The Distribution of Ambiguity Attitudes

Hans-Martin von Gaudecker, Axel Wogrolly, Christian Zimpelmann

November 7, 2022

This paper analyzes the stability and distribution of ambiguity attitudes using a broad population sample. Using high-powered incentives, we collected six waves of data on ambiguity attitudes about financial markets—our main application—and climate change. Estimating a structural stochastic choice model, we obtain three individual-level parameters: Ambiguity aversion, likelihood insensitivity, and the magnitude of decision errors. These parameters are very heterogeneous in the population. At the same time, they are stable over time and largely stable across domains. We summarize heterogeneity in these three dimensions using a discrete classification approach with four types. Each group makes up 20-30 % of the sample. One group comes close to the behavior of expected utility maximizers. Two types are characterized by high likelihood insensitivity; one of them is ambiguity averse and the other ambiguity seeking. Members of the final group have large error parameters; robust conclusions about their ambiguity attitudes are difficult. Observed characteristics vary between groups in plausible ways. Ambiguity types predict risky asset holdings in the expected fashion, even after controlling for many covariates.

Keywords: ambiguity attitudes; temporal stability; domain specificity; socio-demographic factors; cluster analysis; household portfolio choice
JEL: D81, G41, C38, D14,

Affiliations: Gaudecker: University of Bonn and IZA – Institute of Labor Economics, hmgaudecker@uni-bonn.de; Wogrolly: Private Sector; Zimpelmann: IZA, zimpelmann@iza.org. **Acknowlegements:** We gratefully acknowledge support from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through CRC-TR 224 (Project C01) and under Germany's Excellence Strategy – EXC 2126/1 – 390838866. We would like to thank Steffen Altmann, Peter André, Sarah Auster, Luca Henkel, Thomas Neuber, Jonas Radbruch, and seminar and conference participants at University of Oxford, Copenhagen Business School, Society for Economic Measurement Conference 2019, German Economic Association Annual Meetings 2020, briq workshop on beliefs 2021, RUD 2021, D-TEA 2021, and Barcelona Summer Forum 2022 for very helpful comments. **Replication:** The replication material for the study is available at https://github.com/ChristianZimpelmann/replication-ambig-beliefs.

1 Introduction

People face ambiguity in many domains. How likely is it that the return on a portfolio of stocks is larger than some threshold for a certain horizon? What are the odds that an offered job will be sufficiently better than the current one? Will climate change make living at the current place of residence much harder during one's lifetime? In a large class of models, decisions in the face of ambiguity depend on two core parameters. Ambiguity aversion is the average dislike for ambiguous events compared to risky events with known probabilities. Likelihood insensitivity measures how strongly decisions react to changes in subjective beliefs about the ambiguous event; an alternative interpretation of this parameter is the degree of ambiguity. Decision-making under risk emerges as the special case where both parameters are irrelevant.

To what extent ambiguity aversion and likelihood insensitivity represent fundamental personal traits is, however, largely an open question. How stable are they over time and across domains? Do they vary in expected ways with observable characteristics in broad population samples? What is the connection between ambiguity attitudes and decisions in everyday life? This paper sheds light on these questions. In doing so, we address methodological questions on how to deal with decision errors when eliciting ambiguity attitudes and on how to best describe heterogeneity when traits are interdependent.

Six bi-annual waves of data on ambiguity attitudes in the domain of the stock market form the basis of our analysis. We collected this data in a probability-based sample of the Dutch population using substantial financial incentives (expected hourly compensation corresponded to 51 €). In one wave, we also included the domain of climate change. In total, we analyze data from almost 2,200 individuals or 11,000 person × wave observations.

In each wave, respondents faced a series of choices between receiving a prize with some known probability or receiving it in case an ambiguous event occurred. As an example, one such event consisted of an investment in a stock market index yielding a positive return over the upcoming six months. For seven events like this per wave, our design yields data on individuals' *matching probabilities*. For the matching probability, an individual is indifferent between receiving the prize with that probability and receiving it if the ambiguous event occurs.

Descriptively, five salient features emerge for matching probabilities. First, the sum of average matching probabilities for an event and its complement is less than one. This implies that, on average, subjects are averse to ambiguity. Second, average matching probabilities are *sub-additive* in the sense that the sum of matching probabilities of two mutually exclusive events exceeds the matching probability of their union. This means individuals are ambiguity averse for high-probability events and ambiguity seeking for low-probability events on average. Third, matching probabilities differ widely across subjects. Fourth, a non-negligible fraction of choice patterns violates set-monotonicity; i.e., choices reveal a higher matching probability for an event that is a strict subset of another. Such patterns cannot be rationalized by deterministic theories of choice under uncertainty. Fifth, the rate of set-monotonicity violations is highest for pairs of choices where—based on a separate question on the historical behavior of the stock market—individuals judge the past frequency of the event forming the subset to be large relative to that of the superset.

To account for these facts, we set up a stochastic choice model with three parameters of interest. Ambiguity aversion and likelihood insensitivity control the deterministic part of the model; the third parameter is the relative weight of its stochastic component. In a first step, we structurally estimate the model for each individual × wave observation separately. The stylized facts on matching probabilities are reflected in the marginal parameter distributions. On average, individuals are ambiguity averse. Likelihood insensitivity is quantitatively very important for the majority of observations. All parameters display large heterogeneity. For example, a substantial fraction of subjects display ambiguity seeking behavior on average. Most choice sequences cannot be fully rationalized by the deterministic model and the importance of the stochastic component turns out to be a key feature for describing different individuals' choice sequences.

We show that all three parameters are stable over time and largely stable across domains. Over time, the stability of ambiguity aversion and likelihood insensitivity is comparable to what previous literature finds for risk preferences. When accounting for attenuation due to measurement error, we find that there are no systematic changes in the sense that individuals' parameters in one time period are the best predictors for parameters in another period. Looking across the domains of finance and climate change, ambiguity aversion and the magnitude of decision errors are completely transferable in this sense. Transferability is lower for likelihood insensitivity. These results suggest that ambiguity aversion is a domain-invariant preference parameter but that individuals have different degrees of trust in their probability judgments in different domains (or that they perceive a different level of ambiguity).

Imposing stability of preferences, we find that a clustering approach is a useful way to describe parameter heterogeneity. From an ex-ante perspective, it does not place any restrictions on the joint distribution of the three parameters and thus accounts for the non-separable nature of the model. Empirically, we find that four groups—each of which has a share of 20-30 %—summarize broad choice patterns well. One type is fairly close to the behavior of subjective expected utility maximization; ambiguity aversion and likelihood insensitivity play limited roles. For two groups, likelihood insensitivity is large. They differ in their attitude toward ambiguity. The first of the two displays substantial aversion on average, the other one a slight preference for it. For the three groups described so far, the deterministic part of the model has high explanatory power. The stochastic element plays a much more important role for the last group, which is thus characterized by very noisy decision-making; choice patterns in that group do not reveal much about ambiguity attitudes.

Individual characteristics differ in sensible ways across the four groups. For example, subjects behaving close to subjective expected utility maximization are the most educated, display the highest level of numeracy, and the lowest risk aversion. The groups classified to be ambiguity averse and ambiguity seeking, respectively, are similar in many dimensions of observed characteristics, often assuming intermediate positions. There are exceptions for the ambiguity averse group, which has a high share of females, the lowest financial wealth, and ceteris paribus the highest risk aversion. Finally, the members of the group whose decision-making is noisiest are the oldest, and they have the lowest average levels of education and numeracy.

The preference groups predict portfolio choice behavior. This holds true even after conditioning on a large set of observable characteristics, including financial wealth and risk aversion. We consider two measures of portfolio choice: Whether people hold risky assets and the share invested into these. The group closest to subjective expected utility maximization has the riskiest portfolios according to both measures; the ambiguity averse group takes the least amount of risk.

Our paper is related to various strands of the literature. The importance of distinguishing between uncertainty and risk has been introduced by Keynes (1921) and Knight (1921). Ellsberg (1961) showed deviations from the subjective expected utility paradigm in a controlled empirical setting. Based on those considerations, a burgeoning theoretical literature has produced tractable models of choice under ambiguity (e.g., Gilboa and Schmeidler, 1989; Ghirardato and Marinacci, 2001; Chateauneuf, Eichberger, and Grant, 2007). Our empirical specification is directly based on these models.

Recent advances in measurement techniques (Baillon, Huang, Selim, and Wakker, 2018; Baillon, Bleichrodt, Li, and Wakker, 2021) have made it possible to elicit ambiguity attitudes for domains that go beyond highly stylized settings such as the famous Ellsberg urns. We adapt these methods for use in a broad population survey by simplifying individual decisions, which are all binary choices.

We contribute to the literature examining empirical estimates of ambiguity attitudes. Early papers summarized in Trautmann and van de Kuilen (2015) have mostly focused on working out stylized facts such that on average, behavior is ambiguity seeking for low probability gain events and ambiguity averse for high probability events. More recent studies based on laboratory experiments have focused on limitations to measurement (Baillon, Halevy, and Li, 2022b), the interpretation of parameters (Henkel, 2022), their stability over time (Duersch, Römer, and Roth, 2017) and across domains (Li, Müller, Wakker, and Wang, 2018), or learning (Baillon, Bleichrodt, Keskin, l'Haridon, and Li, 2018). Most directly related to our paper are cross-sectional studies in broader samples. They document large heterogeneity of attitudes (Dimmock, Kouwenberg, Mitchell, and Peijnenburg, 2015; Anantanasuwong, Kouwenberg, Mitchell, and Peijnenburg, 2020) and show connections of ambiguity preferences with portfolio choices (Dimmock, Kouwenberg, Mitchell, and Peijnenburg, 2016; Anantanasuwong et al., 2020). We replicate many of these findings. Based on our unusually large dataset, we are able to estimate the parameters more precisely and unify several conflicting pieces of prior evidence.

We show that one reason for us to be able to do so is that we make use of an explicit stochastic choice model. Doing so has a long tradition in the estimation of risk preferences (e.g., Harless and Camerer, 1994; Hey and Orme, 1994; Loomes and Sugden, 1995; Gaudecker, Soest, and Wengström, 2011; Apesteguia and Ballester, 2021) whereas prior work on ambiguity attitudes has focused on deterministic components of choice.

Another reason is that prior work looking at parameter heterogeneity and behavioral consequence has focused on marginal parameter distributions. This approach has limits because the preference parameters are inherently non-separable. If a decision-maker does not perceive any ambiguity for a given event, her ambiguity aversion does not play a role. Similarly, if the stochastic component is very important, changing the parameters of the deterministic component will hardly alter the power of the model to explain data. Modelling parameter heterogeneity as a discrete distribution in nonlinear models is a common approach in other strands of the literature (e.g., Heckman and Singer, 1984; Keane and Wolpin, 1997). We make use of clustering techniques introduced more recently into econometrics (Bonhomme and Manresa, 2015), which are computationally favorable.

In the next section, we sketch a framework for interpreting decisions under ambiguity and describe our design and the resulting data, including the descriptive facts on matching probabilities. Section 3 presents our structural model and the results for wave-by-wave parameter estimates, establishing the properties for their stability over time and across domains. In Section 4, we classify individual-specific parameters into types and describe these types' relation to personal characteristics and portfolio choice behavior. That section also examines robustness to various specification choices and provides a detailed comparison with the literature. We discuss the findings in Section 5.

2 Ambiguity framework and data

In this section, we first sketch the framework we use to define ambiguity attitudes. We focus on the interpretation of two key parameters. Next, we introduce our version of the paradigm by Baillon, Huang, et al. (2018), which we implemented in the LISS panel. In Section 2.3, we describe some stylized facts in our data on ambiguity attitudes, which include up to six waves for 2,177 respondents, collected over a period of three years. These key facts will guide our empirical strategy in Section 3.1 below. In between those two sections, we briefly describe additional variables that will be important for our analyses in Section 2.4.

2.1 Definition of ambiguity attitudes and parameter interpretation

We focus on prospects—i.e., state-contingent outcomes as in Wakker (2010)—which pay out $x > 0$ if event $E \in \Omega$ occurs and nothing otherwise, denoting such prospects as $x_E 0$. Decision-makers value monetary quantities according to a utility function $u(\cdot)$. We normalize $u(0) = 0$ and assume that $u(x) > 0$. Using the biseparable utility framework of Ghirardato and Marinacci (2001), a decision-maker evaluates the prospect $x_E 0$ as $W(E) \cdot u(x)$. Her event weighting function $W(E)$ satisfies $W(\emptyset) = 0$, $W(\Omega) = 1$, and set-monotonicity in the sense that $B \subseteq A \implies W(B) \leq W(A)$.

