UNIVERSITÀ di VERONA

Department of ECONOMICS

Working Paper Series
Department of Economics
University of Verona

Consistency of Risk Preference Measures and the Role of Ambiguity: An Artefactual Field Experiment from China

Pan He, Marcella Veronesi, Stefanie Engel

WP Number: 3 March 2016

ISSN: 2036-2919 (paper), 2036-4679 (online)

Consistency of Risk Preference Measures:

An Artefactual Field Experiment from Rural China

Citation:

He, P., M. Veronesi, and S. Engel (2017), "Consistency of Risk Preference Measures: An Artefactual Field Experiment from Rural China." ***Journal of Development Studies***

(doi: 10.1080/00220388.2017.1336542)

ttps://www.tandfonline.com/doi/abs/10.1080/00220388.2017.1336542

Pan He [a]

Marcella Veronesi *

Stefanie Engel [b]

[a] Institute of Environmental Engineering, ETH Zurich. Email: pan.he@ifu.baug.ethz.ch.

[b] Alexander von Humboldt Professorship for Environmental Economics, Institute for Environmental Systems Research, University of Osnabruck. Email: stefanie.engel@uos.de.

[*] Corresponding author: Department of Economics, University of Verona, and Center for Development and Cooperation (NADEL), ETH Zurich. Address for correspondence: Via Cantarane 24, 37129 Verona, Italy. Fax: +39 045 802 8529. Phone: +39 045 802 8095; Email: marcella.veronesi@univr.it.

Acknowledgements. We acknowledge the financial support of the China Scholarship Council and the North-South Centre of ETH Zurich. We thank James Andreoni, Michael Bennet, Michael Kuhn, Peter Martinsson, and Jintao Xu for their comments and suggestions.

Consistency of Risk Preference Measures:

An Artefactual Field Experiment from Rural China

Abstract

A variety of measures have been developed to elicit individual risk preferences. How these measures perform in the field, in particular in developing countries with non-student subjects, is still an open question. We implement an artefactual field experiment in rural China to investigate (i) consistency across incentivised experimental risk measures, (ii) consistency in risk preferences elicitation between non-incentivised survey measures and incentivised experiments, and (iii) possible explanations for risk preference inconsistency across measures. We find that inconsistent risk preferences across survey and experimental measures may be explained by ambiguity preferences. In the survey, subjects may mix risk and ambiguity preferences.

Keywords: risk preferences, ambiguity preferences, field experiments, socio-economic survey, China.

JEL Codes: C93, D81, O1

1. Introduction

Risk preferences play an important role in individual decisions and behaviours of strong relevance for development studies, such as investments, production decisions, and technology adoption.[1] Various measures have been developed to elicit individual risk preferences (Charness, Gneezy, and Imas, 2013). Some measures rely on simple survey questions on willingness to take risks in general or in specific domains, or questions on hypothetical gambles, lotteries, and investments to elicit subjects' risk attitudes.[2] Other measures are based on complex experimental designs with real monetary incentives, often developed and tested in the laboratory with educated students.[3]

The question whether survey measures can be used as a substitute for experimental measures is of particular relevance when conducting field research in a developing country setting. First, field research in developing countries, particularly in rural areas, tends to be more logistically difficult, so that conducting sound economic experiments can be difficult. Lack of computer labs, electricity and/or computer literacy often make it necessary to conduct pencil-and-paper versions, which in turn require simplified experimental protocols and smaller session sizes with the associated risk of less reliable experimental data.[4] Second, lower education levels may make it more difficult for respondents in developing countries to understand the more complex experimental setup. Cardenas and Carpenter (2008) emphasise that in developing countries 'Literacy is a large problem but it is numeracy that might be something that field experiments may stumble on and not be able to recover from.' (p. 329) By contrast, asking a risk survey question is simpler and less costly in terms of money and time. Survey measures, however, are not incentivised, which raises the concern on whether they can reveal accurate risk preferences (Charness *et al.*, 2013).

In this study, we investigate the consistency of risk preference measures in the context of a rural area of a developing country and with non-student subjects. In particular, we address three questions: first, whether subjects behave in a consistent manner across different incentivised experimental risk measures; second, whether there is consistency in risk preferences elicited through non-incentivised survey measures versus incentivised experiments; and third, possible explanations for risk preference inconsistency across measures.

Most previous studies focus on only one risk measure. Recently, some studies have compared different elicitation measures (see Table A1 of the appendix for a review). However, the wide majority of studies are based on experiments conducted in the computer laboratory in developed countries with students. A common criticism of such experiments is that students are not representative of non-student populations, and therefore laboratory experimental findings may be invalid in the field (Falk and Heckman, 2009). This concern is particularly valid for developing country contexts with less-educated subjects. Whether elicited preferences are consistent across measures in such a setting is still an open question.

In addition, the overall evidence on what measure is better in predicting risky choices in real life, and whether there is consistency in risk preferences elicited by using a simple survey question or risk experiments is mixed. Dohmen *et al.* (2011), Hardeweg, Menkhoff, and Waibel (2013), and Vieider *et al.* (2014) find that self-reported risk preferences elicited through the survey measure are significantly related to experimental results, and that the survey measure is a better predictor of risky behavior. Verschoor, D'Exelle, and Perez-Viana (2016) show that incentivised risk measures are predicting real world behavior in some domains (purchasing fertilizers) but not in others (growing cash crops). Other studies show low or no correlations between risk preferences elicited through the survey measure and the experimental measure (for example, Charness and Viceisza, 2016; Deck, Lee, Reyes, and Rosen, 2013; Ding, Hartog, and Sun, 2010; Eckel and Grossman, 2002, 2008; Lönnqvist, Verkasalo, Walkowitz, and Wichardt, 2015; Pennings and Smidts, 2000). Only two studies focus on non-student subjects in developing countries, and find contradictory results. Hardeweg *et al.* (2013) conduct a within-subject comparison with a large sample size (934 farmers in Thailand), and find that the risk preferences elicited through the survey measure are consistent with those elicited through the experiment and that the survey measure outperforms the experimental measure in predicting risky behavior. Charness and Viceisza (2016) use a between-subject comparison with a small sample size (91 farmers in Senegal), and find that subjects behave differently in two experimental measures and that the survey measure is unlikely to reveal accurate risk attitudes.

In this study, we do not investigate what measure is better in predicting risky behaviour but whether elicited risk preferences are consistent across incentivised experimental risk measures, and between survey and experimental measures. We contribute to the literature by implementing an artefactual field experiment on risk preferences with a large sample of non-student subjects (farmers) in a developing country context (rural China). We adopt a within-subject comparison of a risk measure based on a survey question on willingness to take risks in general (Dohmen *et al.*, 2011), and two incentivised experimental measures based on the Holt and Laury (2002) (henceforth HL) experiment, and the Andreoni and Harbaugh (2010) (henceforth AH) experiment.

We first compare elicited risk preferences at the aggregate level. Results from all the elicitation measures indicate that subjects are on average risk averse. Following a within-subject comparison, however, we find a large level of inconsistency between the two experimental measures, and between the survey and the experimental measures. We test some possible explanations for the inconsistency. Dave, Eckel, Johnson, and Rojas (2010) suggest that the inconsistency may be related to the difference of cognitive difficulties across elicitation measures. Our empirical results do not support this explanation in line with findings by Reynaud and Couture (2012). We also hypothesise another possible and previously unexplored explanation for the inconsistency, that is the role played by ambiguity preferences in the survey question. We conduct an ambiguity experiment developed by Lauriola and Levin (2001) to elicit subjects' ambiguity preferences. We find that the inconsistency between the survey and experimental measures may be related to the fact that the survey measure may reveal a mix of risk and ambiguity preferences instead of pure risk preferences.

The paper is organised as follows. Section 2 describes the experimental design and the measures used to elicit risk and ambiguity preferences. Section 3 presents and discusses the results. Section 4 concludes.

2. Experimental Design

In this section, we first describe the sampling and procedures of the experiments, and then, the experimental measures used to elicit risk and ambiguity preferences.

