Measuring ambiguity preferences: A new ambiguity preference survey module[†]

Elisa Cavatorta*
King's College London, University of London
Strand, London WC2R 2LS, United Kingdom

David Schröder**
Birkbeck College, University of London
Malet Street, Bloomsbury, London WC1E 7HX, United Kingdom.

Published on the *Journal of Risk and Uncertainty*, (2019) 58:71-100
https://link.springer.com/article/10.1007/s11166-019-09299-0

Abstract

Ambiguity preferences are important to explain human decision-making in many areas in economics and finance. To measure individual ambiguity preferences, the experimental economics literature advocates using incentivized laboratory experiments. Yet, laboratory experiments are costly, time-consuming and require substantial administrative effort. This study develops an experimentally validated ambiguity preference survey module that can reliably measure ambiguity preferences when carrying out laboratory experiments is impractical. This toolkit may have wide applications, including end-of-session lab questionnaires, large scale surveys and financial client assessments.

JEL Classification: C81, C83, C91, D81

Keywords: ambiguity, preference measurement, decision making, experimental economics, survey validation

[†]We thank the editor Kip Viscusi and an anonymous referee for their constructive comments. We also thank Johannes Abeler, Aurélien Baillon, Han Bleichrodt, Syngjoo Choi, Peter Dürsch, Roy Kouwenberg, Daniel Martin, Uwe Sunde, Matthias Sutter, Jean-Marc Tallon, Stefan Trautmann, Peter Wakker, Matthias Wibral, seminar participants at Paris-1, as well as conference participants at IMEBESS, the Maastricht behavioral and experimental economics symposium, FUR XVI, Verein für Socialpolitik, ESRC Workshop on Preferences and Personality, NIBS, D-TEA, and the Royal Economic Society for helpful discussion and comments. This work was supported by a BEI research grant by Birkbeck College. Any remaining errors are ours.

*King's College London, Strand, London WC2R 2LS, United Kingdom. E-mail: elisa.cavatorta@kcl.ac.uk. Phone: +44 20 7848 7495.

**Birkbeck College, University of London, Malet Street, Bloomsbury, London WC1E 7HX, United Kingdom. E-mail: d.schroeder@bbk.ac.uk. Phone: +44 207 631 6408.

1 Introduction

In many circumstances of everyday life, people take decisions in uncertain environments. In most situations, the probabilities of the possible outcomes are only vaguely known to decision makers, if at all. Since the seminal works of Knight (1921) and Ellsberg (1961), the absence of precise information on probabilities is referred to as ambiguity, and has been recognized as a form of uncertainty distinct from the standard notion of risk.

Preferences towards ambiguity – and ambiguity aversion in particular – have shown to be an important determinant of individual decision-making.[1] Incorporating ambiguity preferences in economic models helps to explain a variety of phenomena in economics and finance that cannot be attributed to risk aversion alone.[2]

Given the importance of ambiguity preferences for a wide range of decisions in everyday life, it is important for empirical researchers to measure these preferences. The experimental economics literature advocates measuring ambiguity preferences using decision tasks incentivised with money in laboratory settings. However, these procedures are costly, time-consuming, and often complex to implement.

The objective of this paper is to provide a tool to measure ambiguity preferences when carrying out laboratory experiments is impractical, such as in large scale surveys or field studies. We develop an ambiguity preference survey module by combining hypothetical thought experiments and attitudinal questions that can reliably predict ambiguity preferences. Individual ambiguity preferences are measured using a standard, incentivised decision task following the experimental tradition. The hypothetical thought experiments are taken from the economics literature, while the attitudinal questions are predominantly from the psychology literature on ambiguity (in)tolerance. These measures operationalise the concept of ambiguity as vagueness about outcome probabilities (typical in economics), as well as aversion to novelty, complexity and insolubility. Applying a rigorous selection procedure that considers all possible combinations of thought experiments and attitudinal questions, this paper identifies a concise set of predictors that explain a large part of the variation in ambiguity preferences elicited experimentally.

The ambiguity preference survey module consists of five survey items. The first item is a hypothetical, path-dependent thought experiment in the spirit of the original Ellsberg (1961) two-

[1] Empirical studies show that ambiguity aversion can explain, among others, patterns in stock market participation (Bianchi and Tallon, 2018; Dimmock, Kouwenberg and Wakker, 2016; Dimmock, Kouwenberg, Mitchell and Peijnenburg, 2016) and health behaviours among adolescents (Sutter et al., 2013).

[2] For example, theoretical models show that ambiguity aversion can explain the lower than expected investment in financial markets (e.g. Dow and Werlang, 1992; Bossaerts et al., 2010) and real investment projects (e.g. Nishimura and Ozaki, 2007), the economic evaluation of climate change (e.g. Weitzman, 2009), the equity premium puzzle (e.g. Collard et al., 2018) or low reservation wages for job searchers (e.g. Nishimura and Ozaki, 2004).

colour urn experiment. At the same time, this item is the best single predictor of ambiguity preferences. The survey module includes four attitudinal questions. In addition, we also propose a more parsimonious survey module that consists of a smaller subset of three items in total, the thought experiment and two attitudinal questions. Yet, this smaller module comes at the cost of a lower explanatory power. To make the ambiguity preference module easy to use in practical applications, we also propose a ready-to-use ambiguity preference score that allows a simple and quick assessment of ambiguity preferences once the survey data has been collected.

We test the validity of the ambiguity preference module using a variety of tests. First, we examine how well the preference module can explain ambiguity preferences in-sample. Second, we resort to an entirely different sample of subjects of similar size to test the predictive power of the preference module out of sample. Finally, we use a test-retest procedure to establish a benchmark against which the quality of the preference module can be compared. Taken together, these tests show that the ambiguity preference module allows the reliable measurement of ambiguity preference for different samples of subjects.

This paper is related to several streams of literature. Most of all, the paper complements a number of studies that propose survey modules to measure economic preferences without relying on incentivized experiments. So far, this literature has focused on risk preferences, impatience and trust, but has neglected ambiguity preferences. For example, Dohmen et al. (2011) show that a self-reported willingness to take risks correlates with experimentally elicited risk preferences. Questions of this type are now routinely used to measure risk preferences in large surveys, including the German Socio-Economic Panel and other surveys (Guiso and Jappelli, 2008; Vieider et al., 2015; Donkers et al., 2001).

Similarly, time preferences are collected in surveys using self-reported attitudes on impatience and hypothetical choice tasks of present and future rewards (Cole et al., 2013; Bernard et al., 2014). Trust can be measured using attitudinal trust questions, which has been shown to correlate with behaviour in trust games (Fehr et al., 2003). Most closely related to this paper is a recent study by Falk et al. (2016) that examines the ability of hypothetical decision tasks and survey questions to predict preferences for risk, time discounting, altruism, trust, positive and negative reciprocity in incentivised tasks. Like our paper, they propose a survey module to measure these six preferences. This preference module is then used to measure economic preferences in 76 countries around the globe (Falk et al., 2018).

Our paper complements these studies by providing a module for ambiguity preferences. The ambiguity preference module allows researchers to extend empirical studies on ambiguity preferences to large scale field studies based on the general population. This is especially important in light of recent advances in theoretical models assuming ambiguity, whose predictions have rarely been tested empirically.

Furthermore, by analysing the predictive power of the ambiguity preference module using out-of-sample tests, this paper makes also a methodological contribution when designing preference survey modules. While previous papers mainly concentrate on achieving a high explanatory power of their preference modules in-sample, these modules are rarely tested on different, independent samples. By showing that our ambiguity preference module can reliably explain ambiguity preferences for a different sample of similar size, we further substantiate the validity of the preference module.

This paper is also related to some recent studies that measure ambiguity preferences in large representative samples. In most of these studies, however, the ambiguity preference elicitation is not experimentally validated. For example, Butler et al. (2014) and Bianchi and Tallon (2018) assume that their non-incentivized Ellsberg-style thought experiments reveal true preferences. In fact, by showing that these thought experiments are significantly correlated to ambiguity preferences obtained in incentivized decision tasks, our study gives empirical support to their assumptions. Studies that measure ambiguity preferences for large samples of the population using incentivized decision tasks are rare. Four notable examples are the papers by Cohen et al. (2011), Dimmock et al. (2015), Dimmock, Kouwenberg, Mitchell and Peijnenburg (2016) and Dimmock, Kouwenberg and Wakker (2016). Yet, given the financial and administrative costs, such large-scale field studies are likely to remain the exception rather than the rule.[3]

The remainder of the paper develops as follows. Section 2 describes the experimental design of this study, including a detailed description of the various measures of ambiguity preferences. Next, section 3 presents the experimental results. The ambiguity preference module is derived in section 4 together with a series of validity tests. Section 5 provides some discussion and conclusions.