Following Abdellaoui, Baillon, Placido, and Wakker (2011), we assume that decision weights depend on subjective probabilities $\Pr_{\text{subj}}(E)$ and the source of uncertainty S (e.g., an urn with an unknown distribution of balls, the future evolution of the stock market, or the path that will be taken by the earth's climate). $W(E)$ then boils down to how decision weights depend on subjective probabilities for a particular source of uncertainty; it is thus called the source function (Wakker, 2010). In this model, Baillon, Huang, et al. (2018) and Baillon, Bleichrodt, Li, et al. (2021) define two parameters describing ambiguity attitudes, both of which are zero for subjective expected utility maximizers:

$$\text{Ambiguity aversion} \qquad \alpha^S = \mathbb{E}[\Pr_{\text{subj}}(E) - W(E)], \tag{1}$$

$$\text{Likelihood insensitivity} \qquad \ell^S = 1 - \frac{Cov(W(E), \Pr_{\text{subj}}(E))}{Var(\Pr_{\text{subj}}(E))}. \tag{2}$$

Ambiguity aversion is the average amount by which subjective probabilities exceed decision weights. Decision-makers with $\alpha^S = 0$ are ambiguity neutral on average; negative values indicate a dominance of ambiguity seeking behavior. Likelihood insensitivity captures the extent to which individuals' decision weights change when the underlying subjective probabilities change. In certain multiple-prior models (Ghirardato, Maccheroni, and Marinacci, 2004; Dimmock, Kouwenberg, Mitchell, et al., 2015; Alon and Gayer, 2016), likelihood insensitivity can be interpreted as the perceived level of ambiguity. See Online Appendix A for more details on the ambiguity framework and different interpretations.

For our main results, we further assume that $W(E)$ is *neo-additive* (Chateauneuf, Eichberger, and Grant, 2007):

$$\begin{aligned} W(E) &= \tau_0^S + \tau_1^S \cdot \Pr_{\text{subj}}(E) \text{ for } \Pr_{\text{subj}}(E) \in (0,1) \\ W(E) &= 0 \text{ for } \Pr_{\text{subj}}(E) = 0 \\ W(E) &= 1 \text{ for } \Pr_{\text{subj}}(E) = 1 \\ 0 &\leq \tau_1^S, 0 \leq \tau_0^S \leq 1 - \tau_1^S \end{aligned} \tag{3}$$

Neo stands for "non-extreme outcome", i.e., weights are zero (one) for events the decision-maker considers impossible (certain); they are linear in $\Pr_{\text{subj}}(E)$ in between. We chose this functional form because of its tractability and good empirical performance (Li et al., 2018). For the neo-additive weighting function, α^S and ℓ^S have very simple representations:

$$\alpha^S = \frac{1 - 2\tau_0^S - \tau_1^S}{2}, \tag{4}$$

$$\ell^S = 1 - \tau_1^S. \tag{5}$$

Alternatively, Baillon, Huang, et al. (2018) and Baillon, Bleichrodt, Li, et al. (2021) show that α and ℓ can be estimated under different assumptions using the empirical analogues of the moments in Equations (1)-(2). We will pursue that as a robustness check and comment on the relative merits in Section 3.1, after having introduced the structure of our data.

2.2 Measuring ambiguity attitudes

In order to measure ambiguity attitudes, we adapt the method developed by Baillon, Huang, et al. (2018) and Baillon, Bleichrodt, Li, et al. (2021) for use in a general population. Our main source of uncertainty is the Amsterdam Exchange Index (AEX), the most widely known stock market index in the Netherlands. We expect individuals to differ in their perception of the AEX. For some, probabilities may be close to objective. Others might perceive substantial uncertainty regarding its evolution.

Eliciting attitudes about ambiguous events is cognitively demanding for participants. To keep this burden low, we confront subjects with binary choices only. Going through a tutorial introducing the choice situations and potential payoff consequences was mandatory in the initial survey round. In later waves, the tutorial was optional, but advertised prominently. Compared to a choice list format as in Baillon, Huang, et al. (2018), we expect this procedure to reduce complexity as subjects can focus on one question at a time.

Individuals make a series of choices, all of which share the structure shown in Figure 1. Each decision is between a bet on an event relating to the performance of the AEX over the subsequent six months and a lottery with known probabilities. In the example in Figure 1, Option 1 pays out € 20 if a hypothetical € 1,000 investment in the AEX is worth more than € 1,100 six months in the future. Option 2 is a lottery and pays € 20 with probability 50 %. The lottery is introduced as a wheel of fortune during the tutorial and it is spun when determining payoffs.

Depending on her choice between the AEX event and the lottery, a subject is presented with another choice with the same AEX event and a different lottery. If the subject choose the AEX event, we increase the winning probability of the lottery and vice versa. For each event, subjects make three to four binary choices (see Online Appendix Figure B.1 for the entire decision tree). Our data identify an interval for the *matching probability* where the length of the interval will be between 0.01 and 0.1, depending on the path taken in the decision tree.

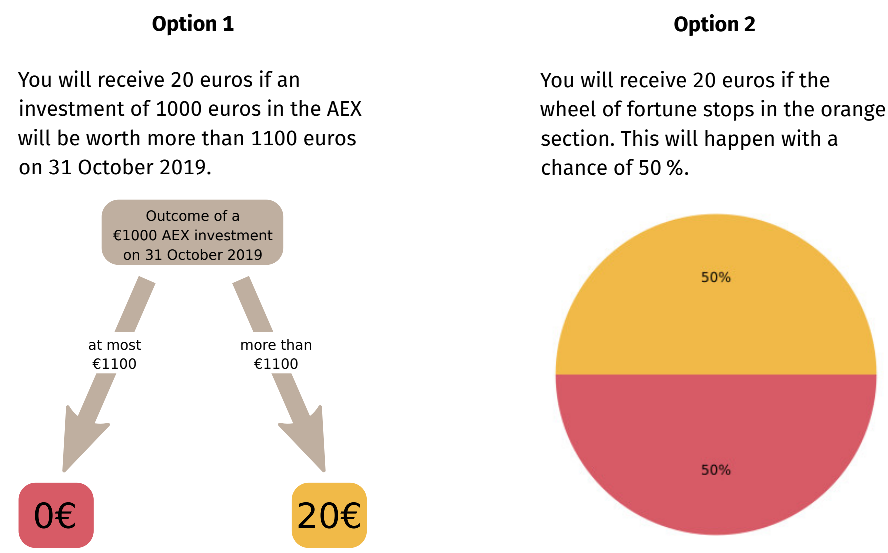


Figure 1. Exemplary binary choice situation

Notes: Labels are translated from Dutch to English. The date refers to the data collection during the month of May 2019.

Definition 1 (Matching probability)**.** The matching probability $m(E)$ of an event E is the probability p that makes a decision-maker indifferent between a pay-out of x if event E occurs and a bet on a lottery that pays x with probability p and zero otherwise.

For the ambiguity model sketched in the previous section and many others, Dimmock, Kouwenberg, and Wakker (2016) show that matching probabilities are useful for analyzing ambiguity attitudes because they are independent of utility function parameters and any weighting of probabilities.

The remainder of our design closely follows Baillon, Huang, et al. (2018). We partition the space of possible values the AEX investment can take into three events: $E_1^{AEX}: Y_{t+6} \in (1100, \infty]$, $E_2^{AEX}: Y_{t+6} \in [0, 950)$, and $E_3^{AEX}: Y_{t+6} \in [950, 1100]$, see Figure 2. This partition leads to balanced historical 6-month returns of the AEX with empirical frequencies in the 1999-2019 period of 0.24, 0.28, and 0.48, respectively. We elicit matching probabilities for each of these events along with their complements. Additionally, we include the event $E_0^{AEX}: Y_{t+6} \in (1000, \infty]$. As it comprises all outcomes for which the AEX is not declining, it is arguably the most intuitive event and should ease the entry for participants.

If we selected one of the answered questions for pay-out ex-post, the chained design would not be incentive compatible. Inspired by Bardsley (2000) and Johnson, Baillon, Bleichrodt, Li, van Dolder, et al. (2021), we let subjects start a random number generator to select the question to be paid out before they make any decisions. The selected question was displayed as a meaningless sequence of characters.

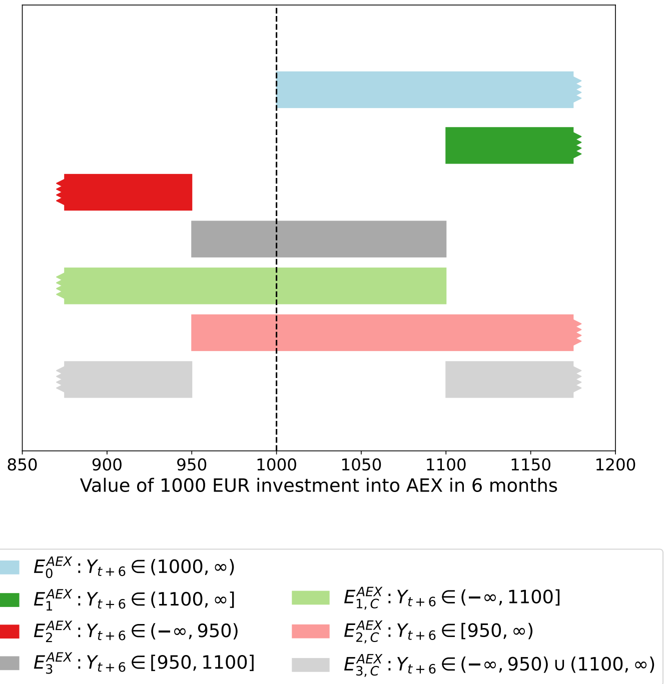


Figure 2. Events of AEX performance used in the experiment

If the subject did not encounter the selected choice situation during the questionnaire—i.e., she took a different branch in the decision tree—we presented it after all other decisions had been made. Pre-selection of the choice to be paid out makes it difficult for subjects to hedge against the encountered ambiguity (Baillon, Halevy, and Li, 2022a).[1] For every subject in our experiment, we either played out a lottery or checked the evolution of the AEX after six months, i.e., no additional randomness was introduced by paying only a fraction of subjects. Expected incentive payments for a expected utility decision-maker using empirical frequencies for stock returns were € 13.50. At the median response time, this amounts to an hourly wage of € 51.

We implement the elicitation in the LISS (Longitudinal Internet studies for the Social Sciences) panel administered by CentERdata (Tilburg University, The Netherlands). The LISS panel is a representative sample of Dutch individuals who partic-

1. Baillon, Halevy, and Li (2022b) showed that measuring ambiguity attitudes might not be possible at all for when paying out one choice at random. In their data, some subjects appear to integrate all decisions, creating a hedge against ambiguity. We do not think that this is much of a concern in our data because there is no direct hedge for the event E_0^{AEX}, described just below. Any strategy integrating the seven different events in a way that would yield a perfect hedge against ambiguity would require substantial cognitive effort. Furthermore, individuals did not have the required information on the structure of the design in the first wave and we do not see a sharp decline in ambiguity aversion in the subsequent wave (see Table 4 below). Hence, we feel comfortable with the assumption that respondents isolated their decisions across events.

ipate in monthly Internet surveys. The panel is based on a true probability sample of households drawn from the population register. Households that could not otherwise participate are provided with a computer and Internet connection. A longitudinal survey is fielded in the panel every year, covering a large variety of domains. Respondents are financially compensated for all questions they answer. On top of that, every respondent had the chance of earning an additional €20 in our experiment.

We collected six waves of data from November 2018 to May 2021. In November 2019, we additionally included a similar design where the source of uncertainty was the climate in the Netherlands over the subsequent winter. For example, $E_1^{climate}$ referred to the average temperature over the months of December, January, and February being at least 1° Celsius higher than the average temperature over the previous five winters. We chose the evaluation dates for the AEX (and the temperature) such that we could determine payoffs at the start of the subsequent wave. For the example in Figure 1, the subsequent questionnaire in November 2019 would start by showing that screen. We then revealed the value of an AEX investment made on the date somebody took the questionnaire (or the temperature) and played out the lottery (by spinning the "wheel of fortune" on the right). Each participant whose choice turned out to be winning received €20.

2.3 Data on ambiguity attitudes

In line with the domain of our application, we invited the financial deciders of households to participate. Initial invitations went out to 2,773 individuals, 2,407 of whom completed the questionnaire in at least one wave. Unless they dropped out of the LISS panel altogether, we invited respondents for each new wave regardless of their participation status in prior waves. We exclude subjects who seemingly did not spend time with the contents of the questionnaire. In particular, we drop a subject's data for one wave if she chose the same option (AEX or lottery) in all choices and her response time was below the 15th percentile. This condition affects 2 % of person × wave observations. To keep a similar sample for all our analyses—including those geared at stability over time—we require two waves with choice data meeting our inclusion criteria. Our final sample consists of 2,177 unique subjects, with 1,702–1,991 responses per wave; see Online Appendix Table D.1 for more details.

Since event-specific average matching probabilities are fairly stable across waves (see Online Appendix D.2, where we provide detailed statistics on matching probabilities), Table 1 pools all waves for summary statistics at the event-level. Table 2 shows statistics on set-monotonicity violations. We observe five salient features.