2.1 Sampling and Procedures

Our study was conducted in the Hubei Province of China in March and April of 2012 using a sample of farmers as part of a project on the adoption of biogas technology. We conducted two pre-tests on 20 farmers to test for the comprehension of the survey and the experimental instructions.[5] We then randomly selected 685 households in 12 villages, excluding the villages where we conducted the pre-tests to avoid contamination. Village leaders helped us inform the decision-maker of each household and persuade him/her to participate in the study. A show-up fee of 5 CNY was used to incentivise participation.[6] 597 farmers participated in the study, generating a response rate of about 87 per cent. We exclude from the analysis one subject that did not complete the experimental tasks. In the end, our final sample includes 596 farmers. Table 1 presents the descriptive statistics. The average farmer is 48 years old and has a middle school degree (nine years). The majority of farmers are male decision-makers (74%). About half of the respondents have worked off-farm in the last year. The average household consists of four members, owns 0.70 hectares of land, and has an annual income of CNY 25,000.

[TABLE 1 HERE]

Each subject was assigned an identification number to guarantee the anonymity, and faced first an experimental session and then, a survey section. At the beginning of the experimental session, each subject received a brief introduction of the tasks, and information for example, on the expected duration of the study, and how the earnings were delivered. After the introduction, subjects were allowed to decide whether to participate or leave.[7] The experimental session included three tasks eliciting subjects' risk and ambiguity preferences through the Holt and Laury (2002) experiment, the Andreoni and Harbaugh (2010) experiment, and an ambiguity experiment based on Ellsberg's two-color problem (Ellsberg, 1961). We used a within-subject design, which implies that the same subjects participated in all the three tasks. To control for order effects, the two risk preference tasks were conducted before the ambiguity preference task in half of the villages and the order was switched in the other half. The Holt and Laury experiment

(HL) was conducted before the Andreoni and Harbaugh experiment (AH) in half of the villages and the order was switched in the other half.

At the beginning of each task, subjects were explained the experiment in detail.[8] Subjects were told that one decision in each task would be randomly selected by the experimenter to decide their earnings. This random selection was implemented at the end of the experimental session after all decision sheets of all experiments were collected so that subjects should treat each decision equally, and to avoid any potential wealth effects. In addition, tests were conducted to ensure the comprehension of the tasks. For example, in the HL experiment, subjects were asked how much they could earn if their choice in the first paired lottery was Option 1 and the randomly drawn number was three out of 1-10. If participants gave a wrong answer, the experimenters explained the experiment and tested them again.[9] After all subjects finished one task, the next task began.

After the experimental session, all subjects had to complete a survey that included questions on individual and household characteristics such as age, education and household income, and a risk elicitation question described in the next sub-section. After completing the survey, one subject at the time was invited to another room to receive the total earnings of all the experimental tasks. The highest total possible earning of the three tasks was CNY 50, which is equivalent to the daily wage of a farmer working at a factory. The total duration of the study including the distribution of the payoffs was between 2.5 and 3 hours. Table S1 of Supplementary Materials reports the average time taken to complete each experiment by village.

2.2 Survey Question on Willingness to Take Risks

The survey measure of risk preferences is a risk preference elicitation question measured on a 5-point Likert scale: 'In general, how would you rate your willingness to take risks? (1 = very unwilling; 2 = unwilling; 3 = neutral; 4 = willing; 5 = very willing)' (see for example, Charness *et al.*, 2013). We chose a 5-point Likert scale following Weijters, Cabooter, and Schillewaert (2010). After comparing different Likert item formats, Weijters *et al.* (2010) recommend a fully labelled 5-point scale format for studies among a non-student population because this leads to less misresponses than a scale with more response categories. The use of a 5-point scale is especially recommended in the case of a general population with low levels of literacy and little experience with questionnaires as it could be in rural areas of developing countries.

We use this question to elicit subjects' self-reported willingness to take risks in general. Subjects choosing the Likert scale 1 (very unwilling to take risk) or 2 (unwilling to take risk) are classified as risk averse, subjects choosing the Likert scale 3 (neutral) are classified as risk neutral, and subjects choosing the Likert scale 4 (willing to take risk) or 5 (very willing to take risk) are classified as risk loving. This survey question is not incentivised, and it is included in the post-experiment survey. In order to mediate the potential effects of previous experiments, we place this survey question at the end of the questionnaire.

2.3 The Holt and Laury (2002) (HL) Task

The Holt and Laury (2002) (HL) method elicits risk preferences by asking subjects to make choices in 10 binary lotteries as shown in Table 2. In each binary lottery there are two options, Option 1 and Option 2, and subjects need to choose one of them. Each option has two outcomes, a higher outcome x_1 and a lower outcome x_2.[10] Outcomes in Option 1 have lower variations with respect to outcomes in Option 2. This implies that Option 1 is less risky than Option 2. The probabilities of receiving the higher outcomes in the two options are the same, and increase from 1/10 in the first lottery to 10/10 in the last lottery. At the beginning, subjects may choose the less risky option. As the probabilities of receiving the higher outcomes increase, subjects may switch to the riskier option in a certain lottery. This switching point reveals subjects' risk preferences. Risk neutral subjects would switch to the riskier option in the fifth lottery, risk loving subjects before the fifth lottery, whereas risk averse subjects after the fifth lottery. Based on the switching point, an interval of utility parameters r with a constant relative risk aversion (CRRA) utility function $u(x) = x^{1-r}/(1 - r)$ can be obtained (Holt and Laury, 2002), which indicates the risk preferences of subjects ($r < 0$: risk loving; $r = 0$: risk neutral; $r > 0$: risk aversion).

After subjects complete all choices, one lottery is randomly selected to decide subjects' earnings.[11] Since each lottery has the same chance of being chosen, subjects should reveal their true risk preferences in each lottery. After the lottery is selected, one number between 1 and 10 is randomly picked. Subjects' earning is the corresponding outcome in the preferred option. For instance, assume that lottery 3 is selected to decide subjects' earnings, and that in this lottery subjects choose Option 2. Option 2 corresponds to receiving CNY 20 if the number reads 1-3, and CNY 0.5 if the number reads 4-10. Assume that the number reads 6, then subjects obtain a payoff corresponding to CNY 0.5.

[TABLE 2 HERE]

2.4 The Andreoni and Harbaugh (2010) (AH) Task

In each choice set of the Andreoni and Harbaugh (2010) (AH) task, subjects face a number of combinations of an outcome x and a probability p of receiving x under a budget constraint:

$$r_1 p + r_2 x = m \tag{1}$$

where r_1 is the price of the probability p, r_2 is the price of the outcome x, and m is the experimental budget. Subjects need to choose one favourite combination in each choice set.

In the original laboratory experiment (Andreoni and Harbaugh, 2010), r_2 is set equal to 1 in all choice sets. The value of p ranges from 0 to m/r_1, with a unit of increase $\Delta p = 1/100$. This implies that (p, x) combinations range from $(0, m)$ to $(m/r_1, 0)$, generating $100*m/r_1 + 1$ combinations. When subjects change the probability p on a computer screen, they can immediately see the corresponding change of the outcome x. In a field experiment without computers, some adjustments have to be made. We simplify the AH method following the procedure by Andreoni, Kuhn, and Sprenger (2015). In each choice set, we decrease the number of combinations to seven. The (p, x) combinations are: $(0, m)$, $(m/6r_1, 5m/6)$, $(2m/6r_1, 4m/6)$, $(3m/6r_1, 3m/6)$, $(4m/6r_1, 2m/6)$, $(5m/6r_1, m/6)$, $(m/r_1, 0)$. Since the first and the last combinations mean that subjects definitely obtain zero payoffs, we only present the remaining five combinations in the task table (see Table S2 of Supplementary Materials). Table 3 presents an example of choice set. For instance, assume the subject chooses Option 4 in this choice set, then she/he has a 64/100 chance of receiving CNY 8. If this choice set is selected to decide the experimental payoffs, then one number between 1 and 100 is drawn. If this number is not higher than 64, then the subject obtains CNY 8; otherwise the subject obtains nothing.

[TABLE 3 HERE]

In this study, the AH task contains nine choice sets in total. Following Andreoni and Harbaugh (2010), r_2 is equal to one in all choice sets, but m and r_1 are different in each choice set as shown in Table 4. At the end of the task, one choice set is randomly chosen to determine subjects' earnings.

[TABLE 4 HERE]

As an alternative to expected utility theory, prospect theory suggests that the probability is transformed by a weighting function to affect the value of the prospect (Kahneman and Tversky, 1979; Liebenehm and Waibel, 2014). One advantage of the AH method is that it allows to test whether the independence axiom or probability weighting better fits the experimental data (Andreoni and Harbaugh, 2010). If the independence axiom holds, the chosen outcome x is decided only by the budget m and the price of outcome r_2, and is not affected by the change in the price of probability r_1. If probability weighting holds, x is positively affected by r_1. When we regress x on r_1 with fixed effects for m/r_2, a positive coefficient on r_1 indicates that probability weighting holds while a zero coefficient indicates that the independence axiom holds. Andreoni and Harbaugh (2010) perform this test on the aggregate data and find a significant and negative coefficient, which rejects both the independence axiom and probability weighting. They argue that the change in r_1 causes only a small change in x and the small elasticity suggests an incomplete rejection of the independence axiom. Using our field experimental data, at the aggregate level we also obtain a significant and negative coefficient (-0.005), confirming that the independence axiom is more supported than probability weighting, and the assumption of a CRRA utility is reasonable.