2 Experimental design

This section presents the research design, followed by a detailed description of the incentivized tasks, hypothetical thought experiments and attitudinal questions to measure ambiguity and risk preferences. Then we present the experimental procedure and provide a short description of the participants.

[3]For example, the study by Dimmock, Kouwenberg, Mitchell and Peijnenburg (2016) paid USD 23,850 in real incentives.

2.1 Research design

This study evaluates whether hypothetical thought experiments and attitudinal survey questions can offer a reliable alternative to measure individual ambiguity preferences as obtained in incentivized laboratory decision tasks.

Our research design involves two parts. In the first part, we measure each subject's ambiguity preferences using a standard task with monetary rewards. Such incentives aim to ensure that the subjects' choices reveal their true preferences. The preferences obtained from the incentivized task form our experimental ambiguity preference benchmark. In the second part, we measure the subjects' ambiguity preferences using hypothetical thought experiments and attitudinal questions. Then we assess whether these measures of ambiguity preferences can reliably predict the ambiguity preference benchmark.

It is well known that ambiguity preferences depend on many factors. For example, Curley and Yates (1989) show that ambiguity preferences depend on the probability of the risky alternative offered. When facing unlikely events, subjects tend to exhibit ambiguity-seeking preferences because of overweighing of low likelihood events. Furthermore, Cohen et al. (1987) show that ambiguity preferences are contingent on the outcome domain. When confronted with potential losses, subjects are less ambiguity averse, or even show ambiguity-seeking preferences (Chakravarty and Roy, 2009; Kocher et al., 2015). To minimize the impact of such factors, the incentivized task and hypothetical thought experiments measure ambiguity preferences in the gain domain, using non-extreme probability ranges.

This paper measures individual ambiguity preferences in the sense of the Ellsberg (1961) thought experiment, defining ambiguity as the absence of information on exact probabilities. While there are other sources of ambiguity (Fox and Tversky, 1995; Abdellaoui et al., 2011; Baillon et al., 2018), the ambiguity notion in the tradition of Ellsberg is the most important ambiguity concept in the experimental and behavioural economics literature and the paper focuses on that concept. By considering a large set of hypothetical thought experiments and attitudinal questions, our research design allows us to measure Ellsberg-type ambiguity preferences. However, it is less suited to measure a separate aspect of decision-making under ambiguity: the tendency of subjective overweighting (and underweighting) of extreme likelihood ambiguous events, known in the literature as a-insensitivity. This concerns the observed tendency for subjects to transform extreme likelihoods towards un-extreme, mid-range, probabilities. The implication of a-insensitivity is that it may lead subjects to exhibit more ambiguity-seeking behaviour for low-likelihood events and more ambiguity-averse behaviour for high-likelihood events. While a-insensitivity is an important issue, its separate measurement, in addition to ambiguity preferences, imposes the use and selection of additional probability-based incentivized tasks and probability-based hypothetical thought experiments for low, medium, high likelihood events.[4] This would have significantly increased the number of probability-based tasks relative to the others and increased the duration of the experiment. To avoid introducing an overly complicated design with many probability-based tasks by construction, we use non-extreme probability ranges at the cost of being unable to elicit a-insensitivity.[5]

2.2 Incentivized tasks

The literature has proposed many different designs to measure ambiguity preferences.[6] We rely on the well-established design to measure ambiguity preferences using binary choice lists between risky and ambiguous lotteries. The lotteries are presented in the form of two-colour urns, similar to Ellsberg (1961). Binary choice lists are considered easier to understand than the BDM mechanism (Becker et al., 1964), and are thus likely to result in more precise estimates of ambiguity preferences.

The ambiguity preference task presents subjects with a decision table with 11 choices between drawing a ball from either a risky or an ambiguous urn. The composition and payoff structure of the ambiguous urn is identical in all 11 situations. In contrast, the expected payoff of the risky urn increases from one situation to the next. This change is induced by increasing the probability of winning some prize, while leaving the potential prize constant. The task has been implemented in student and non-student samples as early as by Lauriola and Levin (2001). As Dimmock et al. (2015) show, this particular design allows measuring ambiguity preference independent of the subject's utility function, and thus risk preferences. The point at which subjects switch from preferring the ambiguous urn over the risky urn reveals their ambiguity preferences.[7]

To verify that ambiguity preferences are distinct from risk preferences, we also include a task to elicit risk preferences. We use a standard multiple choice list, taken from Chakravarty and

[4]One example is the design by Dimmock, Kouwenberg and Wakker (2016) who use three sets of incentivized dynamic choice tasks to elicit ambiguity preferences for three different levels of ambiguity. They show that a-insensitivity is related to financial behaviour.

[5]In our considerations, there was also a trade-off between the duration of the experiment (and the incentives paid) and the number of participants our budget allowed.

[6]For a review, see Camerer and Weber (1992), Trautmann and van de Kuilen (2016), and Carbone et al. (2017).

[7]We also considered measuring ambiguity preferences using a binary choice list where the expected payoff of the risky urn changes by monotonically decreasing its prize (but not its probability), see for example Chakravarty and Roy (2009). However, in such a setting the switching point depends on the subjects' risk preferences. Although it is possible to first estimate risk preferences, such a procedure increases the overall measurement error of the estimated ambiguity preferences.

Roy (2009) that proved intelligible in the piloting phase. In this task, subjects are presented a decision table with 10 choices between drawing a ball from a low-risk urn and high-risk urn. As the list proceeds, the low-risk urn remains identical while the expected payoff from the high-risk urn increases monotonically. The point at which subjects switch from the low-risk urn to the high-risk urn reveals information on subject risk preferences. Both tasks are reproduced in the online appendix A.

2.3 Hypothetical thought experiments

Following Ellsberg (1961), the literature has proposed a large variety of hypothetical thought experiments to measure ambiguity preferences. Non-incentivized elicitation methods have the advantage of lower costs and reduced complexity, thereby reducing the time needed to measure preferences. In this study, we implement two common thought experiments to measure ambiguity preferences:[8]

i. The first thought experiment replicates the seminal two-colour Ellsberg (1961) urn experiment. Following Butler et al. (2014), subjects are offered five answer possibilities to allow differentiating between different degrees of ambiguity preferences.

ii. Second, we design a dynamic version of the Ellsberg (1961) two-colour thought experiment. Subjects are presented four (in some cases five) sequential choices between drawing a ball from a risky or an ambiguous urn. Depending on prior choices, the composition of the risky urn is made more or less attractive from one situation to the next, converging to a point of (approximate) indifference between both urns. This gives a matching probability similar to the incentivized ambiguity task.[9]

Online appendix B presents the two thought experiments in detail.

2.4 Attitudinal survey questions

While using survey questions to measure risk preferences is common (Guiso and Paiella, 2008; Dohmen et al., 2011), there is yet no study in economics that explicitly uses questionnaires to measure ambiguity preferences. We hence resort to survey questions from validated self-reported attitudinal scales in the psychology literature on ambiguity (in)tolerance. In these scales subjects

[8] Any incentivized task can be transformed into a non-incentivized task by removing the monetary incentives. However, the complexity remains higher than in usual thought experiments.

[9] Similar path dependent designs are used in Baillon et al. (2012), Baillon and Bleichrodt (2015), Dimmock et al. (2015) and Dimmock, Kouwenberg, Mitchell and Peijnenburg (2016).

are asked to indicate the extent to which they agree or disagree with a list of statements on a scale from 1 to 7.

The concept of ambiguity (in)tolerance used in psychology is more general than in economics. While it comprises the typical notion of aversion to vague outcome probabilities, the concept also refers to aversion to novelty, complexity and insolubility (Budner, 1962). Aversion to situations with unclear structure (e.g., where there is no clear solution) is an indication of aversion to ambiguity. For example, the scales contain statements such as "Practically every problem has a solution".

We implement the Intolerance of Ambiguity Scale by Kirton (1981), one of the most widely used and renowned scales on ambiguity attitudes in psychology.[10] In addition, we include selected items from more recent scales, like the Ambiguity Tolerance scales (Budner, 1962; Norton, 1975; McLain, 2009) and the Uncertainty Response Scale by Greco and Roger (2001) that correspond to the notion of ambiguity as absence of exact probabilities used in the economics literature. Since it has been argued that ambiguity attitudes are related to optimism and pessimism (Chateauneuf et al., 2007) and self-esteem (Heath and Tversky, 1991), we include attitudinal questions on optimism/pessimism from the Extended Life Orientation test by Chang et al. (1997) and a self-esteem measure by Robins et al. (2001). Together with some of our own additions, the survey questionnaire contains 46 attitudinal questions in total. The complete list is included in the online appendix C.

2.5 Experimental procedure

The experiment was conducted in May and October 2013 at Birkbeck College, University of London. The laboratory sessions were implemented in z-tree (Fischbacher, 2007). Each session consisted of four parts. The first part included the attitudinal questionnaire as described in section 2.4. The hypothetical thought experiments (see section 2.3) followed in part 2. The third part included a standard demographic questionnaire. Finally, the last section consisted of the incentivized tasks, see section 2.2. The first task was the risk task, followed by the ambiguity task.