First, the sum of the average matching probabilities of an event and its complement is less than 1. Similar to findings for Ellsberg (1961) urns, this pattern indicates that matching probabilities are not equal to subjective probabilities; individuals are ambiguity averse on average. This is in line with findings in Dimmock,

Table 1. Matching probabilities, empirical frequencies, and judged historical frequencies

	Mean	Std. Dev.	$q_{0.1}$	$q_{0.5}$	$q_{0.9}$	Empir. Freq. '99-'19	Judged Freq., '99-'19
$E_0^{AEX} : Y_{t+6} \in (1000, \infty)$	0.49	0.27	0.075	0.45	0.93	0.63	0.52
$E_1^{AEX} : Y_{t+6} \in (1100, \infty]$	0.35	0.25	0.03	0.35	0.65	0.24	0.31
$E_{1,C}^{AEX} : Y_{t+6} \in (-\infty, 1100]$	0.51	0.29	0.075	0.45	0.97	0.76	
$E_2^{AEX} : Y_{t+6} \in (-\infty, 950)$	0.37	0.26	0.03	0.35	0.75	0.28	0.22
$E_{2,C}^{AEX} : Y_{t+6} \in [950, \infty)$	0.55	0.29	0.15	0.55	0.97	0.72	
$E_3^{AEX} : Y_{t+6} \in [950, 1100]$	0.56	0.28	0.15	0.55	0.97	0.48	0.47
$E_{3,C}^{AEX} : Y_{t+6} \in (-\infty, 950) \cup (1100, \infty)$	0.42	0.27	0.075	0.45	0.85	0.52	

Notes: Events were asked about in the order $E_0^{AEX} \cdot E_1^{AEX} \cdot E_2^{AEX} \cdot E_3^{AEX} \cdot E_{1,C}^{AEX} \cdot E_{2,C}^{AEX} \cdot E_{3,C}^{AEX}$, see Figure 2. Matching probabilities are set to the midpoint of the interval identified by the design. Data for 2,177 subjects are pooled across all waves. The next-to-last column shows the frequency of each event over half-year horizons in the 1999-2019 period. The last column contains subjects' average estimates thereof, which were elicited in May 2019 (see Section 2.4). Judged frequencies are available for 1952 subjects in our sample. Online Appendix D.2 provides more statistics on matching probabilities including variation across waves.

Kouwenberg, and Wakker (2016) and Anantanasuwong et al. (2020, both studies are also based on broad population samples) while Baillon, Huang, et al. (2018) observe ambiguity seeking choices on average in a time pressure task among students.

Second, mean matching probabilities are sub-additive for composite events. E.g., the sum of the matching probabilities of E_1^{AEX} and E_2^{AEX} is well above the average matching probability of their union, $E_{3,C}^{AEX}$. Sub-additivity implies that on average, subjects are likelihood-insensitive. This is a very robust finding in studies based on Ellsberg urns (e.g. Dimmock, Kouwenberg, and Wakker, 2016), as well as natural events (e.g. Li, 2017; Baillon, Bleichrodt, Keskin, et al., 2018).

Third, there is large variation across individuals for all matching probabilities. Interdecile ranges vary between 0.57 and 0.82, with an average of 0.74. This fact reveals large heterogeneity in response patterns. Standard deviations in our sample line up with related designs in Dimmock, Kouwenberg, and Wakker (2016) and Li (2017), who report values between 0.24 and 0.33.

Fourth, violations of set-monotonicity are prevalent. From Figure 2, it is easy to see that eight pairs of events bear the potential of such violations.[2] The first column of Table 2 shows that the set-monotonicity violation rate over all waves and superset-subset pairs is 14 %. Slicing the data in a different way, for each wave, 55 % of individuals violate set-monotonicity at least once (see Table D.4 in the Online Appendix). While substantial, such frequencies are anything but uncommon in general subject pools (see, for example Gaudecker, Soest, and Wengström (2011) for risky

2. The superset-subset pairs are $E_0^{AEX} \supset E_1^{AEX}$, $E_{1,C}^{AEX} \supset E_2^{AEX}$, $E_{1,C}^{AEX} \supset E_3^{AEX}$, $E_{2,C}^{AEX} \supset E_0^{AEX}$, $E_{2,C}^{AEX} \supset E_1^{AEX}$, $E_{2,C}^{AEX} \supset E_3^{AEX}$, $E_{3,C}^{AEX} \supset E_1^{AEX}$, and $E_{3,C}^{AEX} \supset E_2^{AEX}$.

Table 2. Judged historical frequencies and set-monotonicity violations

	Dependent variable: Set-monotonicity violation			
	(1)	(2)	(3)	(4)
Intercept	0.14^{***}	0.17^{***}		
	(0.0024)	(0.003)		
Judged frequencies (superset - subset)		-0.076^{***}	-0.044^{***}	-0.037^{***}
		(0.0055)	(0.0054)	(0.006)
Superset-subset pair fixed effects	No	No	Yes	Yes
Individual fixed effects	No	No	No	Yes
Observations	15616	15616	15616	15616

Notes: This table reports OLS regressions on the subject × superset-subset pair level. The dependent variable is the rate of set-monotonicity violations, averaged across waves. Set-monotonicity is violated if the lower bound of the interval elicited for the matching probability of the subset is strictly larger than the upper bound of the corresponding interval of the superset. The first column reports the average set-monotonicity violation rate. The remaining columns include the distance in judged historical frequencies over the 1999-2019 period for the two events in a superset-subset pair (elicited in May 2019, see Table 1 and Section 2.4 below). Column 3 adds superset-subset pair fixed effects and column 4 additionally adds individual fixed effects. Standard errors are clustered at the individual level. Sample: All waves meeting our inclusion criteria (i.e., there is variation across options and/or completion time is outside the fastest 15 %, see Section 2.3) for individuals with at least two such waves, who completed the May 2019 survey.
$* - p < 0.1, ** - p < 0.05, *** - p < 0.01$.

choices or Dimmock, Kouwenberg, and Wakker (2016) and Anantanasuwong et al. (2020) for ambiguity attitudes). We view violations of set-monotonicity as prima-facie evidence for decision errors. That is, they are unlikely to reflect preferences but rather carelessness or difficulties in understanding the tasks.

Fifth, set-monotonicity errors occur more often when individuals judge the past frequency of the event that forms the the subset to be large relative to that of the superset. In May 2019, we asked individuals to state the empirical frequency of the events we also use during elicitation of ambiguity attitudes. The remaining columns of Table 2 add the difference in judged historical frequencies between superset-subset pairs as an explanatory variable. The relation is clearly negative, no matter whether we add fixed effects for superset-subset pairs and individuals. The negative coefficients imply that for superset-subset pairs where the difference between the judged frequency of the superset and the subset is large, the likelihood of set-monotonicity errors tends to be low. For example, from Table 1 we see the average frequencies $E_{1,C}^{AEX} = 0.69$, $E_2^{AEX} = 0.22$, and $E_3^{AEX} = 0.47$. The resulting average set-monotonicity violations are 0.1 for $E_{1,C}^{AEX} \supset E_2^{AEX}$ and 0.24 for $E_{1,C}^{AEX} \supset E_3^{AEX}$ (see Online Appendix Table D.4). Hence, if two events are rather similar in subjects' memory, set-monotonicity violations are more likely to occur.

The first four stylized facts are also present in the data collected with climate change as the source of uncertainty (see Online Appendix Table D.3). We cannot check the fifth stylized fact because we did not ask about historical frequencies. Mean matching probabilities of complementary events add up to less than 1. Matching probabilities are sub-additive for composite events on average. Interdecile ranges

Table 3. Descriptive statistics on key variables

	N Subj.	Mean	Std. Dev.	$q_{0.25}$	$q_{0.5}$	$q_{0.75}$
Female	2177	0.5				
Education: Lower secondary and below	2177	0.26				
Education: Upper secondary	2177	0.34				
Education: Tertiary	2177	0.4				
Age	2177	57	16	45	59	69
Monthly hh net income (equiv., thousands)	2110	2.2	1	1.6	2.1	2.8
Total hh financial assets (equiv., thousands)	1727	39	120	2.6	12	34
Owns risky financial assets	1727	0.2				
Share risky financial assets (if any)	338	0.35	0.26	0.12	0.29	0.52
Risk aversion index	2121	−0	1	−0.68	−0.026	0.67
Numeracy index	2064	0	1	−0.55	0.27	0.78
Understands climate change	1936	0.54	0.21	0.5	0.5	0.75
Threatened by climate change	1936	0.55	0.22	0.4	0.6	0.6

Notes: Sample: Individuals with at least two waves meeting our inclusion criteria (i.e., there is variation across options and/or completion time is outside the fastest 15 %, see Section 2.3). Net income and assets are pooled within partners and equivalized, data from 2018. Risk aversion and numeracy are normalized to have mean zero and unit variance. The variables concerning climate change are normalized such that they vary between 0 and 1.

are even larger than for the AEX, with an average of 0.85. Set-monotonicity violations are just as prevalent as in the case for the AEX (see Online Appendix Table D.4).

2.4 Background characteristics

The LISS panel allows individual-level linkage of our choice data with a variety of information collected about the LISS panel members. This includes background information from regular surveys and additional questionnaires we ran ourselves. Table 3 shows the socio-demographic composition of our sample, variables relating to personal finances, and additional measures we collected. More detailed statistics can be found in Online Appendix D; our questionnaires are documented in Online Appendix B.

Socio-economic characteristics. The gender split is even. The average age is close to 57 years with ample variation. The share of subjects with tertiary education is 40 %; another 34 % hold an upper secondary degree. Net household income—pooled within households and equivalized using the square root of adults in the household—amounts to € 2,200 per month. Financial assets are equivalized in the same way. Our measure includes assets kept in joint accounts and assets assigned to the respondent (i.e., the person identifying as being most familiar with the household's finances); it does not include assets solely owned by the partner.

Risky asset holdings. 20 % of our sample directly hold risky assets which include among others individual stocks, funds, and bonds (we provide more detail in Online Appendix D.4). Conditional on owning risky assets, the average share is 35 %.

Judged historical frequencies of past AEX returns. In May 2019, we asked individuals to judge how frequently the AEX events used in our designs ($E_0^{AEX}, E_1^{AEX}, E_2^{AEX}, E_3^{AEX}$) occurred over the previous 20 years. Although there is substantial individual heterogeneity, the last column of Table 1 shows that the average judged frequencies are not too far from the empirical frequencies. Subjects underestimate the frequency of positive returns on average but think that returns greater than 10 % occurred more often than they did.

Risk Aversion. We measure households' risk aversion using the preference survey module developed by Falk, Becker, Dohmen, Huffman, and Sunde (2022). The module includes a general risk question and a quantitative component that is based on elicited certainty equivalents for risky lotteries. We combine the qualitative and quantitative components as suggested in Falk et al. (2022). Risk aversion bears the same relation to observed characteristics as in prior literature (e.g., Dohmen, Falk, Huffman, Sunde, Schupp, et al., 2011; Gaudecker, Soest, and Wengström, 2011): Older, lower income, and female subjects tend to be more risk averse (see Online Appendix Table D.5).

Numeracy. We measure three dimensions of numeracy: First, a basic numeracy component that is, e.g., used in the English Longitudinal Study of Ageing (Steptoe, Breeze, Banks, and Nazroo, 2013); second, a financial numeracy component that involves interest rates and inflation (a subset of the questions of Rooij, Lusardi, and Alessie (2011)); third, a probability numeracy component proposed by Hudomiet, Hurd, and Rohwedder (2018), which tests both basic understanding of probabilities and more advanced concepts such as independence and additivity. We aggregate the three components into a numeracy index, giving equal weight to each component. Our aggregated measure of numeracy is related to socio-demographics in similar ways as has been shown for its components in other settings (e.g., Rooij, Lusardi, and Alessie, 2011; Hudomiet, Hurd, and Rohwedder, 2018, also see Table D.5)

Knowledge of and concern about climate change. To help analyze ambiguity attitudes toward climate change, we asked subjects to report (i) their perceived understanding of the causes and implications of climate change and (ii) whether climate change is a threat to them and their family on Likert scales. We normalize the variables such that they vary between 0 and 1.

3 Estimation strategy, marginal parameter distributions, and stability

The stylized facts in Section 2.3 showed that on aggregate, behavior is indicative both of ambiguity aversion and likelihood insensitivity. At the same time, heterogeneity in matching probabilities is large. Decision errors are frequent and more likely for events that people judge to have been closely related in the past. Our empirical strategy, described next, takes these features into account in a stochastic model of choice. Its key parameters are ambiguity aversion, likelihood insensitivity, and the variance of decision errors.

Section 3.2 describes the distributions of wave-by-wave estimates of these parameters. We find that all of them are important in determining behavior and that they are very heterogeneous across subjects. Section 3.3 shows that there are no systematic changes within individuals over time.

Section 3.4 adds the survey using climate change and asks to what extent the estimated parameters are stable across completely different sources of uncertainty. Ambiguity aversion turns out to be transportable directly and this is largely true for decision errors, too. In contrast, likelihood insensitivity is more specific to a particular source of uncertainty.