As shown in Andreoni and Harbaugh (2010), subjects choose the combination that maximises the utility:

$$U(p, x) = px^{\alpha} \quad (2)$$

where α is the utility parameter indicating risk preferences ($\alpha < 1$: risk aversion; $\alpha = 1$: risk neutral; $\alpha > 1$: risk loving). Solving the problem of maximizing (2) subject to (1), we can obtain

$$x = \alpha \times \frac{r_1 p}{r_2}. \quad (3)$$

For the purpose of comparing the AH task to the HL task, in this study we report and use the CRRA coefficient $r = 1 - \alpha$. The parameters α can be obtained by estimating (3) using ordinary least squares (OLS).

2.5 The Ambiguity Preferences Task

Most measures of ambiguity preferences are developed based on Ellsberg's two-colour problem (Ellsberg, 1961). In these measures, subjects are asked to make choices between a risky option and an ambiguity option. In the experiment proposed by Lauriola and Levin (2001), subjects are presented with 41 choice questions. Both risk and ambiguity options have two outcomes: receiving a payment or nothing. Subjects know the probabilities of the outcomes in the risky options, but do not know the probabilities in the ambiguity options. The probabilities of receiving outcomes in the risky options change in steps ($\Delta p = 0.025$). In our field experiment, we set the probability change of the risky option as $\Delta p = 0.1$ to make a shorter list feasible for the implementation in the field.

Table 5 shows the 11 pairs of risky and ambiguity options. The possible payoffs in the risky option and the ambiguity option are the same. The probability of receiving the payoff in the risky option increases from 0/10 in the first row to 10/10 in the last row. The probability of receiving the payoff in the ambiguity option, however, is unknown (marked as '?') and will be randomly determined at the end of the task by drawing one number out of 0-10. This implies that the probability in the ambiguity option has a uniform distribution with an expected center of 5/10 (Lauriola and Levin, 2001). In the first row, subjects may choose the ambiguity option. As the probability in the risky option increases, subjects may switch to the risky option at a certain row. This switching point shows subjects' ambiguity preferences.

Ambiguity averse subjects switch before the seventh row, whereas ambiguity loving subjects switch at the seventh row or after. One shortage of this method is that it cannot distinguish ambiguity neutrality from small levels of ambiguity loving and ambiguity aversion. Therefore, we follow the standard method (see for example, Kahn and Sarin, 1988; Lauriola and Levin, 2001) and classify subjects into ambiguity averse and ambiguity loving subjects ignoring ambiguity neutral subjects.

[TABLE 5 HERE]

3. Results

In this section, we first report results from aggregate data collected using different risk elicitation measures (section 3.1). Second, we conduct a within-subject comparison of different risk elicitation measures (section 3.2). Third, we examine some possible reasons for risk preference inconsistency across measures (section 3.3).

3.1 Risk Preferences Elicited Through Different Measures at the Aggregate Level

Figure 1 displays the distribution of subjects' willingness to take risks in general, which is elicited using the survey question described in section 2.2 on a Likert scale ranging from one to five. About 27 per cent of subjects report the midpoint of the scale, and about 8 per cent of subjects choose the two extreme points.[12] The shares of risk averse, risk neutral, and risk loving subjects are 40 per cent, 27 per cent, and 33 per cent, respectively.

[FIGURE 1 HERE]

In the HL experiment, 552 subjects (92%) report a unique switching point that allows us to generate an interval of constant relative risk aversion (CRRA) coefficient *r*. Figure 2 shows the distribution of CRRA coefficient *r*. The distribution follows a pattern similar to that presented in the original Holt and Laury (2002) study. The majority (68%) of subjects appear to be risk averse, about 18 per cent risk neutral, and the remaining 14 per cent risk loving.

[FIGURE 2 HERE]

Figure 3 presents the distribution of CRRA coefficient *r* estimated in the AH experiment. The figure shows a considerable variation of the values of *r*. Most of the values are in the range (-1, 1), however, some extreme values can be observed suggesting that those subjects are highly risk loving. A relatively large proportion (55%) of subjects are risk averse in the experiment while the percentages of risk neutral and risk loving subjects are smaller, 7 per cent and 38 per cent, respectively.

[FIGURE 3 HERE]

Table 6 summarises the percentages of risk averse, risk neutral, and risk loving subjects elicited using the survey question, the HL experiment, and the AH experiment at the aggregate level. All three measures reveal that risk averse subjects are the majority, which is consistent with results from other studies (Akay, Martinsson, Medhin, and Trautmann, 2012 in rural Ethiopia; Carlsson, Martinsoon, Quin, and Sutter, 2013, and Liu, 2013 in rural China; Holt and Laury, 2002 in U.S. with students). The percentages of risk averse subjects elicited using the survey question and the AH experiment are smaller than those elicited in the HL experiment. In particular, 40 per cent of subjects report that they are risk averse in the survey compared to 68 per cent in the HL experiment and 55 per cent in the AH experiment. In addition, the survey and AH measures reveal larger fractions of risk loving subjects (33% and 38%) compared to the HL measure (15%).

[TABLE 6 HERE]

3.2 Within-Subject Comparison of Risk Preferences across Measures

We first perform a within-subject comparison of the two incentivised experimental measures (HL and AH experiments). Then, we investigate how many subjects have CRRA coefficient *r* in the same range in the two experiments.[13] Figure 4 presents the number of subjects in each combination of CRRA coefficient intervals in the HL and AH experiments. The size of the point indicates the number of subjects. The shape of the point indicates different types of risk preference consistency. The triangle indicates that CRRA coefficients are in the same interval. Only 18 per cent of subjects are in this group. The square presents subjects without the same CRRA coefficient interval but with the same classified risk attitudes (risk aversion, risk neutrality, risk loving). This group contains 21 per cent of subjects. A large proportion (61%) of subjects (indicated by the solid circle) shows different risk attitudes in the two experiments. The ranked correlation between CRRA coefficients estimated in the HL and AH experiment is 0.049. The correlation is low and insignificant. In summary, subjects' estimated risk preferences are not stable across experiments as found also by the majority of previous studies comparing different incentivised experimental measures (for example, Berg, Dickhaut, and McCabe, 2005; Dulleck, Fooken, and Fell, 2015; Isaac and James, 2000).

[FIGURE 4 HERE]

We then compare the non-incentivised survey measure with the incentivised experimental measures. Table 7 shows the percentage of subjects in each combination of classified risk preferences between the survey and the HL experiment, and between the survey and the AH experiment. Only 34 per cent of subjects are in the same risk preference category across the survey and the HL experiment. Similarly, only 35 per cent of subjects can be grouped into the same risk preference category across the survey and the AH experiment. The ranked correlation between risk preferences elicited through the survey and the HL experiment is weakly significant (at the 10% statistical level), and the correlation is low and negative (ρ = -0.072). The ranked correlation between risk preferences elicited through the survey and the AH experiment is also negative (ρ = -0.044) and insignificant.

[TABLE 7 HERE]

3.3 Possible Explanations for Risk Preference Inconsistency across Measures

We have shown that the different measures yield to different categorization of risk preferences. Few studies investigate why different measures reveal inconsistent risk preferences. One potential reason could be that elicitation measures differ in cognitive difficulty, and subjects have different levels of cognitive skills (Anderson and Mellor, 2009; Dave *et al.*, 2010; Reynaud and Couture, 2012). We test this explanation following Reynaud and Couture (2012) and using education years as a proxy for cognitive skills to predict the consistency of risk preferences across measures. We control for some other individual characteristics such as age, gender, household income, household size, and land area.[14] We include village fixed effects and cluster the standard errors at the village level. Table 8 displays the marginal effects of a probit model where the dependent variable is a dummy variable equal to one if subjects have same classified risk preferences (risk aversion, risk neutrality, and risk loving) across measures, zero otherwise. We find that education has an insignificant effect on the consistency between survey and experimental risk measures and between the two experimental risk measures. This result is in contrast with findings by Dave *et al.* (2010) from Canada while it confirms previous findings by Reynaud and Couture (2012) from rural France, which suggests that risk preference inconsistency may not be explained simply by differences in cognitive difficulties across measures. However, we also recognise that it also depends on whether years of education can fully reflect cognitive skills, which are multidimensional (see Carroll, 1993 for discussions). We leave for future research to elicit subjects' cognitive abilities by using formal test scores.