This particular sequence was chosen to reduce potential spill-over effects between the four parts. More precisely, subjects' answers in one part might be influenced by answers in previous parts – for example in an attempt to be consistent.[11] However, because of monetary incentives, we

[10]Kirton (1981) is widely used in empirical work in social psychology. For a review, see Furnham and Ribchester (1995). The scale develops from earlier works by Budner (1962), Rydell and Rosen (1966) and Mac Donald Jr. (1970).

[11]For example, individuals aim to be consistent in order to avoid cognitive dissonance (Festinger, 1957; Falk

expected choices in the decision tasks to be uninfluenced by non-incentivised thought experiments and attitudinal questions. Hence, placing the incentivized decision tasks at the end minimises potential spill-over effects across parts. In two pilot sessions we reversed the order to test for order effects, and found none.

In the ambiguity tasks, participants were asked to select the colour of the winning ball. This ensures that subjects had no reason to believe that the experimenter had any strategic incentive to manipulate the colour of the balls in the ambiguous urn (Chow and Sarin, 2002; Charness et al., 2013).

The payment modality of the incentivised tasks was common knowledge. Subjects were informed that one situation from each task would be randomly selected by the computer at the end of the session. Then the computer would randomly draw one ball from the urn chosen. This procedure ensures that subjects state their true preferences.[12] Earnings from the tasks were calculated in terms of points, and then converted at a rate of 2:1 into GBP. On average, subjects earned GBP 18.45, which includes a fixed show-up fee of GBP 10.[13] Earnings were paid in private at the end of the sessions.

2.6 Participants

There were 121 participants in the study, all of them students of three University of London colleges. The participants were recruited via electronic mail and announcements at the beginning of graduate and undergraduate lectures of various study programmes. The sample contains 54 (45%) male and 67 (55%) female subjects, with an average age of about 26 years. The sample of participants predominantly came from Birkbeck College (University of London) which has a professional and international student population. This gives the sample more variability than that usually observed in experimental subject pools of undergraduate students. For example, almost a fifth of the subject sample is older then 30. For the detailed sample statistics see the online appendix D.

and Zimmermann, 2017).

[12] In a recent paper, Bade (2015) presents some problems with this random incentive mechanism when subjects are ambiguity averse. Yet, to our knowledge, the random incentive mechanism remains the best incentive scheme to measure ambiguity preferences and it is commonplace in laboratory and field studies.

[13] The lowest payment was GBP 10, the highest payment GBP 26. Since the sessions lasted for about 40 minutes, the payoffs are sizable.

3 Experimental results

3.1 Ambiguity task

Table 1 presents the subjects' choices between drawing a ball from the risky or the ambiguous urn in the incentivized ambiguity task. In around 55% of the situations, subjects prefer drawing a ball from the risky urn over drawing a ball from the ambiguous urn. In binary choice lists, a typical strategy is a threshold strategy. Since the relative attractiveness of the lotteries changes monotonically from situation to situation, many participants prefer one urn over the other up to a switching point, from which they prefer the other urn. Since there is no chance of winning anything in the risky urn (urn 1) in situation 0, the ambiguous urn (urn 2) is the natural choice. However, as the probability of winning a prize in the risky urn increases, it becomes more attractive.

Table 1: Descriptive statistics of incentivized ambiguity task

Panel A: Choices and switching points

Risky choices			Switching point			Switching point (consistent DMs)		
Risky choices	Obs.	Fraction	Switching point	Obs.	Fraction	Switching point	Obs.	Fraction
0/11	0	0.0%	0	0	0.0%	0	0	0.0%
1/11	0	0.0%	1	0	0.0%	1	0	0.0%
2/11	0	0.0%	2	2	1.7%	2	2	1.7%
3/11	0	0.0%	3	9	7.4%	3	8	6.9%
4/11	7	5.8%	4	28	23.1%	4	27	23.3%
5/11	26	21.5%	5	49	40.5%	5	48	41.4%
6/11	50	41.3%	6	25	21.5%	6	25	21.6%
7/11	28	23.1%	7	6	5.0%	7	6	5.2%
8/11	8	6.6%	8	1	0.8%	8	0	0.0%
9 /11	2	1.7%	9	0	0.0%	9	0	0.0%
10/11	0	0.0%	10	0	0.0%	10	0	0.0%
11/11	0	0.0%	11	0	0.0%	11	0	0.0%
Total	121	100.0%	Total	121	100.0%	Total	116	100.0%

Panel B: Summary statistics

	Observations	Mean	Standard deviation	Lowest	Highest
Risky choices	121	55.3%	0.10%	36.4%	81.8%
Switching point	121	4.91	1.08	2	8
Switching point (consistent DMs)	116	4.90	1.04	2	7

The table summarizes the results of incentivized ambiguity task. In panel A, the three columns on the left report the number of situations where subjects preferred drawing a ball from the risky urn (urn 1) over drawing a ball from the ambiguous urn (urn 2). The columns on the right report the switching points of the subjects. More precisely, it indicates the last situation before a subject switches from the ambiguous urn (urn 2) to the risky urn (urn 1). Five subjects exhibit multiple switching points, i.e., they display inconsistent choices. In this case, we calculate the average switching point. The three columns on the right exclude these inconsistent decision makers. Panel B presents the summary statistics of the fraction of risky choices and the switching points. For a detailed description of the task, see online appendix A.

Yet, subjects may switch from one choice to the other more than once. This behaviour is difficult to reconcile with standard preference models under ambiguity. The percentage of subjects that switch more than once is fairly small at around 4%, in line with the study on risk preferences by Holt and Laury (2002).

The columns on the right of table 1 present the switching points of the subjects, i.e., the last situation *before* a subject switches from the ambiguous urn (urn 2) to the risky urn (urn 1). In case a subject exhibits multiple switching points, we follow Falk et al. (2016) and calculate the subject's average switching point. Since multiple switching pointy may indicate a misapprehension of the task, we also present the switching points for the sample of consistent decision makers.

3.2 Risk task

Table 2 summarizes the results from the risk task. In about 56% of the situations, subjects prefer drawing a ball from the relatively safe urn over drawing a ball from the very risky urn. Again, the large majority of subjects (95%) exhibits a threshold strategy. Since the expected earnings of the risky urn are zero in situation 0, all subjects prefer the safe urn. As the relative attractiveness of the risky lottery increases monotonically from situation to situation, many subjects switch at some stage to the risky urn. In case a subject exhibits multiple switching points, we again calculate the subject's average switching point. The two columns on the right present the switching points for all decision makers and consistent decision makers separately.

3.3 Ambiguity preference benchmark

This section derives the experimental ambiguity preference benchmark from the switching points of the incentivized ambiguity task. Following Dimmock et al. (2015), we use the subject's so-called matching probability (or risk equivalent) as the ambiguity preference benchmark. The matching probability m is defined as the subjective probability at which a subject is indifferent between a risky and the ambiguous lottery. If $m < 0.5$, a subject is ambiguity-averse; if $m > 0.5$ a subject is ambiguity-seeking.[14]

We construct a subject's matching probability from the switching point in the incentivized ambiguity task. Since the probability of winning a prize in the risky urn increases in 10% steps,

[14]Because of symmetry of the two colours and an implicitly assumed exchangeability condition (Chew and Sagi, 2006), a matching probability of 0.5 is the subjective probability of an ambiguity-neutral decision maker. From this, it follows that a matching probability larger than 0.5 corresponds to ambiguity seeking preferences, and that a matching probability smaller than 0.5 corresponds to ambiguity averse preferences. See Dimmock et al. (2015) for a theoretical derivation.

Table 2: Descriptive statistics of incentivized risk task

Panel A: Choices and switching points

Choices			Switching point			Switching point (consistent DMs)		
Safe choices	Obs.	Fraction	Switching point	Obs.	Fraction	Switching point	Obs.	Fraction
0/10	0	0.0%	0	0	0.0%	0	0	0.0%
1/10	1	0.8%	1	1	0.8%	1	1	0.8%
2/10	1	0.8%	2	1	0.8%	2	1	0.8%
3/10	4	3.3%	3	4	3.3%	3	4	3.5%
4/10	23	19.0%	4	23	19.0%	4	23	20.0%
5/10	36	29.8%	5	37	30.6%	5	34	29.6%
6/10	22	18.2%	6	21	17.4%	6	19	16.5%
7/10	18	14.9%	7	18	14.9%	7	18	15.7%
8/10	10	8.3%	8	10	8.3%	8	9	7.8%
9/10	2	1.7%	9	2	1.7%	9	2	1.7%
10/10	4	3.3%	10	4	3.3%	10	4	3.5%
Total	121	100.0%	Total	121	100.0%	Total	115	100.0%

Panel B: Summary statistics

	Observations	Mean	Standard deviation	Lowest	Highest
Save choices	121	56.4%	0.17%	10%	100%
Switching point	121	5.64	1.66	1	10
Switching point	115	5.63	1.68	1	10

The table summarizes the results of incentivized risk task. In panel A, the three columns on the left report the number of situations where subjects preferred drawing a ball from the safe urn (urn A) over drawing a ball from the risky urn (urn B). The columns on the right report the switching points of the subjects. More precisely, it indicates the last situation before a subject switched from the safe urn (urn A) to the risky urn (urn B). Six subjects exhibit multiple switching points, i.e., they display inconsistent choices. In this case, we calculate the average switching point. The three columns on the right exclude these inconsistent decision makers. A switching point of 0 means that the subject always preferred the risky urn; a switching point of 10 means that the subject always preferred the safe urn. Panel B presents the summary statistics of the fraction of safe choices and the switching points. For a detailed description of the task, see online appendix A.

