UC Santa Barbara
Departmental Working Papers

Title

Comprehension and Risk Elicitation in the Field: Evidence from Rural Senegal

Permalink
https://escholarship.org/uc/item/5512d150

Authors
Charness, Gary
Viceisza, Angelino

Publication Date
2012-07-27

eScholarship.org Powered by the California Digital Library
University of California

Comprehension and Risk Elicitation in the Field: Evidence from Rural Senegal

Gary Charness and Angelino Viceisza*

July 27, 2012

ABSTRACT: In the past decade, it has become increasingly common to use simple laboratory games and decision tasks as a device for measuring both the preferences and understanding of rural populations in the developing world. In this paper, we report the results observed with three distinct risk elicitation mechanisms, using samples drawn from the rural population in Senegal, West Africa. We test the understanding of and the level of meaningful responses to the typical Holt-Laury task, to an adaptation of a simple binary mechanism pioneered by Gneezy and Potters in 1997, and to a non-incentivized willingness-to-risk scale. We find a low level of understanding with the Holt-Laury task and an unlikely-to-be-accurate pattern with the willingness-to-risk question. Our analysis indicates that the simple binary mechanism has substantially more predictive power than does the Holt-Laury mechanism. Our study is a cautionary note regarding utilizing either relatively sophisticated risk-elicitation mechanisms or non-incentivized questions in the rural developing world.

Keywords: Risk elicitation, laboratory experiments in the field, comprehension, rural Senegal.

JEL Classifications: B49, C91, C93, O13, O20

ACKNOWLEDGEMENTS: We would like to acknowledge seminar participants at IFPRI and the 2012 Foundations and Applications of Utility, Risk and Decision Theory (FUR) Conference for their meaningful comments. We would also like to thank Peter Ouzounov for his assistance during the design and implementation of the experiments and Pierre Ngom, Assane Thioune, and Seydou Tandjigora for their assistance during the execution of the experiments. Angelino Viceisza gratefully acknowledges financial support from the German Federal Ministry for Economic Cooperation and Development through the funding initiative for International Agricultural Research Centers and from the IFPRI Mobile Experimental Economics Laboratory (IMEEL).

* Gary Charness, Professor, Department of Economics, University of California, Santa Barbara, charness@econ.ucsb.edu; Angelino Viceisza, Research Fellow, International Food Policy Research Institute (IFPRI), Washington, D.C., a.viceisza@cgiar.org.

1. INTRODUCTION

It has become increasingly common in recent years for researchers in development economics to utilize tools developed by experimental economists, in order to measure important traits, attitudes, and characteristics of the poor in rural settings in the developing world. Some examples include Ashraf (2009); Attanasio et al. (2011); Castillo and Carter (2011); Harrison, Humphrey, and Verschoor (2010); Hill, Maruyama, and Viceisza (2012); and Giné et al. (2010).[1] The importance for policy recommendations of characterizing risk attitudes is well established.

As Harrison (2011) mentions, welfare evaluation of any proposed policy with risky outcomes should take into account people's risk attitudes. Consider for example the decision to adopt a new technology such as a new seed variety. This decision involves risk (and possibly, ambiguity) and is expected to depend on the decision-maker's risk attitude. So, governments and NGOs contemplating the introduction a new seed variety program, or other programs more generally should care about risk attitudes.[2]

A variety of techniques have evolved for testing risk preferences. These are typically incentivized, although non-incentivized questions have also been used in Western societies in recent years. The elicitation mechanisms range from the simple to the complex.[3] Some researchers favor the theoretical elegance of the more sophisticated approaches, while others favor the simpler mechanisms on the basis of the ease of comprehension and the greater probability of obtaining meaningful responses.

---

[1] For additional examples, refer to the discussion by Viceisza (2012).

[2] Suppose an NGO implements a pilot seed variety program. Researchers are often requested to evaluate the impact of such a program, by for example designing a randomized controlled trial (RCT) around it. It is common to ask: how does the impact of this RCT, or the adoption decision, vary with the average participant's risk preference? To answer this question, researchers turn to risk attitude data. If these data are imprecise or biased, incorrect inferences can emerge.

[3] Two of the most complex mechanisms are those put forward in Tanaka, Camerer, and Nguyen (2010) and Offerman et al. (2009).

In this study, we implement three elicitation devices in rural Senegal. One device is the well-known method used in Holt and Laury (2002), where people make a series of 10 choices between two systematically varied alternatives. A second device is an adaptation of a method used by Gneezy and Potters (1997), where there is a simple choice of how much to invest in a risky asset with a positive expected profit from investing. The third mechanism we use is a non-incentivized survey question; we are unaware of other studies that consider this type of non-incentivized instrument outside of Western societies.