[TABLE 8 HERE]

In our study, we examine another possible and previously unexplored reason for the inconsistency between the survey and experimental measures: the survey question might elicit a mix of risk and ambiguity preferences. The probability of outcomes in risky tasks is known, while the probability of outcomes in ambiguity tasks is unknown. Experimental designs allow distinguishing between risk and ambiguity preferences. In a survey question, instead, where subjects give a self-assessment of their willingness to take risks in general, it is difficult to distinguish between risk and ambiguity. When subjects' risk and ambiguity preferences differ, it may happen that self-reported willingness to take risks is different from risk preferences elicited by way of the experimental measures. The shift of risk preferences from the experimental measure to the survey measure can be categorised into two opposite directions. Direction 1 is towards risk loving, which includes risk aversion to risk loving, risk aversion to risk neutrality, and risk neutrality to risk loving. If our hypothesis of subjects mixing risk and ambiguity preferences in the survey holds, we would expect to observe such shifts more often when subjects are ambiguity loving. Direction 2 is towards risk aversion, which includes risk loving to risk aversion, risk loving to risk neutrality, and risk neutrality to risk aversion. Such shifts are expected to be more common among subjects that are ambiguity averse.

We use the ambiguity experiment described in section 2.5 to elicit subjects' ambiguity preferences. We find that 538 subjects (90%) have a unique switching point so that we are able to classify their ambiguity attitudes into ambiguity aversion and ambiguity loving. Figure 5 shows the distribution of switching row in the ambiguity experiment. We first examine the 343 subjects with inconsistent classified risk preferences across the HL experiment and the survey. Among these subjects, 195 are ambiguity averse (about 57%) and 148 (about 43%) are ambiguity loving. In addition, 95 subjects (about 28%) shift towards risk aversion in the survey and 248 (about 72%) towards risk loving. As expected, the proportion of ambiguity averse (loving) subjects shifting towards risk aversion (loving) in the survey is significantly larger (at the 5% statistical level) than the proportion of ambiguity loving (averse) subjects shifting towards risk aversion (loving): about 31 per cent vs 23 per cent in the case of the shift towards risk aversion, and 77 per cent vs 69 per cent in the case of the shift towards risk loving. This result is also confirmed by the comparison between the risk preferences elicited in the AH experiment and the survey where 353 subjects exhibit inconsistent classified risk preferences. Among these subjects, 204 are ambiguity averse (about 58%) and 149 are ambiguity loving (about 42%). The proportion of ambiguity averse (loving) subjects shifting towards risk aversion (loving) in the survey is about 49 per cent (60%)

while the proportion of ambiguity loving (averse) subjects shifting towards risk aversion (loving) is about 40 per cent (51%). These proportions are significantly different at the 5 per cent statistical level in line with our expectations that subjects may mix ambiguity and risk preferences in the survey question.

[FIGURE 5 HERE]

In addition, we formally test for the effect of ambiguity preferences on risk preference inconsistency by estimating a probit model on inconsistent subjects. The dependent variable is a dummy variable equal to one if subjects shift towards risk aversion (direction 2), and zero if subjects shift towards risk loving (direction 1). The independent variable is a dummy variable equal to one if the subject is ambiguity averse and zero otherwise.[15] Our hypothesis is that if the risk survey measure elicits a mix of risk and ambiguity preferences, but risk experimental measure elicits pure risk preferences, then ambiguity preferences may affect the consistency between risk preferences elicited through the survey and experimental measures. For example, if a subject is risk loving but ambiguity averse, she might be elicited as a risk averse person through the survey measure and as a risk loving person in the risk experiment. Therefore, the subject shifts from risk loving in the experiment to risk aversion in the survey. The results of Table 9 support our hypothesis. Table 9 shows that compared to ambiguity loving subjects, ambiguity averse subjects are significantly more likely to make the shift from risk loving in both HL and AH experiments towards risk aversion in the survey. This suggests that inconsistent risk preferences across the survey measure and experimental measures may be explained by ambiguity preferences, and that subjects seem to mix risk preferences and ambiguity preferences in the survey.

[TABLE 9 HERE]

4. Conclusions

This paper contributes to the ongoing discussion on the consistency of risk preference elicitation measures by comparing three different elicitation measures on a large sample of non-student subjects

(Chinese farmers) in a developing country using a within-subject design. Due to the lack of field evidence, especially from developing countries with non-student subjects, our study contributes to the literature by expanding laboratory findings on this topic to a broader population, and by providing an alternative potential explanation on the inconsistency across risk measures not previously investigated. The measures used to elicit risk preferences in this study include a non-incentivised survey question and two incentivised experimental measures proposed by Holt and Laury (2002) and Andreoni and Harbaugh (2010).

We examine both the consistency between the two experimental measures and the consistency between the survey and experimental measures. At the aggregate level, all elicitation measures show that the largest proportion of subjects are risk averse, although the percentages of risk averse subjects differ depending on the risk measure. Substantial within-subject inconsistencies of elicited risk preferences are observed between the two experiments. This result is in line with most previous studies finding that subjects' risk preference classification often varies across experiments (see Table A1 of the appendix). We also find within-subject inconsistencies of elicited risk preferences between the survey and experimental measures.

We investigate possible underlying reasons of these inconsistencies. First, we test education years as a proxy for cognitive ability. The effect, however, is not significant. Then, we propose and test a previously unexplored explanation of the inconsistency between the survey and experimental measures: subjects may not distinguish between risk and ambiguity when they give a self-assessment of their willingness to take risks in the survey. Our findings support this explanation: the survey question seems to elicit a mix of risk and ambiguity preferences. We recognise that our analysis is only a first step in analyzing the role of ambiguity as an underlying reason of the inconsistency across survey and experimental risk measures. Further research on this issue would be warranted, including a more refined and in-depth measurement approach since the probability of an event may be more ambiguous for some respondents than others.

Given our findings on the inconsistency of risk preference elicitation measures, further research is also needed to investigate the predictive power of the different measures for real-life decisions and the underlying reasons of this inconsistency in order to develop more suitable measures of risk preferences. In addition, we recognise that using education as a proxy for cognitive skills does not fully capture the multidimensional aspects of such skills. In psychology, there is a distinction between crystallised intelligence, and fluid intelligence (see for example, Nisbett *et al.*, 2012). Crystallised intelligence refers to the knowledge acquired by an individual, and it is captured by ad hoc designed achievement tests (Dohmen *et al.*, 2010; Heckman, Stixrud, and Urzua, 2006; Roberts *et al.*, 2000). Fluid intelligence refers to the rate at which individuals learn, and it is measured by IQ tests such as Raven's progressive matrices (1962). We leave it for future research to formally test for cognitive skills as a potential underlying reason of the inconsistency across risk measures.

References

Akay, A., Martinsson, P., Medhin, H., and Trautmann, S. T. (2012). Attitudes toward uncertainty among the poor: an experiment in rural Ethiopia. *Theory and Decision*, 73, 453-464.

Anderson, L. and Mellor, J. (2009). Are risk preferences stable? Comparing an experimental measure with a validated survey-based measure. *Journal of Risk and Uncertainty*, 39, 137-160.

Andreoni, J. and Harbaugh, W. (2010). Unexpected utility: Experimental tests of five key questions about preferences over risk. University of Oregon Economics Department Working Papers No. 2010-14.

Andreoni, J., Kuhn, M. A., and Sprenger, C. (2015). Measuring time preferences: A comparison of experimental methods. *Journal of Economic Behaviour and Organization*, 116, 451-464.

Barham, B. L., Chavas, J.-P., Fitz, D., Salas, V. R., and Schechter, L. (2014). The roles of risk and ambiguity in technology adoption. *Journal of Economic Behaviour and Organization*, 97, 204-218.

Berg, J., Dickhaut, J., and McCabe, K. (2005). Risk preference instability across institutions: A dilemma. *Proceedings of the National Academy of Sciences of the United States of America*, 102, 4209-4214.

Bruner, D. M. (2009). "Changing the Probability versus Changing the Reward." *Experimental Economics*, 12, 367-385.