3.1 Empirical strategy

We estimate the neo-additive model at the individual level, which allows us to match average levels of ambiguity aversion and likelihood insensitivity while respecting the large heterogeneity in the data. Because frequent set-monotonicity violations increase in the perceived similarity of two events in the past, we augment the deterministic model with an additive error term, also known as a Fechner error (e.g. Loomes and Sugden, 1995). Assuming this error term to be normally distributed, we have

$$m(E) \;=\; W(E) + \varepsilon_E \quad \text{with} \;\; \varepsilon_E \;\sim\; \mathcal{N}\left(0, \left(\sigma^S\right)^2\right), \tag{6}$$

where $W(E)$ is given by (3). Let $m_{\mathrm{lb}}^{\mathrm{ub}}(E) \;:=\; \{m(E)\,|\,\mathrm{lb}(E) \;\le\; m(E) \;\le\; \mathrm{ub}(E)\}$ be the interval identified by the choice sequence. The likelihood that the actual matching probability falls into the interval becomes

$$\Pr\left(m(E) \;\in\; m_{\mathrm{lb}}^{\mathrm{ub}}(E)\right) \;=\; \Pr\left(m(E) \;\le\; \mathrm{ub}(E)\right) - \Pr\left(m(E) \;\le\; \mathrm{lb}(E)\right) \tag{7}$$

We use θ to group the parameters of (3) and (6) for all events in one wave of data:

$$\theta \;:=\; [\tau_0^S, \tau_1^S, \sigma^S, \Pr_{\mathrm{subj}}(E_0), \Pr_{\mathrm{subj}}(E_1), \Pr_{\mathrm{subj}}(E_2)].$$

The likelihood of observing individual i's data in wave t becomes

$$\mathscr{L}(\theta_{i,t}) = \prod_{E \in \{E_0^S, \ldots, E_{3,C}^S\}} \Pr\Big(m(E;\theta_{i,t}) \in \big(m_{\text{lb}}^{\text{ub}}(E)\big)_{i,t}\Big), \tag{8}$$

which we estimate subject to the constraints on τ_0^S and τ_1^S given in (3) and $\Pr_{\text{subj}}(\cdot)$ being proper probabilities (including the cross-event constraints $\Pr_{\text{subj}}(E_0) > \Pr_{\text{subj}}(E_1)$ and $\Pr_{\text{subj}}(E_0) + \Pr_{\text{subj}}(E_2) \leq 1$). When maximizing the sum of the log-likelihoods over events, the objective function is not globally concave due to complex interactions of the parameters (e.g. for a poorly parameterized model the likelihood increases when σ goes to infinity). We, therefore, employ global optimization techniques. See Online Appendix C for further details.

It is easy to see that the neo-additive model, and hence α^S and ℓ^S, are identified in terms of the matching probabilities for the events in our design. $W(E_1) + W(E_2) + W(E_3) = 3\tau_0^S + \tau_1^S$ and $W(E_j) + W(E_j^C) = 2\tau_0^S + \tau_1^S, j \in \{1,2\}$ give three equations with two unknowns. The subjective probabilities drop because the events in the design contain their complements as well. The general reasoning does not depend on the functional form. In fact, Baillon, Huang, et al. (2018) and Baillon, Bleichrodt, Li, et al. (2021) propose indices that estimate α and ℓ directly with moments of matching probabilities (also see Section 2.2).[3]

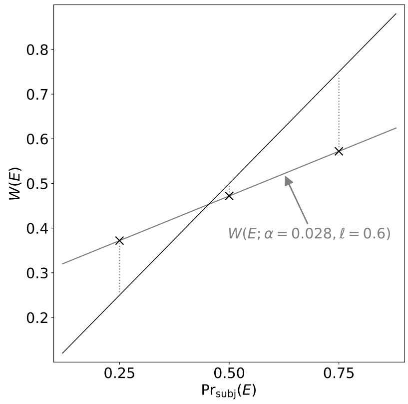
When decision errors are prevalent, however, our estimation strategy adds clarity. Our procedure enforces the theoretical restrictions on the parameters, attributing deviations from the best-fitting deterministic model to the random component of the matching probability in (6). Since there is no random component in the indices approach, researchers are left with the choice between restricting themselves to individuals with valid (α, ℓ)-pairs (e.g., Anantanasuwong et al., 2020) and keeping all observations regardless of whether the estimated parameters make sense (e.g., Dimmock, Kouwenberg, Mitchell, et al., 2015; Dimmock, Kouwenberg, and Wakker, 2016).

When panel data are available and one is willing to impose stability of parameters, it is even more helpful to explicitly account for randomness. In our approach, a large discrepancy between the parameters estimated for two waves will lead to a large variance of the random component. In an approach based on indices, the closest one can do is to average the data across waves. However, with this approach it is impossible to tell apart an individual with perfectly stable preference parame-

3. From a theoretical perspective, imposition of the neo-additive model comes with little loss of generality in our design. Baillon, Bleichrodt, Li, et al. (2021, Theorem 14 and Proposition 21) show that the indices are invariant to the choice of events only under the neo-additive model and ℓ is estimated well only if the neo-additive model is a good approximation of the source function. Using σ^S, we can quantify the quality of the approximation for each individual – while we shall think of it as measuring truly inconsistent behavior, part of it could be due to a nonlinear source function.

	α^{AEX}	ℓ^{AEX}	σ^{AEX}
Mean	0.034	0.58	0.1
Std. dev.	0.16	0.29	0.1
$q_{0.05}$	−0.22	0.084	0.001
$q_{0.25}$	−0.057	0.34	0.009
$q_{0.5}$	0.028	0.6	0.076
$q_{0.75}$	0.13	0.84	0.15
$q_{0.95}$	0.3	0.98	0.3

(a) Statistics



(b) Illustration of median parameters

Notes: Parameters are estimated separately for each of 2,407 individuals × up to 6 waves; all 11,502 estimates are used to produce the statistics in Panel a. See Table E.1 and Figure 4 for the same statistics broken down by wave. Panel b illustrates $W(E)$ at the median parameter estimates from Panel a with subjective probabilities fixed at 0.25, 0.5, and 0.75. The dotted vertical lines depict the difference between $W(E)$ and a bet on a lottery with the same entry probability of the good outcome. The gray line shows the neo-additive source function $W(E) = W\left(\Pr_{\text{subj}}(E); \alpha, \ell\right)$ evaluated at the median parameter estimates from Panel a. Sample: All waves meeting our inclusion criteria (i.e., there is variation across options and/or completion time is outside the fastest 15 %, see Section 2.3) for individuals with at least two such waves.

Figure 3. Marginal distributions of estimated parameters

ters from someone whose behavior changes erratically from one wave to the next, so long as their mean values for α and ℓ are the same.[4]

3.2 Marginal parameter distributions

Panel a of Figure 3 shows the marginal distribution of our parameters of interest, focusing for now on the AEX waves. There is substantial variation in all estimated parameters. The ambiguity parameters are spread over a large part of their support. Ambiguity aversion prevails at both the mean and at the median; we estimate ambiguity seeking behavior at the first quartile. Likelihood insensitivity is substantial with mean and median values around 0.6. The standard deviation of the distribution of the Fechner errors varies from tiny values at the fifth and twenty-fifth percentiles to 0.3 at the 95$^{\text{th}}$ percentile.

We illustrate these numbers with choice behavior in an environment similar to a task in our design, fixing subjective probabilities. A decision maker decides between a lottery yielding €x with probability p and a prospect x_E0 with $\Pr_{\text{subj}}(E) = p$. In our model, behavior is characterized by the difference $W(E) - p$, which yield the

4. Where possible, we have repeated our analyses using the indices from Baillon, Bleichrodt, Li, et al. (2021). We will discuss the results in Section 4.3 among other robustness checks and to connect directly to prior literature.

probability to choose the prospect x_E0 when plugged into the cumulative distribution function of $\mathcal{N}\left(0,\left(\sigma^{AEX}\right)^2\right)$. Figure 3b illustrates this for the median parameter estimates from Figure 3a and $\Pr_{\text{subj}}(E) = p \in \{0.25, 0.5, 0.75\}$. The decision weights $W(E) = W\left(\Pr_{\text{subj}}(E); \alpha, \ell\right)$ are shown as crosses. $W(E) - p$ is the vertical distance between the crosses and the 45°-line.[5]

For $p = \Pr_{\text{subj}} = 0.5$, likelihood insensitivity does not impact choices because $W(E) - p = -\alpha^{AEX}$. At the median value of σ^{AEX}, the probability to choose the prospect x_E0 would be 36 %, which is substantially below 50 %. Hence, the seemingly small value $\alpha^{AEX} = 0.028$ can lead to sizable deviations from subjective expected utility maximization, even at the point where likelihood insensitivity does not play a role. At the 75th percentile of σ^{AEX}, the choice probability still is 42 %. Changing α^{AEX} shifts the line $W(\Pr_{\text{subj}}(E); \alpha^{AEX}, \ell^{AEX})$; the value at the first quartile of α^{AEX} implies ambiguity seeking behavior for $p = \Pr_{\text{subj}} = 0.5$.

For the other two choices depicted in Figure 3b, the probabilities to choose x_E0 amount to 0.95 (for $p = 0.25$) and 0.01 (for $p = 0.75$). When likelihood insensitivity changes, the line for $W(\Pr_{\text{subj}}(E); \alpha^{AEX}, \ell^{AEX})$ rotates in the point $(0.5, W\left(0.5; \alpha^{AEX}, \ell^{AEX}\right))$. Increasing it thus makes both choice probabilities even more extreme; decreasing it brings $W(E) - p$ closer to the 45°-line. At the first quartile of ℓ^{AEX}, the choice probability for $p = 0.25$ ($p = 0.75$) is 0.77 (0.07) when holding the other two parameters at their median values. If ℓ^{AEX} was at its fifth percentile, the decision-maker would exhibit ambiguity aversion for $p = 0.25$ as well and choose x_E0 with probability 0.46.

This analysis has shown that there is rich heterogeneity, but the model makes sharp predictions for a wide range of estimated values of σ^{AEX}. One limitation of the analysis in this section is that the marginal distributions naturally do not capture the co-variation of the three parameters.[6] We will address this in Section 4 below, where we also place our results in the literature. To lend credibility to our approach in Section 4, however, we first establish that there is no systematic variation in individual parameters over time.

3.3 Parameter stability over time

Figure 4 depicts the same quantiles of the parameter estimates' distributions as in Figure 3a, but separately for each wave (the corresponding numbers are listed in Table E.1 along with means and standard deviations). The shapes of all three parameters' distributions look broadly similar for the AEX waves. Statistical methods

5. Tables E.3–E.5 in Online Appendix E.1 show the values of $W(E) - p$ and the corresponding choice probabilities, varying α^{AEX}, ℓ^{AEX}, and σ^{AEX} along the five quantiles shown in Panel 3a of Figure 3 (for brevity, we do not show choice probabilities for the fifth and twenty-fifth percentile of σ^{AEX} because virtually all of them are zero or one).

6. It does not make sense to consider correlations or other linear measures of co-variation in this setting because the constraints in (3) imply that $|\alpha| \leq \ell/2$, causing a highly nonlinear relationship unless α always has the same sign, which clearly is not the case.

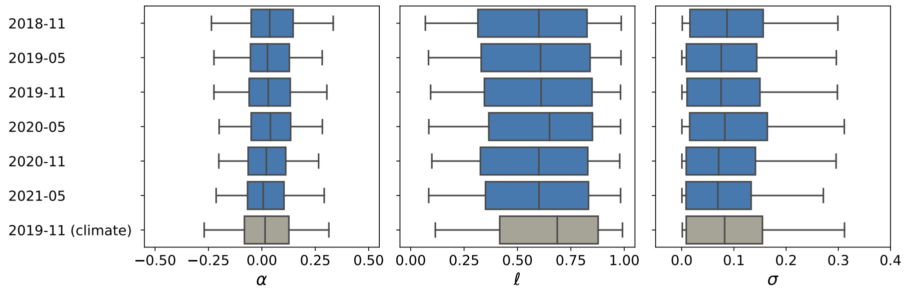


Figure 4. Marginal distributions of estimated parameters, wave by wave

Notes: This figure reports box plots for the distributions of α (left column), ℓ (middle column), and σ. Parameter estimates are obtained from the model described in Section 3.1 separately for each survey wave and individual. Parameters are reported separately for each AEX elicitation and the elicitation on climate change (last row). The boxplots depict the quartiles as well as, indicated by the whiskers, the 5 %/95 % percentiles of each distribution. Sample: All waves meeting our inclusion criteria (i.e., there is variation across options and/or completion time is outside the fastest 15 %, see Section 2.3) for individuals with at least two such waves.

reveal some differences, however.[7] Regressing each of the three parameters on wave dummies shows that on average, ambiguity aversion was largest in the first wave and decreased by about 0.025 until the last wave. This is equivalent to a change from the 54$^{\text{th}}$ percentile to the 48$^{\text{th}}$ percentile in the pooled data. There are no significant changes in average likelihood insensitivity between the early and the late waves. For the standard deviation of Fechner error, we again find a slight downward trend. The decrease is about 0.015 between the first and the last wave; equivalent to a change from moving from the 64$^{\text{th}}$ percentile to the 57$^{\text{th}}$ percentile in the pooled data. For all three parameters, there is one salient feature: In May 2020, all three parameters are significantly higher than predictions based on a linear trend. That data collection took place shortly after a huge increase in volatility of the AEX, associated with the onset of the Covid-19 pandemic. The overall pattern is consistent with a moderate amount of learning—except for likelihood insensitivity—and a transitory shock associated with the uncertainty during the initial phase of the pandemic. Economically speaking, the changes are limited.