Table 3: Ambiguity preference benchmark

Panel A: Ambiguity preference benchmark (m)

	Obs.	Mean	Std. Dev.	Min	Max
All subjects	121	0.441	0.108	0.15	0.75
Only consistent DMs	116	0.440	0.104	0.15	0.65

Panel B: Distribution of ambiguity preferences

	Groups	Ambiguity seeking	Ambiguity neutral	Ambiguity averse
All subjects	2	27%	–	73%
	3	6%	62%	32%
Only consistent DMs	2	27%	–	73%
	3	5%	63%	32%

Panel C: Correlation between benchmark, socio-demographic characteristics and risk preferences

	Rank correlation	p-value
Female	−0.052	0.568
Age	0.078	0.396
Education of mother	0.003	0.973
Risk aversion	0.070	0.448

The table presents the ambiguity preferences of the incentivized task as measured by the subject's matching probability m (or risk equivalent). Panel A reports the descriptive statistics. Panel B describes the distribution of ambiguity preferences when classifying subjects in two or three ambiguity preference groups. The upper row uses a binary classification that distinguishes between ambiguity seeking ($m > 0.5$) and ambiguity averse decision makers ($m < 0.5$). The lower row presents the results based on a classification into three groups, including ambiguity neutrality, where ambiguity neutrality is defined as $m \in [0.45, 0.55]$. Panel C presents the Spearman rank correlation between the ambiguity preference benchmark, selected socio-economic characteristics and an experimental measure of risk aversion. *, **, and *** denote statistical significance at the 10%, 5% and 1% level, respectively.

the task only allows us to determine a subject's matching probability up to intervals of 10%. We therefore define the matching probability as the mid-point of these intervals. For example, if a subject's switching point is 5 (i.e., preferring urn 2 up to situation 5 and then switching to urn 1), then the matching probability is between 50% and 60%. We therefore assign a matching probability of 55%.

Panel A of table 3 presents the summary statistics of the preference benchmark for the entire sample and the sample of consistent subjects. The average matching probability is around 0.44, which corresponds to ambiguity-averse preferences, on average.[15]

Panel B presents the distribution of ambiguity preferences when classifying subjects in two or three ambiguity preference groups. The upper row uses a binary classification that distinguishes between ambiguity seeking ($m > 0.5$) and ambiguity averse decision makers ($m < 0.5$). The

[15]This is similar to Dimmock, Kouwenberg and Wakker (2016) and Dimmock, Kouwenberg, Mitchell and Peijnenburg (2016), which document an average matching probability of 0.40 and 0.48, respectively.

lower row presents the results based on a classification into three groups, including ambiguity neutrality, where ambiguity neutrality is defined as $m \in [0.45, 0.55]$.[16] The benchmark classifies 73% of subjects as ambiguity averse using a binary segmentation of preferences. However, a classification in three preference categories shows that almost two thirds of the sample can be considered ambiguity neutral.

Panel C presents the Spearman rank correlation between the ambiguity preference benchmark and several socio-economic characteristics. Like Binmore et al. (2012) and Dimmock et al. (2015), we do not find any gender differences in ambiguity preferences, nor an effect of the educational background or age. Finally, we replicate the standard result that ambiguity preferences are distinct from risk preferences.

3.4 Thought experiments

Table 4 analyses the results from the two thought experiments. Panel A shows that in the Ellsberg (1961) thought experiment with five answer possibilities, 64% of subjects are ambiguity averse, 28% are ambiguity seeking and 8% ambiguity neutral. In the dynamic Ellsberg urn experiment, we transform the subjects' choices into a matching probability, similar to the incentivized ambiguity task. More precisely, after the last choice, a subject's matching probability is narrowed down to a small interval. As before, we define the matching probability as the mid-point of that interval. Panel B shows that the average matching probability is 44%. This is remarkably similar to the ambiguity preference benchmark, and consistent with the notion of ambiguity aversion.[17]

Panel C presents the distribution of subjects into two or three broad ambiguity preference groups, similar to panel B of table 3. Yet, since the Ellsberg thought experiment explicitly allows for ambiguity neutrality, it is not possible to apply a binary classification for this elicitation method. When using a binary classification, most subjects are ambiguity averse. However, like in the incentivized task, there is a substantial fraction of ambiguity-neutral decision makers.

Panels D and E compare the ambiguity preferences obtained from the thought experiments to the benchmark. With 27%, the dynamic Ellsberg urn exhibits the highest rank correlation with the benchmark. The classic Ellsberg urn experiment correlates to 19% with the ambiguity preference benchmark. This lower correlation is likely to result from the significantly fewer answer possibilities, and hence cross-sectional variation, in this thought experiment.

[16] If a subject is ambiguity neutral ($m = 0.5$), then he is indifferent between switching after situation 4 or 5, which corresponds to an estimated matching probability of 0.45 or 0.55.

[17] In both thought experiments, the majority of subjects is ambiguity averse ($p < 0.01$, binomial probability test).

Table 4: Non-incentivized thought experiments

Panel A: Ellsberg (1961) urn

	Observations	Fraction
Strong preference for ambiguous urn	12	9.9%
Slight preference for ambiguous urn	22	18.2%
Indifferent between both urns	10	8.3%
Slight preference for risky urn	40	33.1%
Strong preference for risky urn	37	30.6%
Total	121	100.0%

Panel B: Dynamic Ellsberg urn

Matching probability	Observations	Fraction
2.5%	1	0.8%
12.5%	4	3.3%
22.5%	5	4.1%
27.5%	5	4.1%
32.5%	11	9.1%
37.5%	25	20.7%
42.5%	6	5.0%
47.5%	20	16.5%
52.5%	27	22.3%
57.5%	10	8.3%
62.5%	1	0.8%
67.5%	2	1.7%
72.5%	2	1.7%
77.5%	1	0.8%
95%	1	0.8%
Total	121	100.0%

	Mean	Standard deviation
Matching probability	43.9%	13.8%

Panel C: Distribution of ambiguity preferences

Elicitation method	Groups	Ambiguity seeking	Ambiguity neutral	Ambiguity averse
Ellsberg (1961) urn	2	–	–	–
	3	28%	8%	64%
Dynamic Ellsberg urn	2	36%	–	64%
	3	14%	39%	47%

Panel D: Correlation between ambiguity preferences and non-incentivized tasks

Elicitation method	Rank correlation	p-value
Ellsberg (1961) urn	0.186^{**}	0.041
Dynamic Ellsberg urn	0.268^{***}	0.003

Panel E: Individual consistency of ambiguity preferences and non-incentivized tasks

Elicitation method	Consistent preferences	Difference from independent distribution	Test of association
Ellsberg (1961) urn	35%	$7.45\%^{*}$	
Dynamic Ellsberg urn	53%	$12.8\%^{***}$	***

Annotation to table 4

The table summarizes the results of the two non-incentivized thought experiments. Panels A and B present the descriptive statistics. Panel C describes the distribution of ambiguity preferences when classifying subjects in two/three broad ambiguity preference groups, similar to table 3. Since the Ellsberg thought experiment explicitly allows for ambiguity neutrality, it has been excluded from the binary classification. Panel D presents the Spearman rank correlation between the ambiguity preference benchmark and the ambiguity preferences obtained from the non-incentivized tasks. Panel E reports the individual consistency of the estimated ambiguity preferences when classifying subjects into three broad ambiguity preferences groups. The first column presents the fraction of individually consistent preferences, i.e., subjects that exhibit the same ambiguity attitude in the thought experiments and the ambiguity preference benchmark. The second column presents the difference of the observed consistency compared to the fraction of individually consistent preferences that one would observe if the ambiguity measures were independently distributed. The last column shows the statistical significance of a two-sided Fisher test of association between each pair of ambiguity preferences. For a detailed description of the thought experiments, see online appendix B.