Blais, A.-R. and Weber, E. U. (2006). A Domain-Specific Risk-Taking (DOSPERT) scale for adult populations. *Judgment and Decision Making*, 1, 33-47.

Cardenas, J. C. and J. Carpenter (2008). Behavioural development economics: Lessons from field labs in the developing world. *Journal of Development Studies*, 44, 311-338.

Carlsson, F., Martinsson, P., Qin, P., and Sutter, M. (2013). The influence of spouses on household decision making under risk: An experiment in rural China. *Experimental Economics*, 16, 383-401.

Carroll, J. B. (1993). *Human Cognitive Abilities: A Survey of Factor-Analytic Studies*. New York, NY: Cambridge University Press.

Charness, G., Gneezy, U., and Imas, A. (2013). Experimental methods: Eliciting risk preferences. *Journal of Economic Behaviour and Organization*, 87, 43-51.

Charness, G. and Viceisza, A. (2016). Three risk-elicitation methods in the field: evidence from rural Senegal. *Review of Behavioral Economics*, 3, 145-171.

Crosetto, P. and Filippin, A. (2013). The 'bomb' risk elicitation task. *Journal of Risk and Uncertainty*, 47, 31-65.

Crosetto, P. and Filippin, A. (2016). "A Theoretical and Experimental Appraisal of Four Risk Elicitation Methods." *Experimental Economics*, 19, 613-641.

Dave, C., Eckel, C. C., Johnson, C. A., and Rojas, C. (2010). Eliciting risk preferences: When is simple better? *Journal of Risk and Uncertainty*, 41, 219-243.

Deck, C., Lee, J., and Reyes, J. (2008). Risk attitudes in large stake gambles: Evidence from a game show. *Applied Economics*, 40, 41-52.

Deck, C., Lee, J., Reyes, J. A., and Rosen, C. C. (2013). A failed attempt to explain within subject variation in risk taking behavior using domain specific risk attitudes. *Journal of Economic Behavior and Organization*, 87, 1-24.

Ding, X., Hartog, J., and Sun, Y. (2010). Can we measure individual risk attitudes in a survey? IZA Discussion Paper No. 4807.

Dohmen, T., Falk, A., Huffman, D., and Sunde, U. (2010). Are risk aversion and impatience related to cognitive ability? *The American Economic Review*, 100, 1238-1260.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., and Wagner, G. G. (2011). Individual risk attitudes: Measurement, determinants, and behavioural consequences. *Journal of the European Economic Association*, 9, 522-550.

Dulleck, U., Fooken, J., and Fell, J. (2015). Within-subject intra- and inter-method consistency of two experimental risk attitude elicitation methods. *German Economic Review*, 16, 104-121.

Eckel, C. C. and Grossman, P. J. (2002). Sex differences and statistical stereotyping in attitudes toward financial risk. *Evolution and Human Behaviour*, 23, 281-295.

Eckel, C. C. and Grossman, P. J. (2008). Forecasting risk attitudes: An experimental study using actual and forecast gamble choices. *Journal of Economic Behaviour and Organization*, 68, 1-17.

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

Falk, A. and Heckman, J. J. (2009). Lab experiments are a major source of knowledge in the social sciences. *Science*, 326, 535-538.

Gneezy, U. and Potters, J. (1997). An experiment on risk taking and evaluation periods. *The Quarterly Journal of Economics*, 112, 631-645.

Harbaugh, W. T., Krause, K., and Vesterlund, L. (2010). The fourfold pattern of risk attitudes in choice and pricing tasks. *The Economic Journal*, 120, 595-611.

Hardeweg, B., Menkhoff, L., and Waibel, H. (2013). Experimentally validated survey evidence on individual risk attitudes in rural Thailand. *Economic Development and Cultural Change*, 61, 859-888.

Harrison, G. W. (1990). "Risk Attitudes in First-Price Auction Experiments: A Bayesian Analysis." *The Review of Economics and Statistics*, 72, 541-546.

He, H., Martinsson, P., and Sutter, M. (2012). Group decision making under risk: An experiment with student couples. *Economics Letters*, 117, 691-693.

Heckman, J. J., Stixrud, J., and Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behaviour. *Journal of Labor Economics*, 24, 411–482.

Hey, J., Morone, A., and Schmidt, U. (2009). Noise and bias in eliciting preferences. *Journal of Risk and Uncertainty*, 39, 213-235.

Hill, R. V. (2009). Using stated preferences and beliefs to identify the impact of risk on poor households. *Journal of Development Studies*, 45, 151-171.

Holt, C. A. and Laury, S. K. (2002). Risk aversion and incentive effects. *The American Economic Review*, 92, 1644-1655.

Isaac, R. M. and James, D. (2000). Just who are you calling risk averse? *Journal of Risk and Uncertainty*, 20, 177-187.

Kahn, B. E. and Sarin, R. K. (1988). Modeling ambiguity in decisions under uncertainty. *Journal of Consumer Research*, 15, 265-272.

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47, 263-291.

Kruse, J. B. and Thompson, M. A. (2003). "Valuing Low Probability Risk: Survey and Experimental Evidence." *Journal of Economic Behavior and Organization*, 50, 495-505.

Lauriola, M. and Levin, I. P. (2001). Relating individual differences in attitude toward ambiguity to risky choices. *Journal of Behavioural Decision Making*, 14, 107-122.

Lejuez, C. W., Read, J. P., Kahler, C. W., Richards, J. B., Ramsey, S. E., Stuart, G. L., Brown, R. A. (2002). Evaluation of a behavioural measure of risk taking: The Balloon Analogue Risk Task (BART). *Journal of Experimental Psychology: Applied*, 8, 75-84.

Liebenehm, S. and Waibel, H. (2014). Simultaneous estimation of risk and time preferences among small-scale cattle farmers in West Africa. *American Journal of Agricultural Economics*, 96, 1420-1438.

Liu, E. M. (2013). Time to change what to sow: Risk preferences and technology adoption decisions of cotton farmers in China. *Review of Economics and Statistics*, 95, 1386-1403.

Liu, E. M and Huang, J. (2013). Risk preferences and pesticide use by cotton farmers in China. *Journal of Development Economics*, 103, 202–215.

Lönnqvist, J.-E., Verkasalo, M., Walkowitz, G., and Wichardt, P. (2015). Measuring individual risk attitudes in the lab: Task or ask? An empirical comparison. *Journal of Economic Behaviour and Organization*, 119, 254-266.

Nisbett, R. E., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D. F., and Turkheimer, E. (2012). Intelligence: New findings and theoretical developments. *American Psychologist*, 67, 130–159.

Pennings, J. M. and Smidts, A. (2000). Assessing the construct validity of risk attitude. *Management Science*, 46, 1337-1348.

Raven, J. C. (1962). *Advanced progressive matrices: Sets I and II* (Revised ed.). London: H.K. Lewis.

Reynaud, A. and Couture, S. (2012). Stability of risk preference measures: Results from a field experiment on French farmers. *Theory and Decision*, 73, 203-221.

Roberts, R. D., Goff, G. N., Anjoul, F., Kyllonen, P. C., Pallier, G., and Stankov, L. (2000). The Armed Services Vocational Aptitude Battery (ASVAB), Little more than acculturated learning (Gc)!? *Learning and Individual Differences*, 12, 81–103.

Tanaka, T., Camerer, C. F., and Nguyen, Q. (2010). Risk and time preferences: Linking experimental and household survey data from Vietnam. *The American Economic Review*, 100, 557-571.

Vieider, F. M., Lefebvre, M., Bouchouicha, R., Chmura, T., Hakimov, R., Krawczyk, M., and Martinsson, P. (2015). Common components of risk and uncertainty attitudes across contexts and domains: Evidence from 30 countries. *Journal of the European Economic Association*, 13, 421-452.

Ward, P. S. and Singh, V. (2015). Using field experiments to elicit risk and ambiguity preferences: Behavioural factors and the adoption of new agricultural technologies in rural India. *Journal of Development Studies*, 51, 707-724.

Weijters, B., Cabooter, E. and Schillewaert, N. (2010). The effect of rating scale format on response styles: the number of response categories and response category labels. *International Journal of Research in Marketing*, 27, 236-247.

