measures.

Treatment	N	CCEI			e_* (SEU)			e_* (MEU)		
		Mean	Median	SD	Mean	Median	SD	Mean	Median	SD
Large volatility	245	0.9720	0.9950	0.0509	1.6194	1.5369	0.9057	1.6097	1.5000	0.9107
Small volatility	256	0.9688	0.9958	0.0552	1.6002	1.4750	0.9243	1.5969	1.4750	0.9261
Combined	501	0.9704	0.9954	0.0531	1.6096	1.5000	0.9144	1.6032	1.5000	0.9177

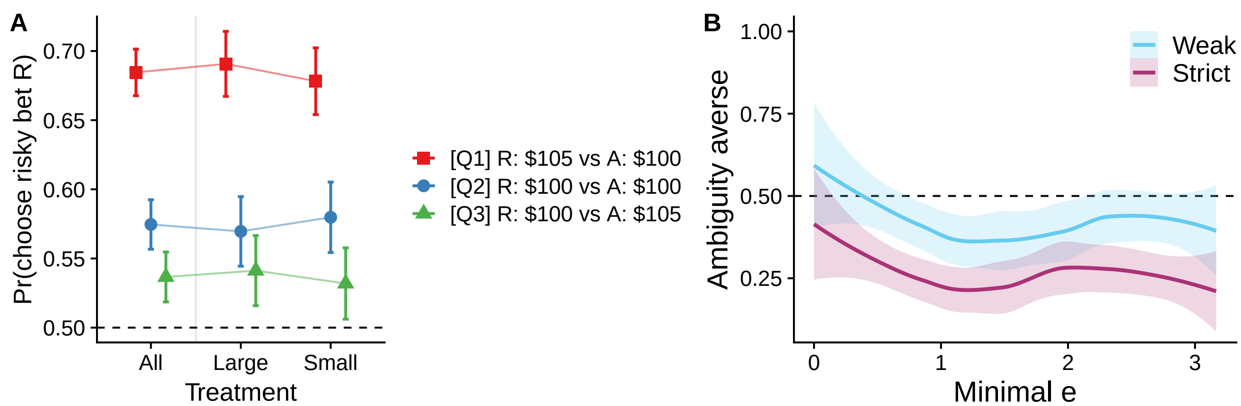

Figure 9: (A) Probability of choosing a risky bet in each question in the standard-Ellsberg task in the panel data. (B) LOESS curves relating e_* and ambiguity attitude.

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 other experiments.[22]

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 $x_2/(x_1+x_2)$, 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"),

[22]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 as measured by CRT is negatively associated with e_* 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."[23] 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. Two other 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).

Finally, we compare the distribution of e_* in the laboratory 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. It is evident from Figure 10 that there is no difference in distributions of e_* (p-values from two-sample Kolmogorov-Smirnov tests are all larger than 0.1). 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 (p-values from two-sample Kolmogorov-Smirnov tests are all below 0.01).