Panel E presents an analysis of the subjects' within-person consistency of the ambiguity preferences obtained from the two thought experiments relative to the ambiguity preference benchmark. More precisely, the panel presents the fraction of subjects that exhibit the same ambiguity attitude in the thought experiments as in the ambiguity preference benchmark. This analysis shows that the dynamic Ellsberg thought experiment performs rather well in classifying subjects similar to the benchmark (53% of the subjects), which is significantly higher than the fraction of consistent preferences one would observe if the preferences were independently distributed.

4 Ambiguity preference module

This section examines whether thought experiments and attitudinal questions can be used to predict ambiguity preferences. We identify a set of thought experiments and attitudinal questions that best predict the ambiguity preference benchmark (section 4.1) and propose an ambiguity preference survey module that reliably measures ambiguity preferences when laboratory experiments are not feasible (section 4.2). We then test the validity of the module (section 4.3) and propose a ready-to-use ambiguity preference module for applied work (section 4.4).

4.1 Item selection procedure

The construction of the ambiguity preference survey module faces the challenge of finding the optimal trade-off between parsimony and a sufficiently high explanatory power of ambiguity preferences. To determine the items that best proxy for ambiguity preferences, we follow a procedure inspired by Falk et al. (2016). The rationale of this procedure is to make as few a-priori assumptions as possible in regards to the item selection and let the selection be data driven.

This procedure consists of two steps. First, for a given number of predictors, we estimate linear regression models for *all* possible combinations of predictors. Considering all possible combinations of thought experiments and attitudinal questions on an equal basis avoids imposing arbitrary selections and takes into account all possible correlation patterns among the predictors. Then we select the set of predictors that maximize the explained variation of the ambiguity preference benchmark, i.e. the R^2. Evaluating the predictive power for a given number of predictors is useful for practical purposes since survey data collection often imposes space and time constraints. In the second step, we use standard information criteria to select the best specification out of the models chosen in the first step. In an attempt to avoid over-fitting the data and identifying a concise preference module, we resort to the Bayesian Information Criterion (BIC), since it contains a larger penalty for additional explanatory variables relative to the Akaike Information Criterion (AIC) or the adjusted R^2.[18]

4.2 The ambiguity preference module

Table 5 presents the selected items for a given number of predictors up to the specification with the lowest overall BIC. Panel A reports the regressions when using all decision makers; panel B reports the regressions when considering the subset of consistent decision makers only. Predictors measured on Likert scales (e.g., survey questions) are coded as binary indicators (equal to 1 if the response value indicates agreement with an ambiguity tolerant statement and equal to 0 otherwise), using the median respondent value in the sample as the threshold for the classification. The estimated coefficients on the binary indicators (multiplied by 100) estimate the differential in terms of risk equivalent (i.e. percentage point differential) between ambiguity tolerant and ambiguity averse people.[19] Predictors measured as continuous variables are treated as continuous variables.[20] Column (1) of table 5 shows the best predictor (in terms of R^2) when

[18] The choice set of alternative item selection procedures is vast. Alternatives vary along the dimensions of i) the selection procedure; ii) the goodness of fit statistics used; iii) the specification used in the iteration. Alternative selection procedures in (i) include stepwise forward selection, stepwise backward selection, or the Lasso technique by Tibshirani (1996). Each of these techniques has considerable drawbacks that make them unsuitable for our purpose. For a critical discussion on these alternative selection procedures see Falk et al. (2016), footnote 13. Alternatives in (ii) include the BIC (which we use), the adjusted R^2, AIC or the Mean Squared Error. These statistics apply different penalties for the estimation of additional parameters. Alternatives in (iii) include non-linear functional forms.

[19] In order to classify responses into binary indicators (i.e. ambiguity tolerant/intolerant), we use the sample median value for each survey question to take into account the distribution of sample responses. Some of the survey question responses have skewed distributions. This is likely to be due to the composition of our subject sample: students. For example, the majority of young people agree to the statement "I would like to live in a foreign country new to me." While the use of the sample median is not indispensable, it creates relative classifications within sample.

[20] Two alternative specifications are possible. A more flexible approach would be to implement a fully non-parametric specification with a binary indicator for each value of the rating scales (=1 if that value is chosen, =0 otherwise). The advantage of this approach is to allow for a flexible, non-linear, pattern of predictors' correlations

using one predictor; column (2) shows the two best predictors when using two predictors, and so on.

For both samples, the best single predictor is the dynamic Ellsberg thought experiment, which alone explains about 5% of the variation in the ambiguity preference benchmark. In fact, this hypothetical thought experiment is very similar in spirit to the hypothetical thought experiment used to measure risk preferences in the streamlined version of the preference module by Falk et al. (2016).

When using the entire sample of 121 subjects (panel A), the best specification in terms of BIC contains five explanatory variables, including the dynamic Ellsberg thought experiment and four survey questions. Together, these items explain about 21% of the variation in ambiguity preferences. When restricting to the sample of 116 consistent subjects (panel B), the best specification (in terms of BIC) includes three predictors, the dynamic Ellsberg thought experiment and two survey questions, a subset of the items selected of the entire data sample. These three predictors achieve an R^2 of about 12%. The explanatory power of both sets of models is comparable to those of risk preferences reported by Falk et al. (2016).

Table 6 presents the exact wording of the ambiguity preference modules for both samples.[21]

It makes sense that the dynamic Ellsberg thought experiment is the best single-item predictor of ambiguity preferences: it has a similar framing as the incentivised task and by producing a finer measurement classification, it provides more explanatory power than qualitative tasks with coarser classifications. In terms of the attitudinal questions, survey items 1 and 2 are part of the Kirton (1981) Intolerance of Ambiguity Scale. Survey items 35 and 41 are taken from the uncertainty response scale by Greco and Roger (2001) and the ambiguity tolerance scale II (McLain, 2009). The former two survey items capture the aversion to complexity and unclear structure, the latter two statements capture the aversion to uncertain outcomes.

By having identified that a dynamic Ellsberg thought experiment is a good predictor of ambiguity preferences, the module gives confidence in the measurement validity of probability-based

with the ambiguity benchmark. As we measured the survey item predictors on 7-point Likert scales, this approach would imply estimating 6 parameters for each ordinal predictor. Given our sample size, such an approach would be inappropriate as we would run out of degrees of freedom quickly. However, the fully non-parametric approach would be advisable for future studies with larger samples. A second alternative approach is to treat rating scales as continuous variables. This approach relies on the assumption of linearity, which may, but need not, be satisfied. As a robustness test, we run auxiliary regressions to test the linearity assumption in our sample and we conclude that the linearity assumption cannot be rejected for the items selected. These tests are available in the online appendix E.

[21] This result shows that the dynamic Ellsberg thought experiment is a better predictor relative to the static two-colour Ellsberg thought experiment. This does not mean that the Ellsberg thought experiment is not a good proxy, but that the dynamic Ellsberg thought experiment explains more variation in the benchmark. Indeed, the static Ellsberg thought experiment is significantly correlated with ambiguity preferences in pairwise tests (see table 4). This provides a validation to existing applications of the static Ellsberg thought experiment, e.g., Butler et al. (2014).

Table 5: Predicting the ambiguity preference benchmark

Panel A: All subjects

Explanatory variables	(1)	(2)	(3)	(4)	(5) best
Dynamic Ellsberg urn	0.0017**	0.0017**	0.0019***	0.0021***	0.0023***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Item 19		-0.0392**			
		(0.019)			
Item 1			0.0501**	0.0545***	0.0485**
			(0.020)	(0.020)	(0.020)
Item 2			0.0501**	0.0524***	0.0488**
			(0.020)	(0.020)	(0.020)
Item 35				0.0276	0.0402*
				(0.021)	(0.021)
Item 41					-0.0425**
					(0.020)
constant	0.3655***	0.3883***	0.3099***	0.2883***	0.2993***
	(0.032)	(0.034)	(0.036)	(0.040)	(0.039)
BIC from selection procedure	-192.54	-192.05	-195.26	-194.66	-195.91
R^2 from selection procedure	4.8%	8.1%	14.0%	16.9%	20.9%
Observations	121	121	121	121	121

Panel B: Consistent subjects

Explanatory variables	(1)	(2)	(3) best
Item 19	-0.0391**	-0.0398**	
	(0.019)	(0.019)	
Dynamic Ellsberg urn		0.0015**	0.0016**
		(0.001)	(0.001)
Item 1			0.0387*
			(0.020)
Item 2			0.0520***
			(0.020)
constant	0.4622***	0.3969***	0.3269***
	(0.015)	(0.033)	(0.036)
BIC from selection procedure	-191.64	-191.46	-192.31
R^2 from selection procedure	3.8%	7.6%	11.9%
Observations	116	116	116

The table presents regressions of the ambiguity preference benchmark (m) on the preferences obtained from non-incentivized thought experiments and the attitudinal questions. In section 4.1, we estimate linear regression models for all possible combinations of predictors. The table above reports the specifications with the highest adjusted R^2 for a given number of predictors *up to* the specification with the lowest overall BIC (the optimal specification). BIC and R^2 statistics of these specifications are reported in the bottom rows of each panel. Predictors measured on Likert scales (i.e., the survey questions, indicated with the term 'Item') are coded as binary indicators (equal to 1 if the response value indicates agreement with an ambiguity tolerant statement and equal to 0 otherwise), using the median respondent value in the sample as the threshold for the classification. Panel A reports the regressions when using all decision makers; panel B reports the regressions when only considering consistent decision makers. T-statistics are given in the parentheses below the estimated coefficients. *, **, and *** denote statistical significance at the 10%, 5% and 1% level, respectively.