The more interesting question is whether the three parameters are stable at the individual level, i.e., whether systematic changes alter the ranking of individuals over time. In a first pass to address this question, we regress the estimates of the last three waves on the respective parameter values of the first three waves.[8] The first column in Table 4 shows that the OLS coefficients are 0.25 for α^{AEX}, 0.36 for

7. See Table E.2 and Figure E.1 in Online Appendix E.1 for the full set of results backing the remainder of this paragraph.

8. In practice, we stack the data so that each combination of dependent and independent variables enters as one row of data. Standard errors are clustered at the individual level. Alternatively, Table E.7 in Online Appendix E.2 reports correlations between parameter estimates for all pairs of survey waves. Naturally, they are very similar to the regression coefficients on average.

ℓ^{AEX}, and 0.32 for σ^{AEX}. To interpret the magnitude of these coefficients—which can be interpreted as correlations since the variance of the parameters does not change much over time—a comparison with results on risk aversion is instructive. Chuang and Schechter (2015) review the literature on the stability of risk aversion parameters. They report that studies with at least 100 observations and at least one month between elicitations find correlations between 0.13 and 0.48. Our results fall in this range which indicates that measures of ambiguity attitudes are of comparable stability to measures of risk attitudes.

However, it is well known that estimated risk aversion parameters are subject to large measurement error (e.g. Friedman, Isaac, James, and Sunder, 2014; Frey, Pedroni, Mata, Rieskamp, and Hertwig, 2017; Schildberg-Hörisch, 2018; Gillen, Snowberg, and Yariv, 2019). There is no reason to expect this to be different for our parameters. We thus follow Gillen, Snowberg, and Yariv (2019) and run ORIV (obviously related instrumental variables) regressions. In our setting, this amounts to instrumenting one wave's parameter estimates with parameter estimates from a second wave to predict parameters in a third wave. The core assumption is that measurement error is uncorrelated across waves. We partition the data so that we predict parameters in waves 4-6 with parameters from waves 1-3.[9] Regressions are run in a stacked dataset using all permutations of selecting the endogenous regressor and the instrument from waves 1–3 and the dependent variable from waves 4–6. Standard errors are clustered on the individual level.

The last two columns in Table 4 show the results of accounting for measurement error in this way. The difference between the columns is that in Column (2), there are no additional regressors. In Column (3), we control for a large set of control variables; coefficients are reported in Online Appendix Table E.8. All F-statistics for the first stage regressions exceed 100. All coefficients of interest are between 0.95 and 0.99; none of them is statistically different from 1. The results indicate that once measurement error is accounted for, the underlying individual-level parameters do not vary systematically over time.

3.4 Parameter stability across domains

A key question arising for any parameter characterizing individual attitudes is how domain-specific it is (see, e.g. Dohmen et al., 2011, for risk attitudes). We address this question using the design with climate as the source of uncertainty, described in the last paragraph of Section 2.2. We noted that the stylized facts for the matching probabilities are broadly similar to those for the AEX at the end of Section 2.3.

We estimate our model for the climate data in the same way as we do for one wave of the AEX data. The last row in Figure 4 shows the distribution of the estimated

9. As demonstrated by Tables E.9–E.11 in Online Appendix E.2, where we split our data is immaterial for the results.

Table 4. Predicting last three waves of ambiguity parameters with first three waves

		OLS	ORIV	
		(1)	(2)	(3)
$\alpha^{AEX}_{\text{last 3 waves}}$	Intercept	0.017^{***}	-0.0097^{**}	
		(0.0025)	(0.0038)	
	$\alpha^{AEX}_{\text{first 3 waves}}$	0.25^{***}	0.95^{***}	0.98^{***}
		(0.01)	(0.07)	(0.09)
	Adj. R^2	0.07		
	1st st. F		148	101
$\ell^{AEX}_{\text{last 3 waves}}$	Intercept	0.37^{***}	0.024	
		(0.0087)	(0.022)	
	$\ell^{AEX}_{\text{first 3 waves}}$	0.36^{***}	0.97^{***}	0.95^{***}
		(0.01)	(0.04)	(0.05)
	Adj. R^2	0.14		
	1st st. F		512	293
$\sigma^{AEX}_{\text{last 3 waves}}$	Intercept	0.066^{***}	-0.0012	
		(0.0019)	(0.0054)	
	$\sigma^{AEX}_{\text{first 3 waves}}$	0.32^{***}	0.99^{***}	0.96^{***}
		(0.01)	(0.05)	(0.08)
	Adj. R^2	0.082		
	1st st. F		250	129
Controls		No	No	Yes
N Subjects		1859	1859	1452

Notes: Table shows OLS and ORIV regressions with the parameter estimates of the May 2020, November 2020, and May 2021 waves as dependent variables and the parameter estimates of the three earlier waves as potential independent variables and instruments. The table is split vertically, such that the first set of rows reports the regressions based on α^{AEX} as dependent and independent variables. The middle set of rows shows the results for ℓ^{AEX} and the last part of the table those for σ^{AEX}. Parameter estimates are obtained from the model described in Section 3.1 separately for each survey wave and individual. In line with the ORIV approach, we use a stacked data set in which all respective combinations of dependent, independent, and (for the ORIV regressions) instrumental variables enter as a separate observation. In all regressions, standard errors are clustered on the individual level and reported in parentheses. Controls are age dummies, gender, education, income and assets dummies, risk aversion, and numeracy. Full regression results reported in Online Appendix Table E.8. Sample: All waves meeting our inclusion criteria (i.e., there is variation across options and/or completion time is outside the fastest 15 %, see Section 2.3) for individuals with at least two such waves in 2018/2019 and at least one such wave in 2020/2021 (This is required for ORIV regressions and we impose the same restriction for the OLS regression). * $- p < 0.1$, ** $- p < 0.05$, *** $- p < 0.01$.

parameters. For α and σ, the distributions are visually similar, although ambiguity aversion is lower in the climate data than its average across the AEX waves (see Table E.6). Likelihood insensitivity regarding temperature changes is notably greater than for the AEX data; the average difference amounts to 0.05.

Parameter stability at the individual level is the more interesting question once more. Table 5 shows regressions for each parameter in the climate domain on parameters from the financial domain elicited in the same wave. The first column of each parameter shows OLS regression with slope coefficients of 0.69, 0.35, and 0.51 for α, ℓ, and σ respectively. This suggests sizable stability across domains, particularly for ambiguity aversion.

Again, there is reason to believe the OLS estimates may be biased. Classical measurement continues to be the same concern as before. However, one may also suspect spurious *positive* correlation because the two elicitations were separated only by a short introduction to the climate change questions. To address this issue, we run two-stage least squares regressions, instrumenting the endogenous regressor from the November 2019 wave with the same parameter from other waves. As in the case of temporal stability, the bias is eliminated if estimation errors are uncorrelated across waves. As in Table 4, the second column of Table 5 reports on a specification without controls and the third column on a specification controlling for many covariates (the full list of coefficients can be found in Online Appendix Table E.12). The coefficients of interest are very similar in both specifications.

The coefficient for ambiguity aversion is precisely estimated and statistically indistinguishable from 1. This supports the interpretation of ambiguity aversion as a stable preference that extends across domains. Anantanasuwong et al. (2020) elicit ambiguity attitudes in a sample of households holding risky assets for events from different financial domains: individual stocks, local and foreign stock indices, and crypto funds. They find that ambiguity aversion parameters are very related across these domains with a correlation coefficient around 0.7, which is very close to what we find in the OLS regression. More closely related to our 2SLS regression, Anantanasuwong et al. (2020) conduct a factor analysis and conclude that ambiguity aversion can be described by one underlying trait. Our results indicate that the stability of ambiguity aversion holds not just within financial contexts, but more generally.

We further find that ℓ also has a substantial transferable component, but the slope coefficient of 0.60 (0.63 when controls are added) is well below 1. Based on the multiple prior interpretation of ℓ as the perceived level of ambiguity, this is expected as perceptions are more likely to differ across domains than preferences. Anantanasuwong et al. (2020) also find weaker dependence across domains for ℓ with correlation coefficients ranging of 0.16 or 0.45, depending on whether they keep subjects with set-monotonicity violations in the sample (their results for α were unaffected by this choice). Correcting for measurement error, we find a substantially

Table 5. Predicting climate ambiguity parameters with AEX parameters

		OLS	2SLS	
		(1)	(2)	(3)
$\alpha^{climate}_{2019-11}$	Intercept	−0.003	−0.016***	
		(0.0033)	(0.0039)	
	$\alpha^{AEX}_{2019-11}$	0.69***	1.04***	1.06***
		(0.03)	(0.05)	(0.07)
	Adj. R^2	0.39		
	1st st. F		215	148
$\ell^{climate}_{2019-11}$	Intercept	0.43***	0.28***	
		(0.015)	(0.024)	
	$\ell^{AEX}_{2019-11}$	0.35***	0.60***	0.63***
		(0.02)	(0.04)	(0.05)
	Adj. R^2	0.13		
	1st st. F		735	406
$\sigma^{climate}_{2019-11}$	Intercept	0.053***	0.022***	
		(0.0027)	(0.005)	
	$\sigma^{AEX}_{2019-11}$	0.51***	0.83***	0.88***
		(0.03)	(0.05)	(0.07)
	Adj. R^2	0.23		
	1st st. F		92	51
Controls		No	No	Yes
N Subjects		1843	1843	1411

Notes: This table shows OLS and 2SLS regressions with the parameter estimates for the decisions about changes in climate (elicited in November 2019) as the dependent variable and the parameter estimates for the decisions about the AEX elicited in November 2019 as independent variable. For the 2SLS regressions, the parameters of all other AEX waves are used as instruments. Parameter estimates are obtained from the model described in Section 3.1 separately for each survey wave and individual. For 2SLS, we use a stacked data set in which all instrumental variables enter as a separate observation and we cluster standard errors on the individual level. Controls are age dummies, gender, education, income and assets dummies, risk aversion, numeracy, and indicators of self-assessed understanding and perceived threat of climate change. The latter two vary between 0 and 1. Full regression results reported in Online Appendix Table E.12. Sample: All waves meeting our inclusion criteria (i.e., there is variation across options and/or completion time is outside the fastest 15 %, see Section 2.3) for individuals with at least two such waves.
* $- p < 0.1$, ** $- p < 0.05$, *** $- p < 0.01$.

higher common component.[10] Turning to the third panel in Table 5, the stability of the standard deviation of the Fechner error is around 0.85 and, thus, in between the values for the other two parameters.

As with stability over time, the comparison with risk aversion is instructive. Dohmen et al. (2011) examine self-reported assessments of risk aversion in several domains like financial matters, sports, or health and report correlations that correspond to R^2 between 0.16 to 0.36 which is comparable to what we find in the OLS columns of Table 5. Dohmen et al. (2011) reason that differences in risky behavior across domains might be more likely to reflect different risk perceptions, rather than differences in actual preferences. This fits well with our results: Ambiguity aversion is very stable, but the perception of ambiguity varies across contexts to a certain degree. One interpretation of our findings is that there can be room for external stimuli—such as providing individuals with more information about a source of uncertainty—to change ℓ while this might not affect α much, unless it is on a constraint implied by ℓ. This aligns well with the findings by Baillon, Bleichrodt, Keskin, et al. (2018) who conduct such an information experiment.

4 Ambiguity types and financial behavior

The previous section has established that each of our three parameters of interest is very heterogeneous across individuals, but remarkably stable over time. The first finding, however, is of limited importance for describing decision behavior and heterogeneity therein. This is due to the non-separable nature of the choice model. The argument might be clearest for the relation between ambiguity aversion α and likelihood insensitivity (or the perceived level of ambiguity) ℓ. For example, individuals who fully trust their probability judgments (who do not perceive any ambiguity) necessarily have an ambiguity aversion parameter of zero. In general, the constraints in (3) imply that $|\alpha| \leq \ell/2$, so ambiguity aversion is bounded by the degree of likelihood insensitivity (the perceived level of ambiguity). In a similar vein, the two preference parameters hardly matter if σ takes on very high values.

In the first part of this section, we thus classify individuals into a discrete set of types, which are characterized by our three parameters of interest. The procedure does not place any restrictions on the dependence between α, ℓ, and σ. This is one of the reasons discrete types are very widely used in nonlinear economic models

10. One potential reason our results on the perceived level of ambiguity are at variance with the results of Anantanasuwong et al. (2020) for their full sample is that they use the indices proposed by Baillon, Bleichrodt, Li, et al. (2021) directly. Table 5 demonstrates that our model-based estimates are likely to be subject to sizable measurement error. In our robustness checks, we show that measurement error affects ambiguity attitudes estimated with BBLW-indices in an even stronger fashion. When replicating Table 5 with index-based estimates, we get an OLS coefficient for ℓ of 0.14, almost the same as that Anantanasuwong et al. (2020, see Table H.4). Unsurprisingly, the 2SLS-measurement-error-adjusted regression slope for the BBLW-indices is in the range of what we find with our model.

(e.g., Keane and Wolpin, 1997). We establish that four types capture a large degree of the observed heterogeneity. In Section 4.2, we show these types are related to socio-demographic characteristics and whether they help predict real-world financial behavior. In Section 4.3, we compare our results to alternative specifications and to the previous literature.