5 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 of deci-

[23] In contrast to these observations, Echenique et al. (2018) find that older subjects have larger e_* for objective expected utility (OEU) (i.e., further away from OEU, not SEU) than younger subjects. This is 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 regression	
	(1)	(2)	(3)	(4)
Dependent variable	e_*	ρ^{dsd}	Violate mon.	Weak AA
Treatment: Large	0.032	0.026	0.003	0.182
	(0.083)	(0.030)	(0.233)	(0.198)
Age: 40-59	−0.025	−0.013	−0.109	−0.134
	(0.110)	(0.040)	(0.316)	(0.263)
Age: 60+	0.054	−0.036	0.365	−0.249
	(0.116)	(0.042)	(0.319)	(0.288)
Financial literacy (UAS #1): High	0.093	0.034	−0.291	0.307
	(0.103)	(0.036)	(0.263)	(0.250)
Financial literacy (UAS #6): High	−0.244	−0.043	0.067	0.204
	(0.100)	(0.035)	(0.268)	(0.247)
CRT score (UAS #1): 1 correct answer	−0.028	−0.019	−0.362	0.436
	(0.098)	(0.034)	(0.264)	(0.232)
CRT score (UAS #1): 2 or 3 correct answers	−0.152	−0.006	−0.609	0.711
	(0.122)	(0.044)	(0.362)	(0.286)
Education: Some college	0.112	0.005	0.142	−0.070
	(0.133)	(0.047)	(0.342)	(0.331)
Education: Assoc. or pro. degree	−0.253	−0.087	−0.167	−0.026
	(0.130)	(0.046)	(0.374)	(0.324)
Education: College or postgraduate	−0.019	−0.032	−0.478	0.574
	(0.121)	(0.043)	(0.346)	(0.299)
Income: 25,000-49,999	0.244	0.098	−0.055	0.470
	(0.138)	(0.046)	(0.368)	(0.335)
Income: 50,000-74,999	0.418	0.126	0.635	0.071
	(0.142)	(0.048)	(0.374)	(0.353)
Income: 75,000-149,999	0.338	0.120	−0.112	0.226
	(0.142)	(0.049)	(0.414)	(0.349)
Income: 150,000+	0.290	0.141	−0.087	0.675
	(0.201)	(0.078)	(0.614)	(0.484)
Male	−0.129	−0.034	−0.130	0.277
	(0.087)	(0.030)	(0.247)	(0.210)
Working	0.053	0.003	−0.311	−0.309
	(0.097)	(0.034)	(0.299)	(0.250)
Constant	1.482	−0.483	−0.779	−1.237
	(0.163)	(0.057)	(0.431)	(0.430)
Observations	490	490	490	490
Adjusted R^2 / Log likelihood	0.031	0.003	-239.470	-309.069

Notes: Robust standard errors are presented in parentheses.

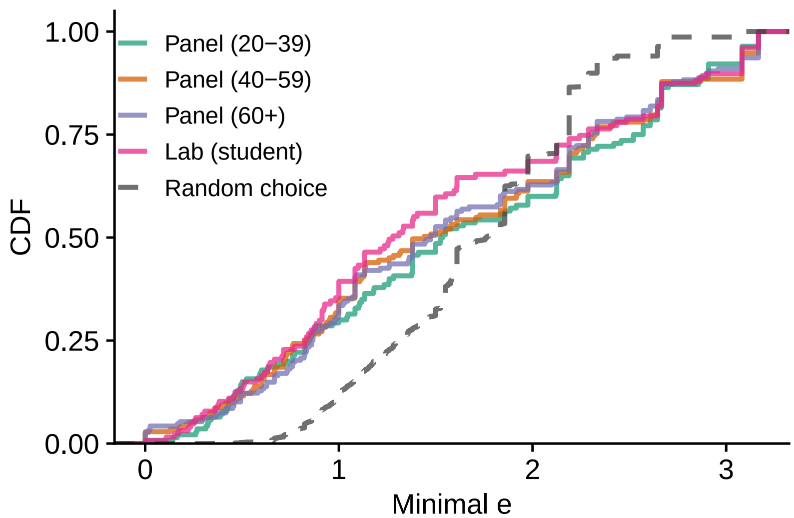

FIGURE 10: Comparing distributions of e_* from the panel study and the laboratory study.

sion making under uncertainty. 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 laboratory 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 choice under risk, not uncertainty) 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.

Further studies are necessary to fully understand the behavior in environments that are more "natural" than traditional artificial Ellsberg-style settings. Our nonparametric 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. (1967): "The Construction of Utility Functions from Expenditure Data," *International Economic Review*, 8, 67–77.

Ahn, D. S., S. Choi, D. Gale, and S. Kariv (2014): "Estimating Ambiguity Aversion in a Portfolio Choice Experiment," *Quantitative Economics*, 5, 195–223.

Anantanasuwong, K., R. Kouwenberg, O. S. Mitchell, and K. Peijnenberg (2019): "Ambiguity Attitudes about Investments: Evidence from the Field," NBER Working Paper No. 25561.

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., S. Lee, and M. Shum (2011): "The Money Pump as a Measure of Revealed Preference Violations," *Journal of Political Economy*, 119, 1201–1223.

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 (2019): "Time Discounting and Economic Decision-making in the Older Population," *Journal of the Economics of Ageing*, 14, 100121.

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. K.-H. QUAH, AND L. RENOU (2020): "Revealed Preferences over Risk and Uncertainty," *American Economic Review*, 110, 1782–1820.

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.

Varian, H. R. (1982): "The Nonparametric Approach to Demand Analysis," *Econometrica*, 945–973.