Table 6: The ambiguity preference modules

Module items	In-sample prediction Observations	In-sample prediction Correlation with benchmark	Out-of-sample prediction Observations	Out-of-sample prediction Correlation with benchmark
Panel A: Long preference module (derived from all subjects)	121	0.408***	99	0.331***
		(< 0.001)		(0.0008)
Thought experiment: Dynamic Ellsberg urn				
Survey item 01: There is a right way and a wrong way to do almost everything.				
Survey item 02: Practically every problem has a solution.				
Survey item 35: I feel relieved when an ambiguous situation suddenly becomes clear.				
Survey item 41: I find it hard to make a choice when the outcome is uncertain.				
Panel B: Short preference module (derived from consistent subjects)	116	0.324***	92	0.323*
		(< 0.001)		(0.001)
Thought experiment: Dynamic Ellsberg urn				
Survey item 01: There is a right way and a wrong way to do almost everything.				
Survey item 02: Practically every problem has a solution.				
Panel C: Single item module	121	0.219**	99	0.262***
		(0.016)		(0.009)
Thought experiment: Dynamic Ellsberg urn				

This table presents the recommended ambiguity preference survey modules. Panel A presents the 5-item module obtained from all subjects. Panel B presents the 3 -item module obtained from consistent subjects. Panel C presents the recommended single-item ambiguity preference measure using the dynamic Ellsberg thought experiment. The in-sample predictions present the correlation between the ambiguity preference benchmark (m) and the fitted values of the regression as presented in table 5. The out-of-sample predictions present the correlation between ambiguity preference benchmark (m) and the fitted values of the regression as presented in table 5 using the 2016 data sample. P-values are given in parentheses. *, **, and *** denote statistical significance at the 10%, 5% and 1% level, respectively. The full script of the module is available on the authors' personal websites and SSRN (Paper ID 3315413).

thought experiments with medium range probabilities. Further research may build on this result and introduce variations in the levels of ambiguity to measure secondary elements of ambiguity attitudes, like a-insensitivity, in large surveys.

4.3 Tests of the ambiguity preference module

4.3.1 In-sample tests

As a first assessment of the quality of the ambiguity preference module, we analyse the module's in-sample fit, i.e., the degree to which the module captures the variation in ambiguity preferences. To this end, we calculate the in-sample correlation between the predicted ambiguity preferences using the preference module and the ambiguity preference benchmark. The predicted ambiguity preferences are calculated as the fitted values of the regressions. Depending on the length of the preference module, the correlations range between 22% for the single-item module up to 41% for the 5-predictor module, see the second column of table 6. These correlations are sizable, and highly significant. They are also comparable to the magnitudes found in similar studies on risk preferences (Falk et al., 2016).

4.3.2 Out-of-sample tests

While the in-sample fit is an important criterion to assess a module's ability to capture the subjects' preferences, it is not free from limitations. Most importantly, adding more predictors to the module always improves the explanatory power in-sample, but can worsen the module's ability to predict preferences for another subject pool (over fitting). Hence, a more important test of any preference module is not the in-sample fit, but its accuracy in predicting preferences out of sample.

To analyse the predictive power of the ambiguity preference module out of sample, we replicate the incentivized tasks and the entire ambiguity preference module on a completely different pool of subjects. This second round of experiments was conducted in March 2016 at the ExpReSS Lab at Royal Holloway, University of London.[22] We call this the *second sample*.

With 99 subjects, the second sample is almost of equal size as the original data sample. Relative to the original sample, the average age of participants is significantly younger (18.3 years) and has a larger fraction of female participants (75%). The diversity in subject characteristics makes the out-of-sample results an even stricter test.

The procedure of the out-of-sample test is as follows. First, we predict ambiguity preferences

[22] The experimental procedure of this second round of experiments is identical to those described in section 2.

of the *second sample* of subjects using their answers to the items included in the preference modules identified in the item selection procedure (and listed in panel A and B in table 6). We use the coefficient estimates in table 5 to predict the second sample subjects' ambiguity preferences. Then, we correlate the predicted ambiguity preferences with the actual ambiguity preference benchmark elicited for the *second sample* of subjects.

The results are presented in the right column of table 6. Depending on the module, the out-of-sample correlations between predicted and actual ambiguity preferences are between 26% and 33%, and in most cases highly significant. Again, these correlations are comparable to those of risk preferences in Falk et al. (2016).

With the exception of the single-item module, the out-of-sample correlation is lower than the in-sample correlation. This is expected, since the module coefficients are obtained from a completely different data set. Yet, the dynamic Ellsberg urn thought experiment is even better in predicting ambiguity preferences out-of-sample than in-sample. This further underlines the reliability of this module item in measuring ambiguity preferences.

4.3.3 Test-retest

The in-sample and out-of-sample correlations are statistically significant. However, to judge whether their magnitudes are sizeable and relevant in economic terms, one needs a benchmark correlation (i.e. the maximum correlation that can be achieved). It is not reasonable to expect a 100% correlation between ambiguity preferences and the preference module because any measure of preferences has in practice some unavoidable measurement error.

This implies that even if the module measured ambiguity preferences equally well as the incentivized task, the correlation will be less than one because of measurement errors. What correlation can one reasonably expect? A reasonable benchmark is the one offered by a test-retest procedure, also adopted in Falk et al. (2016): let the same subject answer the same task twice. The benchmark correlation is the level of correlation between the first and second answer. The underlying idea is that when measuring ambiguity preferences for the same subject twice, one should obtain the maximum level of achievable correlation netted by the measurement error. This correlation is going to be a reasonable benchmark against which to assess in-sample and out-of-sample correlations.

We implement the test-retest procedure on the incentivized tasks using a sub-sample of 26 subjects who participated in the study twice, with a time lag of 5 months. This gives us two measures of ambiguity preferences elicited in an identical way. The test-retest correlation is obtained by correlating the first observation of the subjects' ambiguity preference benchmark with their second observation.

The test-retest correlation is 32.9%, with a p-value of 0.101, i.e., close to significant at the 10% level. The rank correlation is 38.6%, with a p-value of 0.059. This correlation is comparable to the test-retest of risk preferences in Falk et al. (2016). More importantly, the test-retest correlation is close to the in-sample and out-of-sample correlation between preference benchmark and ambiguity preference module (see table 6). Hence, the ambiguity preference module can be considered sufficiently reliable.

4.4 Ambiguity preference score

The ambiguity preference module is easily implementable in all kinds of surveys, including phone-based surveys. Answering the preference module takes less than 5 minutes. In this section, we give two suggestions to create a ready-to-use ambiguity preference score from the responses to the module for empirical applications. The section assumes that a researcher collects responses using the 'best' version of the module as presented in column (5) of table 5, panel A, and wants to combine the data in some ways to obtain a unified score (or index) of ambiguity preferences without having access to incentivized preference measures.

There are several ways to combine the survey module responses. The main challenge is to convert the survey questions (measured on Likert scales) into a unit of measurement comparable to the dynamic Ellsberg thought experiment, such that they can be aggregated. One way to do that is to use the coefficients in table 5 (panel A) to construct this conversion metric. Note that the coefficients associated with the survey questions are similar: they indicate, on average, a differential in ambiguity preferences of 0.045 based on the answers to the survey questions. This differential can be made comparable to a one unit change in the dynamic Ellsberg (one percentage point) by dividing this number by the coefficient of the Ellsberg thought experiment. The same applies to the constant. The resulting equivalence for each survey question is 20 units (rounded at the next integer) and 130 for the constant. Thus, if a subject is classified as ambiguity tolerant based on a survey question, she receives ± 20 points, according to the sign of the coefficient in table 5, 130 for the constant and the numerical value of her answer to the Ellsberg thought experiment (in percent).[23] A low ambiguity preference score corresponds to ambiguity-averse preferences; a high ambiguity score corresponds to ambiguity-seeking preferences.