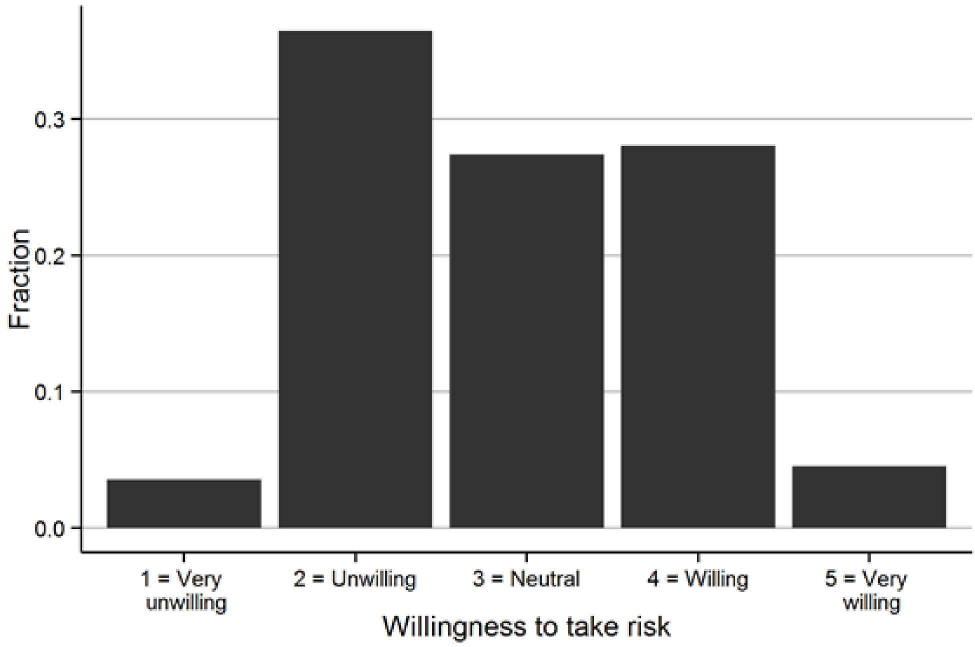

Figure 1. Self-Reported Willingness to Take Risks

Notes: The self-reported willingness to take risks is elicited by a risk preference question measured on a 5-point scale: 'In general, how would you rate your willingness to take risks? (1 = very unwilling; 2 = unwilling; 3 = neutral; 4 = willing; 5 = very willing).'

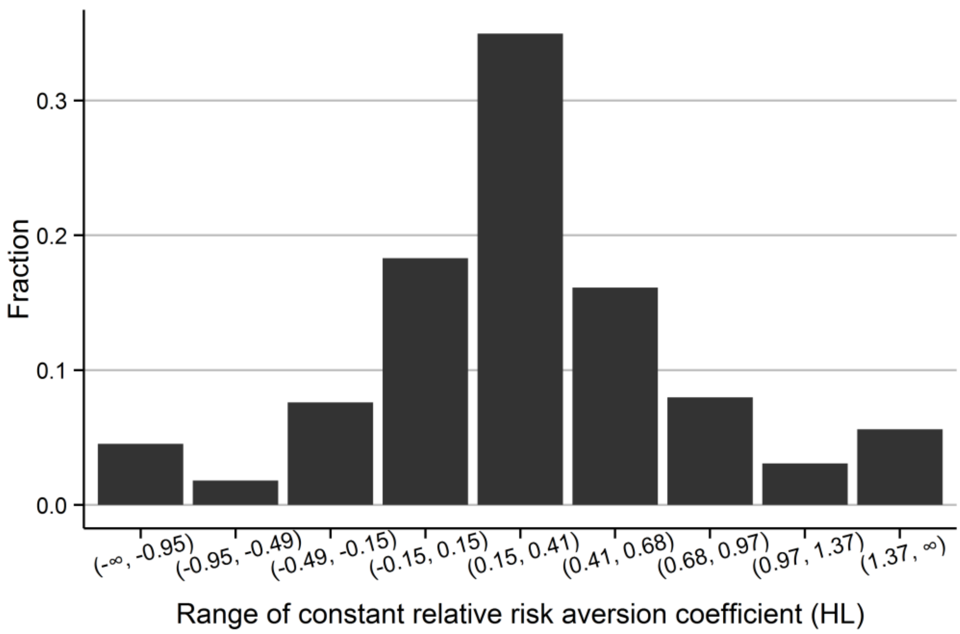

Figure 2. Constant Relative Risk Aversion Coefficient r in the Holt and Laury (HL) Experiment

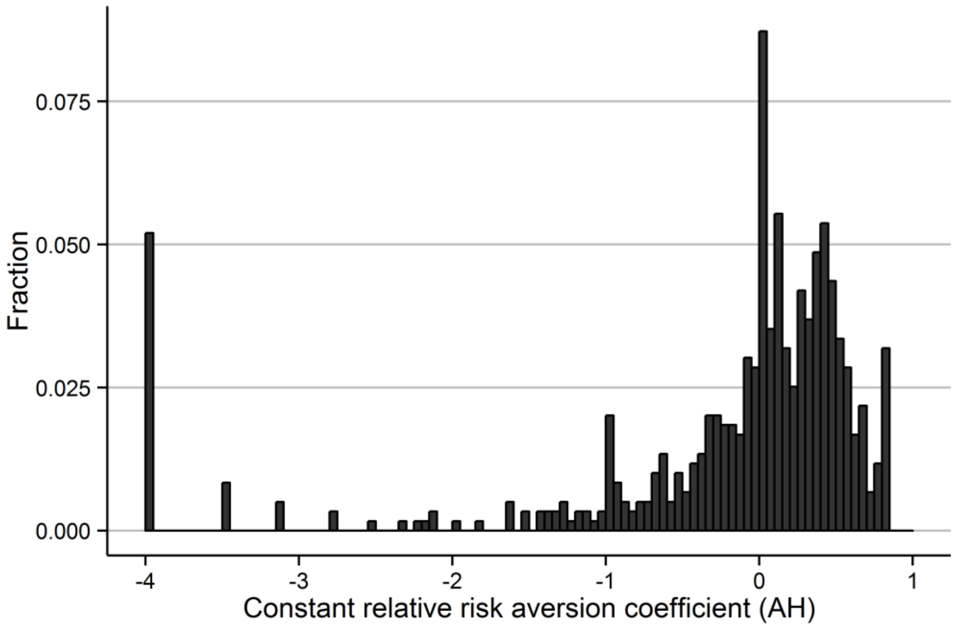

Figure 3. Constant Relative Risk Aversion Coefficient r in the Andreoni and Harbaugh (AH) Experiment

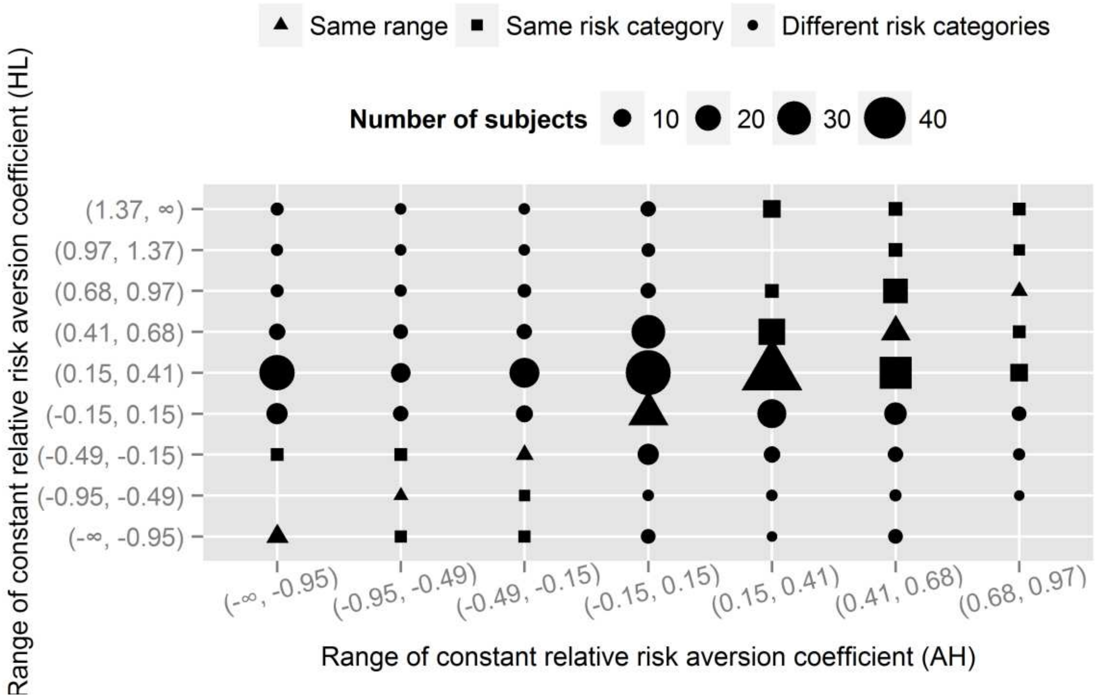

Figure 4. Within-subject Comparison between Constant Relative Risk Aversion Coefficients Estimated in the Holt and Laury (HL) and Andreoni and Harbaugh (AH) Experiments