4.1 Describing heterogeneity in attitudes and error propensities

In a first step, we re-estimate (8), imposing that $\tau_{0,i}^{AEX}$, $\tau_{1,i}^{AEX}$, and σ_i^{AEX} do not vary across waves. Hence, there is no subscript t to the parameters anymore. Doing so changes the interpretation of σ_i^{AEX} because, in addition to the previous types of inconsistencies, it will also capture behavior that is erratic only across waves. Estimates of σ_i^{AEX} will thus be substantially larger than our previously-reported estimates of $\sigma_{i,t}^{AEX}$. We then apply the k-means algorithm (e.g., Bonhomme and Manresa, 2015; see Gaudecker and Wogrolly, 2022, for a related application) to classify individuals into a discrete set of groups. The algorithm assigns individual observations $x_i := [\alpha_i^{AEX}, \ell_i^{AEX}, \sigma_i^{AEX}]$ to groups g such that $\sum_i ||x_i - c_{g(i)}||^2$ is minimized for the group means $c_g = \frac{1}{N_g}\sum_{i \in g} x_i$. We follow common practice and scale each component of x_i to mean 0 and standard deviation 1 in the cross-section to ensure all of them are given equal weight in the optimization. The problem is NP-hard, but several heuristic algorithms exist that work well in practice. The method is widely used in machine learning; we use the implementation in the Python library *scikit-learn* (Pedregosa, Varoquaux, Gramfort, Michel, Thirion, et al., 2011).

In the paper, we report results for $k = 4$ types, striking a balance between qualifying as a "summary" and not merging types that display economically meaningful differences in choice behavior. We provide empirical details and a hint at results for alternative choices of k at the very end of this Section 4.1. Figure 5 shows the distribution of ambiguity profiles in the $(\alpha,\ \ell)$-space with large diamonds indicating group means and small dots indicating individual profiles. We do not visualize the standard deviation of errors σ, but list it in the legend along with the share of each type.

At 30%, the largest share of all subjects is estimated to have an ambiguity aversion parameter $\alpha^{AEX} = -0.0002$, likelihood insensitivity $\ell^{AEX} = 0.28$, and a standard deviation of the Fechner errors $\sigma^{AEX} = 0.14$. For all three parameters, the distance to zero is closest in this group, although the error variance is very similar for three out of the four types. Since subjective expected utility maximizers who do not make any errors would have a zero for each parameter, we label it the "near SEU" type. For the example decisions we used in the previous section—binary choices between a lottery yielding €x with probability p and a prospect x_E0 with

$\Pr_{\text{subj}}(E) = p \in \{0.25, 0.5, 0.75\}$—we obtain choice probabilities for the AEX of 0.7, 0.5, and 0.31.[11]

We label the second-largest group, comprising 27 % of the sample, the "Ambiguity averse". This group is estimated to have an ambiguity aversion parameter $\alpha^{AEX} = 0.15$, likelihood insensitivity $\ell^{AEX} = 0.71$, and a standard deviation of the Fechner errors $\sigma^{AEX} = 0.14$. For our example choices, this group has a slight preference for the ambiguous option if $\Pr_{\text{subj}}(E) = p = 0.25$, choosing the ambiguous prospect with 58 % probability. For probabilities $p = 0.5$ ($p = 0.75$), these choice probabilities are 15 % (1.2 %).

A third group is associated with a likelihood insensitivity parameter $\ell^{AEX} = 0.64$, slightly below the value of the ambiguity averse. The standard deviation of the Fechner errors is also very similar to the previous two groups. The defining feature of this group is $\alpha^{AEX} = -0.054$, implying ambiguity seeking behavior on average. This is how we label them, too. For the example decisions, the choice probability for the ambiguous prospect would be 93 % (64 %, 24 %) at $p = 0.25$ ($p = 0.5$, $p = 0.75$).

For all three groups discussed so far, the error variances are estimated to be very close to each other. So it is no surprise that they partition the (α, ℓ)-space in Figure 5 almost perfectly. This is very different for the last group, members of which are scattered almost all over the triangle with valid ambiguity parameters in Figure 5. Twenty percent of individuals are classified to be in this group; what stands out among the parameters is the large standard deviation of the errors with $\sigma^{AEX} = 0.29$. We thus label it the "High noise" type.

This group is special in a few respects. First, the choice probabilities for the three example probabilities move least in this group. This is not due to the source function being particularly close to the 45°-line, but because the random component in (6) is much more important than in the other groups. Viewed from a different angle, no matter what $\Pr_{\text{subj}}(E)$ is, almost any matching probability (systematic plus random component) would occur with some probability substantially larger than zero. Second, we find the largest fraction of set-monotonicity errors in this group (at 25 % of superset-subset pairs, about twice as often as for the other groups). Third, when we go back to the wave-by-wave estimates from Section 3.2, we find them to be most volatile among the high noise types (see Online Appendix Table F.3). This implies that the large error parameters are due both to erratic behavior within and across waves.

With these types at hand, we are now in a position to describe in detail why we picked $k = 4$, referring to results for $k \in \{3, 5, 8\}$. Tables and figures are relegated to the Online Appendix, Sections F.2–F.4. Reducing k to 3 distributes the group we classified as ambiguity seeking across the other three groups. Most individuals go into the near-SEU group, which comprises almost 40 % of the sample. It covers a very

11. See Table F.1 in Online Appendix F.1; Figure F.1 visualizes the source function including the uncertainty introduced by the Fechner errors.

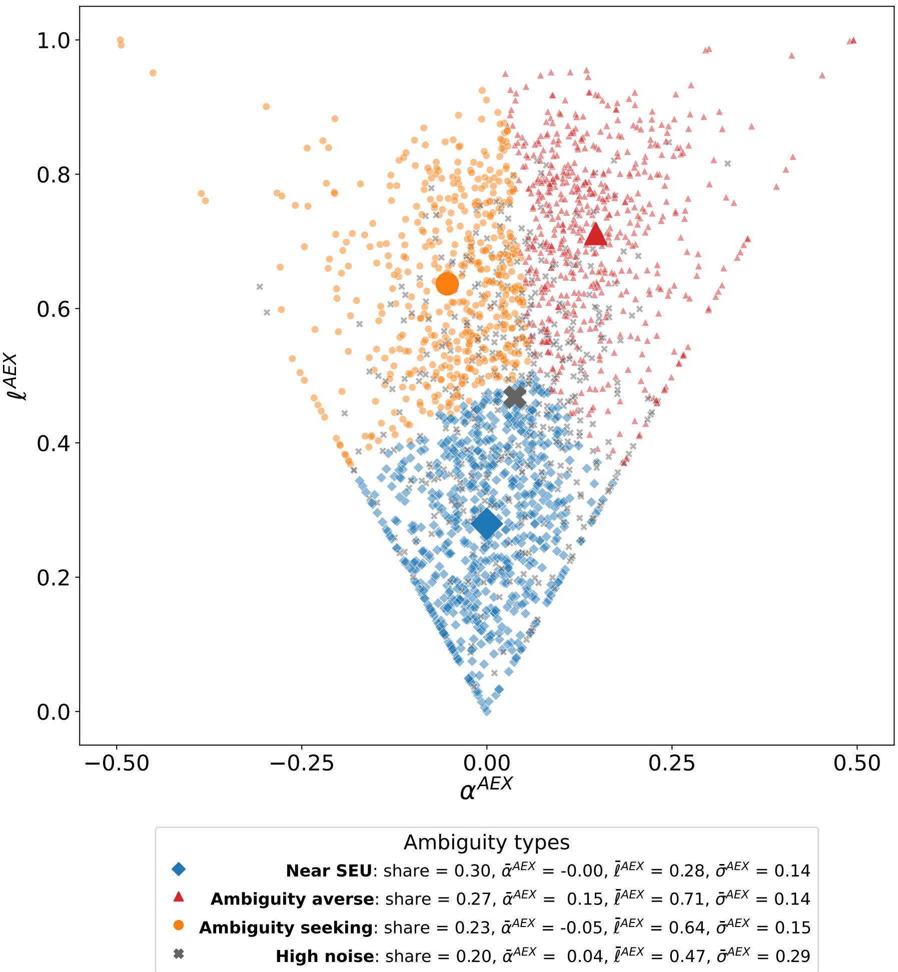


Figure 5. Summarizing heterogeneity in ambiguity profiles with $k = 4$ discrete groups

Notes: The small symbols depict individual preference parameter estimates $(\alpha_i^{AEX}, \ell_i^{AEX})$ obtained from estimating (8) under the assumption that these two parameters and σ_i^{AEX} do not vary across waves. The large symbols are group centers resulting from clustering individuals with the k-means algorithm on the three parameters into four groups. Sample: All waves meeting our inclusion criteria (i.e., there is variation across options and/or completion time is outside the fastest 15 %, see Section 2.3) for individuals with at least two such waves.

	Mean	Std. dev.	$q_{0.05}$	$q_{0.25}$	$q_{0.5}$	$q_{0.75}$	$q_{0.95}$
α^{AEX}	0.035	0.11	−0.13	−0.031	0.032	0.1	0.22
ℓ^{AEX}	0.52	0.22	0.15	0.35	0.53	0.69	0.85
σ^{AEX}	0.17	0.079	0.066	0.12	0.16	0.22	0.33

wide range of behavior – both individuals whose behavior is indistinguishable from SEU-maximization and the subjects at the top left tip of the triangle in Figure 5, i.e., behavior that is most distant from SEU-maximization while consistent, are put in this group. This is not a grouping that makes much sense from a behavioral perspective.

Increasing k to 5 leaves the near SEU and the ambiguity seeking types unchanged. The ambiguity averse and high noise types are split up. The parameters of the original types become slightly more extreme, the parameters of the type in between are all weighted averages of the original types' parameters. Decision behavior is fairly close to the near SEU-type with $k = 4$, but somewhat more erratic. Even when doubling k to 8, there are no groups with clearly different choice behavior from the four types considered in this main text. The four original groups do move somewhat more toward the respective extremes. E.g., in our example decisions, the ambiguity seeking type has choice probabilities for the ambiguous prospect of 94 % / 76 % / 45 % instead of 93 % / 64 % / 24 %. The original labels based on $k = 4$ continue to work for the extreme types and the four additional types are convex combinations thereof.

We conclude that the four types describe overall heterogeneity in choice behavior well, keeping in mind that each group mean summarizes a large volume in $(\alpha^{AEX}, \ell^{AEX}, \sigma^{AEX})$-space. Hence, actual heterogeneity in choice behavior goes well beyond the four types, as is visually clear from Figure 5. Different applications may want to work with much larger k. However, our goal is to have a low-dimensional summary of heterogeneity and $k = 4$ is best suited for this purpose. We now ask how these groups are related to observable characteristics and whether they help explain portfolio choice behavior.

4.2 Ambiguity types: Predictors and consequences

Table 6 describes the groups and their characteristics. There is one column per group. The first two panels repeat the shares and preference parameter estimates from the legend of Figure 5, adding the (very small) standard errors. The lower panel contains average characteristics of groups, including standard errors of these means. We describe the groups without explicitly mentioning the statistical significance of differences, focusing on comparisons where this clearly is the case. As an alternative, we predict group membership in a multinomial regression to partial out the effects of other covariates. Results generally line up, so we relegate the marginal effects to Online Appendix Table F.2.

Near SEU subjects have the highest prevalence of advanced formal education; more than half of them have obtained a tertiary degree and only 13 % are found in the lowest education category. They are among the youngest and somewhat more likely to be male. Monthly income and total financial assets are the highest among all groups, whereas the risk aversion index is the lowest. The numeracy index is 0.63 on average, which is much higher than in any other group and corresponds to

Table 6. Average characteristics of group members

	Ambiguity types			
	Near SEU	Ambiguity averse	Ambiguity seeking	High noise
Share	0.3	0.27	0.23	0.2
α^{AEX}	−0.0002	0.15	−0.054	0.038
	(0.0024)	(0.0031)	(0.0038)	(0.0043)
ℓ^{AEX}	0.28	0.71	0.64	0.47
	(0.0045)	(0.0054)	(0.0056)	(0.0079)
σ^{AEX}	0.14	0.14	0.15	0.29
	(0.0018)	(0.0023)	(0.0024)	(0.0025)
Education: Lower secondary and below	0.13	0.29	0.26	0.42
	(0.013)	(0.019)	(0.02)	(0.024)
Education: Upper secondary	0.31	0.38	0.36	0.29
	(0.018)	(0.02)	(0.022)	(0.022)
Education: Tertiary	0.56	0.33	0.38	0.28
	(0.019)	(0.019)	(0.022)	(0.022)
Age	54	55	57	65
	(0.64)	(0.65)	(0.69)	(0.65)
Female	0.4	0.61	0.52	0.47
	(0.019)	(0.02)	(0.022)	(0.024)
Monthly hh net income (equiv., thousands)	2.5	2.1	2.2	2
	(0.04)	(0.039)	(0.05)	(0.042)
Total hh financial assets (equiv., thousands)	55	23	39	34
	(6.9)	(2.6)	(5.9)	(4.4)
Risk aversion index	−0.1	0.093	0.017	0.0098
	(0.035)	(0.041)	(0.048)	(0.053)
Numeracy index	0.63	−0.2	0.049	−0.72
	(0.024)	(0.038)	(0.042)	(0.056)

Notes: The first row shows the share of individuals classified into a given group. For each group, the mean of several variables are shown. Income and financial assets are in thousands and equivalized for couples. We consider the income of both partners. Total assets include assets kept in joint accounts and those assigned to the respondent (i.e., the person identifying as being most familiar with the household's finances). Risk aversion and numeracy are normalized to have mean zero and unit variance. Sample: All waves meeting our inclusion criteria (i.e., there is variation across options and/or completion time is outside the fastest 15 %, see Section 2.3) for individuals with at least two such waves.

the second tercile in the entire sample. Many of these characteristics point toward this group being the most sophisticated one in statistical and financial matters. This is consistent with subjected expected utility maximization being a benchmark of rationality, from which near SEU subjects fall short the least.