We want to point out that these estimates are based on the median responses of our sample and we should not ignore the possibility that different samples behave differently. One way

[23]As an example, suppose a subject reveals in the dynamic Ellsberg urn experiment a matching probability of 37.5% and answers all 4 survey questions with values revealing ambiguity tolerant attitudes. Then the ambiguity preference score equals 37.5 +20 +20 +20 - 20 + 130 = 207.5.

to standardize the conversion is to attribute an equivalence to each value of the Likert scale. This method yields a score that has the advantage of being median response-invariant and comparable across different samples. To calculate this equivalence, we recover the coefficients associated with one unit change in the Likert scales by re-running the regression specification (5) in table 5 panel A treating the answers to the Likert scales as continuous variables. This approach also allows for more variation in ambiguity preferences across subjects. The average absolute value of the coefficients of the survey questions is 0.01767 and the constant is 0.350. Dividing these numbers by the coefficient of the Ellsberg thought experiment (0.0027) yields an equivalence of 7 points for each value of the Likert scale (and 130 for the constant). Again, a low ambiguity preference score corresponds to ambiguity-averse preferences; a high ambiguity score corresponds to ambiguity-seeking preferences. An ambiguity score of 292 corresponds to ambiguity neutrality. Table 7 presents the ambiguity module together with this conversion rule. Table 8 presents the descriptive statistics of the so-obtained ambiguity preference scores for both samples of subjects. Panel A of table 8 shows that the average and standard deviation of the ambiguity preference score are similar in both data samples: the average is around 262 points, with a standard deviation of about 19.

Panel B examines the correlations of the ambiguity preference score with the ambiguity preference benchmark and self-reported risk preferences using a single-item survey question following Dohmen et al. (2011), controlling for age and gender. The self-reported risk preferences are available only in the follow up dataset. As expected, the in-sample correlation with the ambiguity preference benchmark is substantial, and very close to the correlations between the fitted values and the ambiguity preference benchmark, see table 6. Again, the out-of-sample correlation in the second sample is smaller in magnitude but still strongly significant. Finally, table 8 shows that, as expected, the ambiguity preference score is not capturing self-reported risk preferences following Dohmen et al. (2011). This is in line with the theoretical prediction that risk and ambiguity preferences are not correlated, and gives confidence that, empirically, the ambiguity score is not picking up elements of risk preferences.

5 Conclusions

This paper proposes a novel and experimentally validated ambiguity preference module to measure individual ambiguity preferences. We use responses to a variety of thought experiments and attitudinal survey questions to predict individual ambiguity preferences obtained from a state-of-the-art incentivized experimental task. Applying an iterative selection procedure that

Table 7: Construction of the ambiguity preference score

Ambiguity preference module items	Conversion into ambiguity score						
Dynamic Ellsberg two-color urn thought experiment (see online appendix B)	Score equals risk equivalent (in percent) consistent with the subject's choices (see panel B of table 4)						
Attitudinal survey questions Please respond to the following statements by indicating the extent to which you agree or disagree with them on a scale from 1 (I strongly agree) to 7 (I strongly disagree).	Score depending on answer in the Likert scales:						
	1	2	3	4	5	6	7
Survey items for the 'optimal' module:							
- There is a right way and a wrong way to do almost everything.⋆	49	42	35	28	21	14	7
- Practically every problem has a solution.	7	14	21	28	35	42	49
- I feel relieved when an ambiguous situation suddenly becomes clear.	7	14	21	28	35	42	49
- I find it hard to make a choice when the outcome is uncertain.	49	42	35	28	21	14	7

This table presents the conversion of the ambiguity preference module (left column) into an ambiguity preference score (right column). First, the answers to the thought experiment and the survey questions are transformed into a score for each item. Second, the ambiguity preference score is obtained by adding all individual scores plus 130 (for the constant). Example: Suppose a subject reveals in the dynamic Ellsberg urn experiment a matching probability of 37.5% and answers to all 4 survey questions 1 ("I strongly agree"). Then the ambiguity preference score equals 37.5 +49 + 7 + 7 + 49 + 130 = 279.5
A low ambiguity preference score corresponds to ambiguity-averse preferences; a high ambiguity score corresponds to ambiguity-seeking preferences. The symbol ⋆ indicates that the survey item has reversed order (i.e. low values in the Likert scale are associated with high values in the preference benchmark).

Table 8: Ambiguity preference score

Panel A: Mean (standard deviation) ambiguity preference score

Data set	Original sample	Second sample
‘best’ module	262.74 (19.8)	269.51 (19.3)

Panel B: Correlations

Data set	Original sample	Second sample
Ambiguity preference benchmark	80.7094***	52.7275***
	(14.897)	(16.865)
Age	-0.0738	-0.2612
	(0.200)	(0.760)
Female	-2.9476	5.6460
	(3.233)	(3.674)
Risk preferences (Dohmen et al., 2011)		-1.3995
		(0.880)
constant	233.6830***	249.2658***
	(10.659)	(18.189)
R^2	0.204	0.115
N	121	99

Panel A reports the average and standard deviation (in parentheses) of the ambiguity preference scores. Panel B reports the correlation of the ambiguity preference score with the ambiguity preference benchmark and the risk preferences measured following Dohmen et al. (2011). P-values are given in parentheses. *, **, and *** denote statistical significance at the 10%, 5% and 1% level, respectively.

considers all possible combinations of predictors, the paper identifies one thought experiment and four attitudinal questions that are best in predicting ambiguity preferences. This set forms our ambiguity preference module.

The ambiguity preference module passes a number of validation tests, including in-sample and out-of-sample tests using two completely different subject pools. Taken together, these tests show that the ambiguity preference module allows the reliable measurement of ambiguity preferences for different samples of subjects. The ambiguity preference module thus combines the rigour of laboratory experiments with the convenience of survey questionnaires. It is simple, easy to implement, and cost-effective. It can be used to reliably measure ambiguity preferences when conducting incentivized laboratory experiments is infeasible. It therefore can be a valuable tool for empirical researchers interested in measuring ambiguity preferences when conducting large scale surveys or field studies, where it is often impractical to use incentivized decision tasks. The preference module might also be useful to experimental researchers with time and money constraints, who seek to include a simple control measure of ambiguity preferences in experimental studies.

By experimentally validating an ambiguity preference module to measure ambiguity preferences, these results give some confidence in the ability to elicit ambiguity preferences more widely and consistently. A validated and widely applicable measurement tool to quantify ambiguity preferences is an important and useful contribution to the literature because it may increase the comparability across future studies, thus improving our understanding of the effects of ambiguity preferences on economic behavior.

Finally, the ambiguity preference module may have direct practical applications, for example, in the financial service industry. While it is standard to use simple questionnaires to assess the clients' risk preferences, ambiguity preferences have been left unexplored because of the lack of simple assessment methods. By using the ambiguity preference module, private banks could improve the personality assessment of their clients to tailor better asset allocation strategies.

We do not want to claim that, in order to measure ambiguity preferences, the ambiguity preference module is always preferable over incentivized decision tasks. Yet, given the relevance of ambiguity preferences for economic decision-making, we argue that ambiguity preferences should be measured more often in empirical studies and we provide a toolkit to reliably do so.

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.

Bade, Sophie (2015), 'Randomization devices and the elicitation of ambiguity-averse preferences', *Journal of Economic Theory* **159**, 221–235.

Baillon, Aurélien and Han Bleichrodt (2015), 'Testing ambiguity models through the measurement of probabilities for gains and losses', *American Economic Journal: Microeconomics* **7**(2), 77–100.

Baillon, Aurélien, L. Cabantous and P.P. Wakker (2012), 'Aggregating imprecise or conflicting beliefs: An experimental investigation using modern ambiguity theories', *Journal of Risk and Uncertainty* **44**(2), 115–147.

Baillon, Aurélien, Zhenxing Huang, Asli Selin and Peter P. Wakker (2018), 'Measuring ambiguty attitudes for all (natural) events', *Econometrica* **86**(5), 1839–1858.

Becker, Gordon M., Morris H. DeGroot and Jacob Marschak (1964), 'Measuring utility by a single-response sequential method', *Behavioral Science* **9**(3), 226–232.

Bernard, Tanguy, Stefan Dercon, Kate Orkin, Alemayehu Seyoum Taffesse et al. (2014), The future in mind: Aspirations and forward-looking behaviour in rural ethiopia, *in* 'Centre for the Study of African Economies conference on economic development in Africa, Oxford, UK, March', Vol. 25.

Bianchi, Milo and Jean-Marc Tallon (2018), 'Ambiguity preferences and portfolio choice: Evidence from the field', *Management Science* . forthcoming.

Binmore, Ken, Lisa Stewart and Alex Voorhoeve (2012), 'How much ambiguity aversion? finding indifferences between ellsberg's risky and ambiguous bets', *Journal of Risk and Uncertainty* **45**(3), 215–238.

Bossaerts, Peter, Paolo Ghirardato, Serena Guarnaschelli and William R. Zame (2010), 'Ambiguity in asset markets: Theory and experiment', *Review of Financial Studies* **23**(4), 1325–1359.