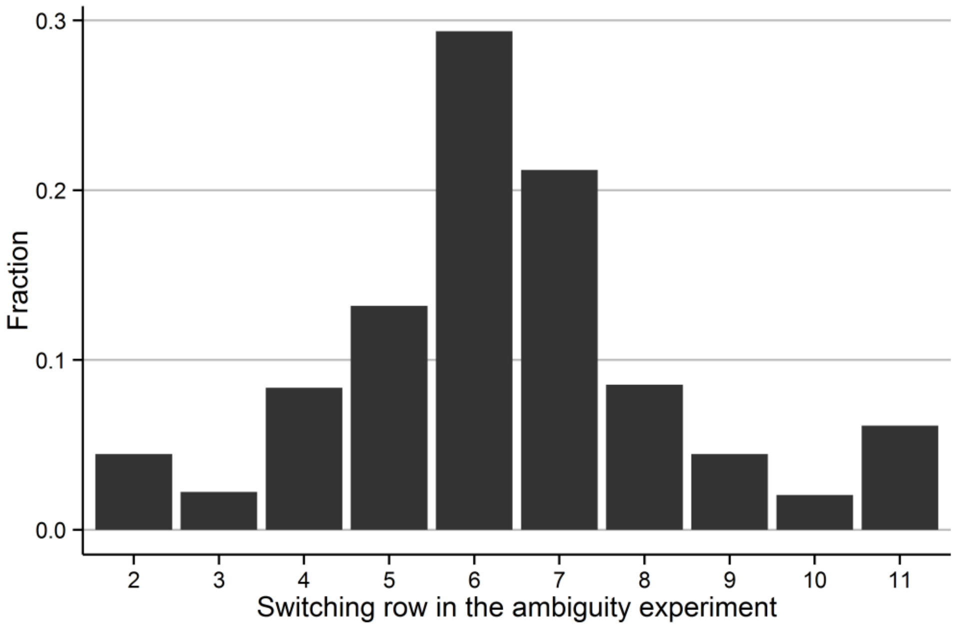

Figure 5. Fraction of Subjects Switching Row in the Ambiguity Experiment

Table 1. Descriptive Statistics

	Mean	S.D.
General willingness to take risks	2.936	0.983
Switching row in HL risk experiment	6.067	1.741
CRRA coefficient *r* in AH risk experiment	-0.238	1.118
Switching row in the ambiguity experiment	6.361	2.019
Covariates		
Age (years)	47.540	8.735
Gender (1 = male, 0 = female)	0.735	0.442
Education (years)	9.074	2.188
Working off-farm (1 = yes, 0 = no)	0.540	0.499
Household size	4.444	1.440
Land size (hectares)	0.702	0.318
Household income (CNY 1000)	25.447	16.220
Observations	596	

Notes: \$1 ≈ CNY 6; S.D.: standard deviation. 'HL' refers to the risk preference measure elicited by the Holt and Laury (2002) task described in section 2.3 while 'AH' by the Andreoni and Harbaugh (2010) task described in section 2.4.

Table 2. Holt and Laury (2002) Task

	Option 1				Option 2				Range of relative risk aversion r
	p	x_1	$1-p$	x_2	p	x_1	$1-p$	x_2	
1	1/10	10.5	9/10	8.5	1/10	20	9/10	0.5	$r < -0.95$
2	2/10	10.5	8/10	8.5	2/10	20	8/10	0.5	$r < -0.95$
3	3/10	10.5	7/10	8.5	3/10	20	7/10	0.5	$-0.95 < r < -0.49$
4	4/10	10.5	6/10	8.5	4/10	20	6/10	0.5	$-0.49 < r < -0.15$
5	5/10	10.5	5/10	8.5	5/10	20	5/10	0.5	$-0.15 < r < 0.15$
6	6/10	10.5	4/10	8.5	6/10	20	4/10	0.5	$0.15 < r < 0.41$
7	7/10	10.5	3/10	8.5	7/10	20	3/10	0.5	$0.41 < r < 0.68$
8	8/10	10.5	2/10	8.5	8/10	20	2/10	0.5	$0.68 < r < 0.97$
9	9/10	10.5	1/10	8.5	9/10	20	1/10	0.5	$0.97 < r < 1.37$
10	10/10	10.5	0/10	8.5	10/10	20	0/10	0.5	$1.37 < r$

Notes: x_1 and x_2 are the two outcomes of each option; p is the probability of receiving outcome x_1.

Table 3. Andreoni and Harbaugh (2010) Task: Example of Choice Set

	Option 1	Option 2	Option 3	Option 4	Option 5
p	16/100	32/100	48/100	64/100	80/100
x	20	16	12	8	4

Notes: x is the possible outcome of each option with a probability p.

Table 4. Nine pairs of the Parameters m and r_1 in Choice Sets of the Andreoni and Harbaugh (2010) Task

Choice set	1	2	3	4	5	6	7	8	9
r_1	12.495	18.75	37.5	75	150	25	37.5	75	150
m	12	12	12	12	12	24	24	24	24

Notes: m is the experimental budget, and r_1 is the price of the probability of receiving a certain outcome.

Table 5. Choice Sets in the Ambiguity Task

	Risky option		Ambiguity Option	
	p	*x*	*p*	*x*
1	0/10	10	?	10
2	1/10	10	?	10
3	2/10	10	?	10
4	3/10	10	?	10
5	4/10	10	?	10
6	5/10	10	?	10
7	6/10	10	?	10
8	7/10	10	?	10
9	8/10	10	?	10
10	9/10	10	?	10
11	10/10	10	?	10

Notes: x represents the outcome, and p the probability of receiving the outcome x.

Table 6. Risk Preferences Elicited Through Different Measures at the Aggregate Level

	Risk aversion	Risk neutrality	Risk loving	Total number of subjects
Survey	40%	27%	33%	595
HL	68%	18%	14%	552
AH	55%	7%	38%	596

Notes: 'Survey' refers to the self-reported willingness to take risks in general, measured on a 5-point scale. 'HL' refers to the risk preference measure elicited by the Holt and Laury (2002) task described in section 2.3 while 'AH' by the Andreoni and Harbaugh (2010) task described in section 2.4.

Table 7. Within-subject Comparison between Survey and Experimental Measures

		HL experiment			AH experiment		
		Risk aversion	Risk neutrality	Risk loving	Risk aversion	Risk neutrality	Risk loving
Survey	Risk aversion	**25%**	8%	7%	**22%**	2%	16%
	Risk neutrality	20%	**5%**	3%	14%	**2%**	11%
	Risk loving	23%	6%	**4%**	19%	2%	**11%**

Note: See footnote of Table 6.

Table 8. Marginal Effects of Education on Risk Preferences Consistency

	Survey vs. HL	Survey vs. AH	AH vs. HL
	(1)	(2)	(3)
Education	-0.022 (0.015)	0.007 (0.009)	-0.011 (0.011)
Village fixed effects	Yes	Yes	Yes
Other controls	Yes	Yes	Yes
Observations	539	582	540
Log likelihood	-333.01	-367.47	-353.84

Notes: The dependent variable is a dummy variable equal to one if subjects have same classified risk preferences across measures, zero otherwise. Average probit marginal effects are reported. 'Survey' refers to the self-reported willingness to take risks in general, measured on a 5-point scale. 'HL' refers to the risk preference measure elicited by the Holt and Laury (2002) task described in section 2.3 while 'AH' by the Andreoni and Harbaugh (2010) task described in section 2.4. Other controls include age, gender, working off-farm, household size, land area, and household income. Standard errors clustered at the village level are presented in parentheses.