The ambiguity averse and the ambiguity seeking groups are similar in their educational attainment, assuming a position in between the extremes. The average age is 55-57 years and similar to that of the near SEU type. Among all groups, the ambiguity averse group has the highest share of women, which is just about average for the ambiguity seeking type. Both groups find themselves in between the near SEU and high noise types for income, although the difference between the ambiguity averse and high noise groups is not significantly different from zero. Total financial asset holdings are the lowest among the ambiguity averse. In terms of risk aversion, the two groups are indistinguishable in statistical terms. If we control for other characteristics in the multinomial logit model, risk aversion is, however, a

Table 7. Ambiguity attitudes and portfolio choice: Marginal effects

	Owns risky assets (Probit)		Share risky assets (Tobit)	
	(1)	(2)	(3)	(4)
Ambiguity averse type	-0.23^{***}	-0.084^{***}	-0.44^{***}	-0.17^{***}
	(0.024)	(0.023)	(0.059)	(0.055)
Ambiguity seeking type	-0.1^{***}	-0.018	-0.15^{***}	-0.028
	(0.028)	(0.024)	(0.05)	(0.046)
High noise type	-0.18^{***}	-0.053^{*}	-0.24^{***}	-0.082
	(0.027)	(0.028)	(0.059)	(0.059)
Controls	No	Yes	No	Yes
Observations	1727	1624	1584	1502
Pseudo R^2	0.054	0.3	0.042	0.28
p-values for differences between				
Ambiguity averse, Ambiguity seeking	0	0.0081	0	0.012
Ambiguity averse, High noise	0.034	0.25	0.0041	0.17
Ambiguity seeking, High noise	0.0079	0.21	0.19	0.36

Notes: The first two columns display Probit regressions where the dependent variable is a dummy indicating whether the subject holds any risky financial assets. In the last two columns, we run Tobit regressions with the share of risky financial assets of all financial assets as the dependent variable. The table reports average marginal effects of a change from the left-out type (near SEU) to the respective type. Controls in columns (2) and (4) are age groups, gender, education, income and assets groups, risk aversion, and numeracy. Full regression results reported in Online Appendix Table F.5. Sample: All waves meeting our inclusion criteria (i.e., there is variation across options and/or completion time is outside the fastest 15 %, see Section 2.3) for individuals with at least two such waves. * $- p < 0.1$, ** $- p < 0.05$, *** $- p < 0.01$

significant predictor of the ambiguity averse group. The numeracy index is lower among the ambiguity averse than the ambiguity seeking.

Finally, subjects classified to be of the high noise type are the least educated and oldest on average. The female share is similar to the overall mean. Income is among the lowest, financial assets are in between those of the other groups. The numeracy index is -0.72 on average, which corresponds to the 22$^{\text{nd}}$ percentile in the overall sample. Remember that a high σ may come about through erratic behavior or because the neo-additive function is a bad approximation. The structure of the covariates lends support to the former interpretation in that high noise subjects score lower on dimensions that predict behavior in cognitively demanding tasks.

Next, we show that our estimated preference types help predict financial decisions. Table 7 contains the results of regressing risky asset holdings on the ambiguity types (Columns 1 and 3) and additionally on control variables, including other potential determinants of financial decisions like risk aversion and numeracy (Columns 2 and 4). In the first two columns, the dependent variable is risky asset ownership and we use a Probit model. The last two columns employ a Tobit model to explain the share of risky assets.

Near SEU-type individuals have the highest propensity to own risky assets; they invest the largest share of their wealth into these. In both dimensions, they are followed by individuals classified to be ambiguity seeking and then by the high-noise types. The ambiguity averse have the lowest propensity to own risky assets and the smallest share invested in them. Differences between groups are significant in the unconditional specifications, the exception being that we cannot statistically tell apart shares invested in risky assets of the ambiguity averse and high noise types in column (3). Once we control for a large number of covariates in columns (2) and (4), coefficients drop everywhere while preserving the ranking of point estimates. Many gaps remain large in economic terms. For example, we estimate an 8 percentage point difference in risky asset participation between the near SEU and ambiguity averse types. Differences between the ambiguity averse on the one hand and near SEU or ambiguity seeking types on the other hand always remain significant. This is not true for most other comparisons.

Our results on portfolio choice behavior are robust to using an alternative measure of risky assets. We obtain this measure by merging our survey data with administrative records at the individual level (see Zimpelmann, 2021, for an extensive comparison of the measures) due to well-known measurement issues with survey reports of household financial assets. The results, shown in Online Appendix Table F.7 look very similar to those reported in Table 7.[12] In particular, the same conclusions hold for unconditional and conditional differences between the ambiguity averse on the one hand, and near SEU or ambiguity seeking types on the other hand. One difference is that the high noise type looks closer to the ambiguity seeking type when using the administrative assets data. One reason could be that erratic response behavior in our survey is correlated with underreporting of assets.

In summary, our results show that ambiguity preferences obtained from small-scale controlled choices help explain an important dimension of real-world financial behavior. Importantly, such strong predictive power of our preference parameter estimates should not be taken for granted. For the case of risk aversion, Charness, Garcia, Offerman, and Villeval (2020) show that measures based on designs comparable to ours often fail to explain anything outside of controlled environments.

4.3 Alternative specifications and relation to the literature

Our results are remarkably robust to various decisions we have made in our main analysis.[13] Including all data instead of requiring two waves meeting minimal quality standards increases the number of individuals by 10 %, but does not lead to any substantive changes in the parameter distributions or the clustering outcomes. The coefficients for portfolio choice behavior attenuate slightly toward zero, but

12. We ran the regressions using the administrative assets data in a remote computing environment at Statistics Netherlands, which is why Table F.7 reports OLS regression results. Comparing Table 7 with OLS regressions using the survey data in Table F.6 shows that this should not affect our conclusions.

13. For the three alternative specifications that we describe in the following, we provide longer descriptions and repeat all relevant tables and figures in Online Appendices G.1–G.3.

all comparisons we have highlighted in the previous section remain significant. The opposite strategy of requiring a balanced panel—i.e., six waves of reasonable data—leads to a drop in the number of individuals by more than 40%. Most statistics remain very close to the values we reported in the main text. One exception is that the average values for ambiguity aversion drop somewhat. In the clustering approach, this is reflected in a lower value of ambiguity aversion for the ambiguity averse type only ($\alpha^{AEX} = 0.12$ instead of 0.15). The long time series per individual lead to more sharply identified differences in types' portfolio choice behavior – most point estimates remain similar, but p-values for the comparisons between groups drop even further.

Another specification choice that is interesting from a modeling perspective concerns the restrictions of the parameters. While the multiple-prior interpretation of our parameters requires $0 \leq \tau_0^S \leq 1 - \tau_1^S$ in (3), an alternative is to take a more descriptive approach, which allows matching probabilities to be hypersensitive to subjective probabilities. Graphically, this means that in the analog to Figure 5, points can now fall below the triangle with valid parameters. Throughout all analysis, the only noticeable change is a drop in the estimated value of ℓ by about 0.02. In the clustering approach, the types have the same average characteristics as before and for 97.5 % of the sample, the assigned groups are identical. This is reflected in the absence of meaningful differences in the group compositions or portfolio choice regressions.

To connect directly with prior literature, we re-run most of our analyses using the indices developed in Baillon, Bleichrodt, Li, et al. (2021). We discussed some a priori considerations in Section 3.1; Online Appendix H has all the tables and figures we refer to in what follows and Section I contains a more detailed comparison with the literature. Closest to our data are other studies estimating ambiguity attitudes in broad population samples (Dimmock, Kouwenberg, Mitchell, et al., 2015; Dimmock, Kouwenberg, and Wakker, 2016; Anantanasuwong et al., 2020). The first two studies use urns as the source of uncertainty; the last considers four different financial assets, among them the development of the AEX. An important difference is that ours is the only data with a panel dimension. The most direct comparison is thus for the wave-by-wave estimates from Section 3.

Using an index-based approach leaves the wave-by-wave estimates of α^{AEX} mostly unaffected. The median rises from 0.028 to 0.033, the change in the mean is similar, and the distribution is spread out slightly more with a standard deviation of 0.18 instead of 0.16. These values are very much in line with the three studies mentioned in the previous paragraph. As prior literature we also regress the ambiguity aversion parameter on potential determinants. The most interesting relation concerns the relation between risk aversion and ambiguity attitudes. The mixed results of previous papers (Dimmock, Kouwenberg, and Wakker, 2016, and Delavande, Ganguli, and Mengel, 2019 find a negative relation; Dimmock, Kouwenberg, Mitchell, et al., 2015, and Anantanasuwong et al., 2020, a positive one) find their reflection in a zero conditional correlation in our data. In contrast, we found risk aversion to be a strong predictor of the ambiguity types in the previous subsection. In terms of ambiguity aversion the implied relationship is nonlinear: The near-SEU types (α^{AEX} near zero) are clearly less risk averse on average than all other types, whose average α is larger (ambiguity averse and high noise types) or smaller (the ambiguity seeking). This result underscores the importance of considering the multidimensional nature of heterogeneity explicitly.

Along several dimensions, likelihood insensitivity is much more volatile than ambiguity aversion. It is more sensitive to the estimation approach we apply in our data and varies more across different studies – this applies to the source of uncertainty, the co-variation with socio-demographic characteristics, and the relation with portfolio choice.

When moving from our wave-by-wave estimates in Section 3 to an index-based approach, ℓ^{AEX} rises substantially. For example, the median increases from 0.6 to 0.88. This rise is a consequence of the fact that set-monotonicity errors are reflected in a more important random component when estimating (6) whereas they lead to $\ell^{AEX} > 1$ under the indices approach. When partitioning the sample into valid and invalid values of the indices, the mean of σ^{AEX} is 0.07 in the former and 0.16 in the latter. The stochastic component picks up other types of imprecisions as well – in the subsample with valid values of $(\alpha^{AEX}, \ell^{AEX})$, the index-based median estimate of ℓ^{AEX} is 0.8.

The values we estimate using indices are larger than urn-based estimates (both Dimmock, Kouwenberg, and Wakker (2016) and Dimmock, Kouwenberg, Mitchell, et al. (2015) find average values of ℓ^{urn} close to 0.4) and slightly below others for the stock market (Anantanasuwong et al., 2020, estimate the median of ℓ^{AEX} to be 1 when including all observations and 0.89 when conditioning on valid indices).

Looking at the correlates of marginal parameter estimates, ℓ falls in both education and numeracy, which is in line with Dimmock, Kouwenberg, and Wakker (2016) and Anantanasuwong et al. (2020) while Dimmock, Kouwenberg, Mitchell, et al. (2015) find a positive relation. While this holds regardless of whether we use our model or the indices-based approach, the latter masks some interesting patterns. For example, the large positive correlation between ℓ^{AEX} and the oldest age group in the indices-based approach seems to be driven in equal parts by likelihood insensitivity and imprecisions. Furthermore, based on our model estimates, women have a larger ℓ^{AEX}, but a smaller stochastic component. Those relations cancel out for the indices-based approach where likelihood insensitivity is unrelated to gender.

While we are not aware of any studies estimating deviations from a benchmark model in the context of choice under ambiguity, several papers estimate parameters related to the standard deviation of σ^{AEX} in the context of choice under risk. The results line up well with ours. Gaudecker, Soest, and Wengström (2011) find higher age, lower wealth, and lower education levels to be associated with a large influence of the random component of utility. In Choi, Kariv, Müller, and Silverman (2014) high age, low education, low income, and low wealth predict deviations from utility maximizing behavior. Echenique, Imai, and Saito (2021) find younger and cognitively able subjects to come closer to expected utility maximization.

Our larger sample size helps add precision to suggestive prior findings on a negative relation of both α and ℓ on the one hand, and portfolio risk on the other hand. Dimmock, Kouwenberg, and Wakker (2016) find some evidence that both parameters predict low stock market participation rates, but statistical significance depends on the precise specification. Similar statements hold for Anantanasuwong et al. (2020) when it comes to predicting risky investment shares in a sample of investors. In our data, the corresponding regressions show clearly negative coefficients for the indices-based approaches, both for ownership of and for shares invested in risky assets. These findings line up well with our prior analysis based on types.