Budner, S. (1962), 'Intolerance for ambiguity as a personal variable', *Journal of Personality* **30**, 29–50.

Butler, Jeffrey V., Luigi Guiso and Tullio Jappelli (2014), 'The role of intuition and reasoning in driving aversion to risk and ambiguity', *Theory and Decision* **77**, 455–484.

Camerer, Colin and Martin Weber (1992), 'Recent developments in modeling preferences', *Journal of Risk and Uncertainty* **5**(4), 325–370.

Carbone, Enrica, Xueqi Dong and John Hey (2017), 'Elicitation of preferences under ambiguity', *Journal of Risk and Uncertainty* **54**(2), 87–102.

Chakravarty, Sujoy and Jaideep Roy (2009), 'Recursive expected utility and the separation of attitudes towards risk and ambiguity: an experimental study', *Theory and Decision* **66**(3), 199–228.

Chang, Edward C., Albert Maydeu-Olivares and Thomas D'Zurilla (1997), 'Optimism and pessimism as partially independent constructs: relationship to positive and negative affectivity and psychological well-being', *Personality and Individual Differences* **23**(3), 433–440.

Charness, Gary, Edi Karni and Dan Levin (2013), 'Ambiguity attitudes and social interactions', *Journal of Risk and Uncertainty* **46**, 1–25.

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**, 538–567.

Chew, Soo Hong and Jacob S. Sagi (2006), 'Event exchangeability: Probabilistic sophistication without continuity or monotonicity', *Econometrica* **74**(3), 771–786.

Chow, Clare Chua and Rakesh K Sarin (2002), 'Known, unknown, and unknowable uncertainties', *Theory and Decision* **52**, 127–138.

Cohen, Michele, Jean-Marc Tallon and Jean-Christophe Vergnaud (2011), 'An experimental investigation of imprecision attitude and its relation with risk attitude and impatience', *Theory and Decision* **71**(1), 81–110.

Cohen, Michele, Jean-Yves Jaffray and Tanios Said (1987), 'Experimental comparison of individual behavior under risk and under uncertainty for gains and for losses', *Organizational Behavior and Human Decision Processes* **39**(1), 1–22.

Cole, Shawn, Xavier Giné, Jeremy Tobacman, Robert Townsend, Petia Topalova and James Vickery (2013), 'Barriers to household risk management: evidence from India', *American economic journal. Applied economics* **5**(1), 104.

Collard, Fabrice, Sujoy Mukerji, Kevin Sheppard and Jean-Marc Tallon (2018), 'Ambiguity and the historical equity premium', *Quantitative Economics* **9**, 945–993.

Curley, Shawn P. and J. Frank Yates (1989), 'An empirical evaluation of descriptive models of ambiguity reactions in choice situations', *Journal of Mathematical Psychology* **33**(4), 397–427.

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), 219–244.

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), 559–577.

Dimmock, Stephen G., Roy Kouwenberg and Peter P. Wakker (2016), 'Ambiguity attitudes in a large representative sample', *Management Science* **62**(5), 1363–1380.

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), 522–550.

Donkers, Bas, Bertrand Melenberg and Arthur Van Soest (2001), 'Estimating risk attitudes using lotteries: A large sample approach', *Journal of Risk and Uncertainty* **22**(2), 165–195.

Dow, J. and Da Costa S.R. Werlang (1992), 'Uncertainty aversion, risk aversion, and the optimal choice of portfolio', *Econometrica* **60**(1), 197–204.

Ellsberg, D. (1961), 'Risk, ambiguity and the savage axioms', *Quarterly Journal of Economics* **75**(4), 643–669.

Falk, Armin, Anke Becker, Thomas Dohmen, Benjamin Enke, David Huffman and Uwe Sunde (2018), 'Global evidence on economic preferences', *Quarterly Journal of Economics* **133**(4), 1645–1692.

Falk, Armin, Anke Becker, Thomas Dohmen, David Huffman and Uwe Sunde (2016), The preference survey module: A validated instrument for measuring risk, time and social preferences. working paper.

Falk, Armin and Florian Zimmermann (2017), 'Consistency as a signal of skills', *Management Science* **63**(7), 2197–2210.

Fehr, E., U. Fischbacher, B. Rosenbladt, J. Schupp and G. Wagner (2003), A nationwide laboratory: Examining trust and trustworthiness by integrating behavioural experiments into representative surveys. working paper.

Festinger, Leon (1957), *A Theory of Cognitive Dissonance*, Stanford University Press, Stanford, CA.

Fischbacher, Urs (2007), 'z-tree: Zurich toolbox for ready-made economic experiments', *Experimental Economics* **10**(2), 171–178.

Fox, Craig R. and Amos Tversky (1995), 'Ambiguity aversion and comparative ignorance', *Quarterly Journal of Economics* **110**(3), 585–603.

Furnham, Adrian and Tracy Ribchester (1995), 'Tolerance of ambiguity: A review of the concept, its measurement and applications', *Current Psychology* **14**(3), 179–199.

Greco, Veronica and Derek Roger (2001), 'Coping with uncertainty: the construction and validation of a new measure', *Personality and individual Differences* **31**, 519–534.

Guiso, Luigi and Monica Paiella (2008), 'Risk aversion, wealth and background risk', *Journal of the European Economic Association* **6**(6), 1109–1150.

Guiso, Luigi and Tullio Jappelli (2008), Financial literacy and portfolio diversification. working paper.

Heath, Chip and Amos Tversky (1991), 'Preference and belief: Ambiguity and competence in choice under uncertainty', *Journal of Risk and Uncertainty* **4**(1), 5–28.

Holt, Charles A. and Susan K. Laury (2002), 'Risk aversion and incentive effects', *American Economic Review* **92**(5), 1644–1655.

Kirton, M. J. (1981), 'A reanalysis of two scales of tolerance of ambiguity.', *Journal of Personality Assessment* **45**(4), 407–414.

Knight, Frank H. (1921), *Risk, Uncertainty, and Profit*, Houghton Mifflin Company, Boston, MA.

Kocher, Martin G., Amrei Marie Lahno and Stefan T. Trautmann (2015), Ambiguity aversion is the exception. Working paper.

Lauriola, Marco and Irwin P. Levin (2001), 'Relating individual differences in attitude toward ambiguity to risky choices', *Journal of Behavioral Decision Making* **14**(2), 107–122.

Mac Donald Jr., A. P. (1970), 'Revised scale for ambiguity tolerance: reliability and validity', *Psychological Reports* **26**, 791–798.

McLain, David L. (2009), ‘Evidence of the propoerties of an ambiguity tolerance measure: the multiple stimulus types ambiguity tolerance scale-ii (mstat-ii)’, *Psychological Reports* **105**(3), 975–988.

Nishimura, Kiyohiko G. and Hiroyuki Ozaki (2004), ‘Search and knightian uncertainty’, *Journal of Economic Theory* **119**(2), 299–333.

Nishimura, Kiyohiko G. and Hiroyuki Ozaki (2007), ‘Irreversible investment and knightian uncertainty’, *Journal of Economic Theory* **136**(1), 668–694.

Norton, Robert W. (1975), ‘Measurement of ambiguity tolerance’, *Journal of Personality Assessment* **39**(6), 607–619.

Robins, Richard W., Holly M. Hendin and Kali H. Trzesniewski (2001), ‘Measuring global self-esteem: Construct validation of a single-item measure and the rosenberg self-esteem scale’, *Personality and Social Bulletin* **27**(2), 151–161.

Rydell, S.T. and E. Rosen (1966), ‘Measurement and some correlates of need cognition’, *Psychological Reports* **19**, 139–165.

Sutter, Matthias, Martin G. Kocher, Daniela Glätzle-Rüetzler and Stefan T. Trautmann (2013), ‘Impatience and uncertainty: Experimental decisions predict adolescents’ field behavior’, *The American Economic Review* **103**(1), 510–531.

Tibshirani, Robert (1996), ‘Regression shrinkage and selection via the lasso’, *Journal of the Royal Statistical Society, Series B (Methodological)* **58**(1), 267–288.

Trautmann, Stefan T. and Gijs van de Kuilen (2016), *Blackwell Handbook of Judgment and Decision Making*, Blackwell, chapter Ambiguity Attitudes.

Vieider, Ferdinand M., Mathieu Lefebvre, Ranoua Bouchouicha, Thorsten Chmura, Rustamdjan Hakimov, Michal Krawczyk and Peter Martinsson (2015), ‘Common components of risk and uncertainty attitudes across contexts and domains: Evidence from 30 countries’, *Journal of the European Economic Association* **13**(3), 421–452.

Weitzman, Martin L. (2009), ‘On modeling and interpreting the economics of catastrophic climate change’, *Review of Economics and Statistics* **91**(1), 1–19.