Table 9. Marginal Effects of Ambiguity Preferences on Inconsistent Risk Preferences

	Survey vs. HL	Survey vs. AH
	(1)	(2)
Ambiguity aversion (1/0)	0.113** (0.054)	0.129*** (0.049)
Village fixed effects	Yes	Yes
Other controls	Yes	Yes
Observations	335	343
Log likelihood	-164.2	-202.69

Notes: The dependent variable is a dummy variable equal to one if subjects shift from risk loving or risk neutrality in the experiment to risk aversion in the survey, or from risk loving in the experiment to risk neutrality in the survey, and zero if subjects shift towards risk loving, that is from risk aversion or risk neutrality in the experiment to risk loving in the survey, or from risk aversion in the experiment to risk neutrality in the survey. Other controls include age, gender, education, working off-farm, household size, land area, and household income. Average probit marginal effects are reported. Standard errors clustered at the village level are presented in parentheses. **, ***: Significant at the 5 per cent and 1 per cent levels, respectively.

Appendix

Table A1. Studies Comparing Different Risk Elicitation Measures

Studies	Sample size	Country	Risk elicitation measures		Results
			Survey questions	Incentivized experiments	
Panel A: Student subjects					
Ding et al. (2010)	121	China	WTR, hypothetical lottery question	Adjusted HL	The correlation of risk preferences elicited via the survey and the experiment is significant at the 5% level.
Harrison (1990)	46	USA		First-price auction, BDM	Compared to the first-price auction, BDM reveals a larger level of risk loving.
Isaac and James (2000)	34	-		First-price auction, BDM	Only a few subjects are stable between the two experiments.
Eckel and Grossman (2002; 2008)	261	USA	WTR	EG	The correlation of risk preferences elicited via the survey and the experiment is insignificant.
Kruse and Thompson (2003)	93	USA	WTR	Risk mitigation experiment	Of the total 93 subjects, only 23 subjects are consistent between the survey and the experiment.
Berg et al. (2005)	48	USA		BDM, English clock auction, first-price auction	Subjects are not stable across elicitation measures.
Deck et al. (2008b)	75	USA	WTR, hypothetical investment and risk job questions	HL, DOND	The correlation of risk preferences elicited between different methods is insignificant.
Anderson and Mellor (2009)	239	USA	hypothetical gamble questions	HL	The majority of subjects are not consistent across elicitation methods.
Bruner (2009)	157	USA		MPL (probability variation), MPL (reward variation)	A large number of subjects are inconsistent across methods.
Hey et al. (2009)	24	USA		PC, BID, ASK, BDM	The correlation of risk preferences elicited via different methods is low and insignificant in most cases.
Harbaugh et al. (2010)	96	USA		Choice and price-based lottery tasks	A large number of subjects are inconsistent across elicitation methods.
Crosetto and Filippin (2016)	444	Germany	WTR	HL, EG, GP, BRET	Elicited risk preferences are inconsistent across methods.
Deck et al. (2013)	203	USA	WTR	HL, EG, DOND, BART	The correlation of risk preferences elicited via different methods is low and insignificant in most cases.
Dulleck et al. (2015)	78	Australia		HL, AH	Only 10% of subjects have the same risk preference intervals in the two experiments.
Lönnqvist et al. (2015)	232	Germany	WTR	HL	The correlation of risk preferences elicited via the survey and the experiment is insignificant.
Vieider et al. (2015)	2939	30 countries	WTR	Adjusted HL	The correlation of risk preferences elicited via the survey and the experiment is significant for most countries.

Table A1. Studies Comparing Different Risk Elicitation Measures (continued)

Panel B: Non-student subjects					
Charness and Viceisza (2016)	91	Senegal	WTR	HL, GP	Subjects fail to understand the HL experiment but not the GP experiment. The risk patterns revealed via the survey and the GP experiment are different.
Hardeweg et al. (2013)	934	Thailand	WTR, hypothetical investment question	Adjusted HL	The correlation of risk preferences elicited via the survey and the experiment is significant at the 1% level.
Pennings and Smidts (2000)	346	Netherlands	WTR	MPL	The correlation of risk preferences elicited via the survey and the experiment is significant at the 1% level.
Dave et al. (2010)	881	Canada		HL, EG	At the aggregate level, subjects are more risk averse in the HL experiment than in the EG experiment.
Dohmen et al. (2011)	450	Germany	WTR, hypothetical investment question	Adjusted HL	The correlation of risk preferences elicited via the survey and the experiment is significant at the 1% level.
Reynaud and Couture (2012)	30	France	WTR	HL, EG	The correlation of risk preferences elicited via the survey and the experiment with low payoffs is weakly significant at the 10% level, the correlation of risk preferences elicited via the survey and the experiment with high payoffs is insignificant.

Notes: Elicitation measures: AH – the Andreoni and Harbaugh (2010) method; ASK – A minimal selling price for lotteries (Hey et al., 2009); BART – the Balloon Analogue Risk Task (Lejuez et al., 2002); BDM – the Becker, Degroot, and Marschack method (Becker et al., 1964); BID – A maximal buying price for lotteries (Hey et al., 2009); BRET – the Bomb Risk Elicitation Task (Crosetto and Filippin, 2013); DOND – the Deal or No Deal method (Deck et al., 2008a); EG – the Eckel and Grossman (2002) method; GP – the Gneezy and Potters (1997) method; MPL – Multiple Price List; PC – Pairwise Choice of lotteries (Hey et al., 2009); WTR – Willingness to take risks (Blais and Weber, 2006; Dohmen et al., 2011).

[1] See for instance, studies by Barham, Chavas, Fitz, Salas, and Schechter (2014), Cardenas and Carpenter (2008), Dohmen *et al.* (2011), Hill (2009), Liu (2013), Liu and Huang (2013), Tanaka, Camerer, and Nguyen (2010), and Ward and Singh (2015).

[2] See for instance, studies by Blais and Weber (2006), Hill (2009), Dohmen *et al.* (2011), or Vieider *et al.* (2015) for measures relying on simple survey questions; and Anderson and Mellor (2009), Ding, Hartog, and Sun (2010), Dohmen *et al.* (2011), or Vieider *et al.* (2015) for measures relying on hypothetical gambles, lotteries, and investments.

[3] See for instance, studies by Crosetto and Filippin (2013), Deck, Lee, and Reyes (2008), Eckel and Grossman (2002), Gneezy and Potters (1997), He, Martinsson, and Sutter (2012), Holt and Laury (2002), or Lejuez *et al.* (2002) for measures based on experiments.

[4] A promising option in overcoming some of these challenges is the use of tablets. However, difficulties could still be encountered if participants are computer illiterate, if there are restrictions in importing and transporting these technologies, and if there is lack of electricity or battery chargers. All these challenges as well as insurance needs can imply additional financial and time costs.

[5] In the pre-test, we mainly tested whether farmers could understand the questionnaire and the experimental instructions. In the first pre-test, we used a parallel gains/loss AH game following the original design by Andreoni and Harbaugh (2010). The AH game on gain worked well. However, due to its unusual design that asks subjects to choose their least preferred option, the AH game on loss confused subjects. The laboratory experiment conducted by Andreoni and Harbaugh (2010) has reported a similar problem. We decided to drop the loss part and only keep the gain part of the AH experiment in the second pre-test. After receiving positive feedback, we started the final experiments four days after the second pre-test.

[6] \$1 ≈ CNY 6.

[7] In total, 34 subjects did not participate in the study.

[8] The experimental instructions are available on request.

[9] The proportion of subjects that did not answer the control questions correctly is very small (about 1.7%). This implies that also the scope for selection bias is very small, and that the inconsistency across the different elicitation measures might not be the result of poor understanding of the tasks.

[10] In this study, we scale up the values of the outcomes in the original Holt and Laury (2002) study to make the maximal outcome equal to that in the Andreoni and Harbaugh (2010) experiment, so that the two experiments are comparable. All scaled-up values of outcomes are adjusted to the nearest multiples of 0.5 for the purpose of easing the implementation.

[11] A parallel gains/loss HL task is conducted but the task over losses is not in the scope of this study and is not reported. Therefore, each lottery over gains has a 1/20 chance of being selected.

[12] Hardeweg *et al.* (2013) find that about 40 per cent of subjects choose the midpoint and 25 per cent of subjects choose the two extreme points. Charness and Viceisza (2016) find a high proportion (27%) at one extreme point.

[13] The CRRA coefficient r estimated in the AH experiment is allocated into intervals of CRRA coefficient r shown in Table 2 for the purpose of comparing the AH experiment and the HL experiment.

[14] Except for land size, the coefficients of these control variables are not significant at conventional statistical levels.

[15] We also control for individual socio-demographic characteristics (age, gender, education, working off-farm, household size, income, and land size), village fixed effects, and cluster the standard errors at the village level. The effects of gender and household size are significant at the 5 per cent statistical level in both models. The effects of age and education are weakly significant (at the 10% statistical level), while the coefficients of the other variables are not significant.