5 Discussion

We have analyzed a large panel dataset containing incentivized choices between lotteries with known probabilities on the one hand and events relating to the stock market or climate change on the other hand. While the vast majority of economic research has dealt with such real-world events in an expected utility framework, our results have demonstrated that nearly all subjects perceive some degree of ambiguity with respect to these events. Even though there is a large common component, the extent of the perceived ambiguity typically differs across the two domains of financial markets and temperature changes. At the same time, the attitude toward ambiguity is remarkably stable across these two sources of uncertainty.

We have argued that it is useful to explicitly estimate a stochastic choice model because random behavior would otherwise be subsumed in the parameters supposedly characterizing ambiguity attitudes. While there is a long tradition of such models in other strands of the literature, to the best of our knowledge we have provided the first application in the context of ambiguity attitudes. Structural estimates at the individual $\times$ wave level have yielded a triplet of ambiguity aversion, likelihood insensitivity (or the perceived level of ambiguity), and the propensity to choose at random as opposed to the best-fitting model.

The properties of these parameters are comparable to parameters relating to risk preferences, which have received much more attention in the literature. In particular, all parameters are highly heterogeneous in the population. At the same time, they are fairly stable over time, with similar properties for risk preferences and ambiguity attitudes. Our IV approach has shown the absence of any systematic changes.

Our core analysis has thus focused on estimating the parameters at the individual level by imposing their stability over time. This means that the random choice component will also pick up variation across waves in addition to within-wave behavior that cannot be explained by the best-fitting deterministic part of the model.

We have argued that the most promising way to describe the three-dimensional distribution of parameters—which are inherently non-separable in our choice model—using clustering techniques recently popularized in the econometric literature.

We found that four ambiguity types are a good way to balance parsimony and capture all economically interesting choice patterns. Predictions for choices differ sharply across these groups. The way the groups differ in both a large set of observed characteristics and portfolio choice behavior makes intuitive sense.

Our results suggest that ambiguity attitudes should be treated on par with risk preferences when it comes to their measurement and their importance in explaining behavior. For example, our results demonstrate much higher explanatory power for portfolio choices than similar studies for risk preferences (see the sobering survey in Charness et al., 2020). We view our applications to portfolio choice as highly suggestive. However, more careful modeling is needed in that respect as well as extending the domains – other relevant areas where ambiguity may play an important role are the labor market, lifestyle decisions in relation to climate change, individual health, or housing choices.

References

Abdellaoui, Mohammed, Aurélien Baillon, Laetitia Placido, and Peter P. Wakker. 2011. "The Rich Domain of Uncertainty: Source Functions and Their Experimental Implementation." *American Economic Review* 101 (2): 695–723. [6]

Alon, Shiri, and Gabi Gayer. 2016. "Utilitarian Preferences With Multiple Priors." *Econometrica* 84 (3): 1181–201. [6]

Anantanasuwong, Kanin, Roy Kouwenberg, Olivia S Mitchell, and Kim Peijnenburg. 2020. "Ambiguity Attitudes for Real-World Sources: Field Evidence from a Large Sample of Investors," 72. [4, 5, 10–12, 16, 22, 24, 32–34]

Apesteguia, Jose, and Miguel A Ballester. 2021. "Separating Predicted Randomness from Residual Behavior." *Journal of the European Economic Association* 19 (2): 1041–76. [5]

Baillon, Aurelien, Yoram Halevy, and Chen Li. 2022a. "Experimental Elicitation of Ambiguity Attitude Using the Random Incentive System." *Experimental Economics* 25: 1002–23. [9]

Baillon, Aurelien, Yoram Halevy, and Chen Li. 2022b. "Randomize at Your Own Risk: On the Observability of Ambiguity Aversion." *Econometrica* 90 (3): 1085–107. [4, 9]

Baillon, Aurélien, Han Bleichrodt, Umut Keskin, Olivier l'Haridon, and Chen Li. 2018. "The Effect of Learning on Ambiguity Attitudes." *Management Science* 64 (5): 2181–98. [4, 11, 24]

Baillon, Aurélien, Han Bleichrodt, Chen Li, and Peter P Wakker. 2021. "Belief Hedges: Applying Ambiguity Measurements to All Events and All Ambiguity Models." *Journal of Economic Theory* 198: 105353. [4, 6, 7, 16, 17, 24, 32]

Baillon, Aurélien, Zhenxing Huang, Asli Selim, and Peter P. Wakker. 2018. "Measuring Ambiguity Attitudes for All (Natural) Events." *Econometrica* 86 (5): 1839–58. [4–8, 11, 16]

Bardsley, Nicholas. 2000. "Control without Deception: Individual Behaviour in Free-Riding Experiments Revisited." *Experimental Economics* 3 (3): 215–40. [8]

Bonhomme, Stéphane, and Elena Manresa. 2015. "Grouped Patterns of Heterogeneity in Panel Data." *Econometrica* 83 (3): 1147–84. [5, 25]

Charness, Gary, Thomas Garcia, Theo Offerman, and Marie Claire Villeval. 2020. "Do Measures of Risk Attitude in the Laboratory Predict Behavior under Risk in and Outside of the Laboratory?" *Journal of Risk and Uncertainty* 60 (2): 99–123. [31, 35]

Chateauneuf, Alain, Jürgen Eichberger, and Simon Grant. 2007. "Choice under Uncertainty with the Best and Worst in Mind: Neo-additive Capacities." *Journal of Economic Theory* 137 (1): 538–67. [4, 6]

Choi, Syngjoo, Shachar Kariv, Wieland Müller, and Dan Silverman. 2014. "Who Is (More) Rational?" *American Economic Review* 104 (6): 1518–50. [33]

Chuang, Yating, and Laura Schechter. 2015. "Stability of Experimental and Survey Measures of Risk, Time, and Social Preferences: A Review and Some New Results." *Journal of Development Economics* 117: 151–70. [20]

Delavande, Adeline, Jayant Ganguli, and Friederike Mengel. 2019. "Uncertainty Attitudes, Subjective Expectations and Decisions under Uncertainty," 51. [32]

Dimmock, Stephen G., Roy Kouwenberg, Olivia S. Mitchell, and Kim Peijnenburg. 2015. "Estimating Ambiguity Preferences and Perceptions in Multiple Prior Models: Evidence from the Field." *Journal of Risk and Uncertainty* 51 (3): 219–44. [4, 6, 16, 32, 33]

Dimmock, Stephen G., Roy Kouwenberg, Olivia S. Mitchell, and Kim Peijnenburg. 2016. "Ambiguity Aversion and Household Portfolio Choice Puzzles: Empirical Evidence." *Journal of Financial Economics* 119 (3): 559–77. [5]

Dimmock, Stephen G., Roy R. P. Kouwenberg, and Peter P. Wakker. 2016. "Ambiguity Attitudes in a Large Representative Sample." *Management Science* 62 (5): 1363–80. [8, 10–12, 16, 32–34]

Dohmen, Thomas, Armin Falk, David Huffman, Uwe Sunde, Jürgen Schupp, and Gert G. Wagner. 2011. "Individual Risk Attitudes: Measurement, Determinants, and Behavioral Consequences." *Journal of the European Economic Association* 9 (3): 522–50. [14, 20, 24]

Duersch, Peter, Daniel Römer, and Benjamin Roth. 2017. "Intertemporal Stability of Uncertainty Preferences." *Journal of Economic Psychology*. Stability of Preferences 60: 7–20. [4]

Echenique, Federico, Taisuke Imai, and Kota Saito. 2021. "Approximate Expected Utility Rationalization." arXiv: 2102.06331. [34]

Ellsberg, Daniel. 1961. "Risk, Ambiguity, and the Savage Axioms." *The Quarterly Journal of Economics* 75 (4): 643–69. [4, 10]

Falk, Armin, Anke Becker, Thomas Dohmen, David Huffman, and Uwe Sunde. 2022. "The Preference Survey Module: A Validated Instrument for Measuring Risk, Time, and Social Preferences." *Management Science*, [14]

Frey, Renato, Andreas Pedroni, Rui Mata, Jörg Rieskamp, and Ralph Hertwig. 2017. "Risk Preference Shares the Psychometric Structure of Major Psychological Traits." *Science Advances* 3 (10): e1701381. [20]

Friedman, Daniel, R Mark Isaac, Duncan James, and Shyam Sunder. 2014. *Risky Curves: On the Empirical Failure of Expected Utility*. Routledge. [20]

Gaudecker, Hans-Martin von, Arthur van Soest, and Erik Wengström. 2011. "Heterogeneity in Risky Choice Behaviour in a Broad Population." *American Economic Review* 101 (2): 664–94. [5, 11, 14, 33]

Gaudecker, Hans-Martin von, and Axel Wogrolly. 2022. "Heterogeneity in Households' Stock Market Beliefs: Levels, Dynamics, and Epistemic Uncertainty." *Journal of Econometrics*. Subjective Expectations and Probabilities in Economics 231 (1): 232–47. [25]

Ghirardato, Paolo, Fabio Maccheroni, and Massimo Marinacci. 2004. "Differentiating Ambiguity and Ambiguity Attitude." *Journal of Economic Theory* 118 (2): 133–73. [6]

Ghirardato, Paolo, and Massimo Marinacci. 2001. "Risk, Ambiguity, and the Separation of Utility and Beliefs." *Mathematics of Operations Research* 26 (4): 864–90. [4, 6]

Gilboa, Itzhak, and David Schmeidler. 1989. "Maxmin Expected Utility with a Non-Unique Prior." *Journal of Mathematical Economics* 18: 141–53. [4]

Gillen, Ben, Erik Snowberg, and Leeat Yariv. 2019. "Experimenting with Measurement Error: Techniques with Applications to the Caltech Cohort Study." *Journal of Political Economy*, 45. [20]

Harless, David W., and Colin F. Camerer. 1994. "The Predictive Utility of Generalized Expected Utility Theories." *Econometrica* 62 (6): 1251–89. [5]

Heckman, James J., and Burton Singer. 1984. "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data." *Econometrica* 52 (2): 271–320. [5]

Henkel, Luca. 2022. "Experimental Evidence on the Relationship between Perceived Ambiguity and Likelihood Insensitivity." Mimeo, University of Bonn. [4]

Hey, John D., and Chris Orme. 1994. "Investigating Generalizations of Expected Utility Theory Using Experimental Data." *Econometrica* 62 (6): 1291–326. [5]

Hudomiet, Peter, Michael Hurd, and Susann Rohwedder. 2018. "Measuring Probability Numeracy." RAND Working Paper WR-1270. [14]

Johnson, Cathleen A., Aurelien Baillon, Han Bleichrodt, Zhihua Li, Dennie van Dolder, and Peter P. Wakker. 2021. "Prince: An Improved Method for Measuring Incentivized Preferences." *Journal of Risk and Uncertainty* 62: 1–28. [8]

Keane, Michael P., and Kenneth I. Wolpin. 1997. "The Career Decisions of Young Men." *Journal of Political Economy* 105 (3): 473–522. [5, 25]

Keynes, John Maynard. 1921. *A Treatise on Probability*. Cambridge: Cambridge University Press. [4]

Knight, Frank H. 1921. *Risk, Uncertainty and Profit*. New York: Hart, Schaffner, and Marx. [4]

Li, Chen. 2017. "Are the Poor Worse at Dealing with Ambiguity?" *Journal of Risk and Uncertainty* 54 (3): 239–68. [11]

Li, Zhihua, Julia Müller, Peter P. Wakker, and Tong V. Wang. 2018. "The Rich Domain of Ambiguity Explored." *Management Science* 64 (7): 3227–40. [4, 7]

Loomes, Graham, and Robert Sugden. 1995. "Incorporating a Stochastic Element into Decision Theories." *European Economic Review*. Papers and Proceedings of the Ninth Annual Congress European Economic Association 39 (3): 641–48. [5, 15]

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. "Scikit-Learn: Machine Learning in Python." *Journal of Machine Learning Research* 12: 2825–30. [25]

Rooij, Maarten van, Annamaria Lusardi, and Rob Alessie. 2011. "Financial Literacy and Stock Market Participation." *Journal of Financial Economics* 101 (2): 449–72. [14]

Schildberg-Hörisch, Hannah. 2018. "Are Risk Preferences Stable?" *Journal of Economic Perspectives* 32 (2): 135–54. [20]

Steptoe, Andrew, Elizabeth Breeze, James Banks, and James Nazroo. 2013. "Cohort Profile: The English Longitudinal Study of Ageing." *International Journal of Epidemiology* 42 (6): 1640–48. [14]

Trautmann, Stefan T., and Gijs van de Kuilen. 2015. "Ambiguity Attitudes." In *The Wiley Blackwell Handbook of Judgment and Decision Making*. Edited by Gideon Keren and George Wu. Chichester, UK: John Wiley & Sons, Ltd, 89–116. [4]

Wakker, Peter P. 2010. *Prospect Theory: For Risk and Ambiguity*. Cambridge: Cambridge University Press. [6]

Zimpelmann, Christian. 2021. "Stock Market Beliefs and Portfolio Choice in the General Population." CRC TR 224 Discussion Paper No. 258. [31]