In the particular case of the Senegalese subjects, risk attitudes are relevant in the following way. A larger project involving RCTs was conducted with the purpose of understanding behavioral dynamics in farmer groups. These internal dynamics typically require 'coordination' in the game-theoretic sense. Consider a group negotiating an overall price that depends on delivery of a certain (minimum) quantity to a buyer. The price a farmer gets by selling his harvest through the group depends on the amount other farmers deliver. In a setting like Heinemann, Nagel and Ockenfels (2009, the strategic uncertainty that emerges from this 'coordination' must be disentangled from the standard uncertainty arising from nature. This means that we need to control for group members' risk preferences. Concern about these issues naturally extends to forms of elicitation. The subjects in our experiments did not participate in this larger project, but were drawn from farmer groups that were in the same general population as the random sample of groups that was selected to be eligible for the aforementioned RCTs.

We are primarily interested in the relative effectiveness of these various elicitation approaches, including how well they are understood by people in a rural, developing country setting. To do so, we compare the patterns in the responses made in each instrument to the patterns observed in responses to the same questions in Western societies. We also use the data testing the Gneezy and Potters mechanism to predict behavior for the Holt and Laury mechanism, and *vice versa*. We go into further detail later in the paper.

Our main results can be summarized as follows. (1) There is much inconsistency in the choices made by individuals responding to the Holt-Laury (HL) questionnaire. (2) The Gneezy-Potters (GP) approach seems to have borne more fruit; our analysis shows that it has better predictive power than the HL mechanism. (3) While the willingness-to-take risk (WTR) question would seem to be easy to understand, responses are quite different from those in Germany with a similar mechanism and, quite surprisingly, suggest that women are much more likely to take on the highest degree of risk.[4] (4) Overall, our results indicate that simpler is better and that incentives appear to matter when eliciting risk preferences in developing nations.

Our study is related to that of Dave et al. (2010), who compare a relatively complex elicitation method to a coarser but simpler elicitation method. They find (with Canadian subjects) a nuanced result: "The simpler task may be preferred for subjects who exhibit low numeracy ... For subjects with higher numerical skills, the greater predictive accuracy of the more complex task more than outweighs the larger noise." While the level of numeracy is probably low in our subject population, the background environment is quite different. A low-income group comprised of unemployed subjects, as in Dave et al. (2010), in a place where there are much wealthier people living nearby and where mobility is possible, may make systematically different responses than people in Senegalese villages, where low income is ubiquitous and nearly inescapable. While both low-income groups are important, there are far more people in this group in rural areas of developing countries. There is also considerable research on risk-related aspects and behavior in these contexts, so our results are particularly

[4] Dohmen et al. (2011) experimentally validate a non-incentivized survey question similar to ours with the German population.

applicable to researchers doing such experimental work.

To our knowledge, there are very few studies comparing different risk elicitation instruments in these contexts. It is this literature at the intersection of experimental and development economics that we seek to complement with our study. Development economists wishing to gather data on risk preferences should take heed; furthermore, the spirit of our methodological results may also apply to the elicitation of other preferences in this environment.

The remainder is structured as follows. In Section 2, we describe the results from previous studies using the elicitation methods we implemented in Senegal, and we describe our implementation in considerable detail. Section 3 contains the experimental and survey results, and we conclude in Section 4.

2. EXPERIMENTAL DESIGN AND IMPLEMENTATION

There have been quite a few experimental studies on risk preferences, as this is one of the building blocks of economic theory and analysis. In this section, we only discuss the main articles that use the elicitation methods that we implement in our own study. For reviews of previous work, we refer the interested reader to Offerman et al. (2009) and Dave et al. (2010).

Holt and Laury (2002) ask participants to make 10 choices of either option A (the safe option) or option B (the risky option), one for each row. Table 1 shows the choices the participants faced in the low-payoff treatment. They also conducted treatments in which the payoffs are 20 times, 50 times, or 90 times the ones shown below.

[Table 1 about here]

It is clear that option B dominates option A in the last row. It would also require very strong risk-seeking preferences to choose option B in the first row. If a person chooses option A in the first row, there should be exactly one switch point. If we observe switching back and forth, this would seem to indicate either a lack of comprehension, frivolous responses, or indifference between alternatives; although this is unlikely since the instrument allows for indifference as an option.

Most studies in the Western world have inconsistency rates between 10 and 15 percent (13 percent in Holt and Laury 2002, 11 percent in Stockman 2006, and 12 percent in Meier and Sprenger 2010). [5] Table 2 shows that the evidence seems to be gloomier in settings more similar to ours. Lammers, Lau and Verbon (2006) find an inconsistency rate of 66.5% in South Africa; Galarza (2009) finds an inconsistency rate of 52% in Peru; and Doerr et al. (2011) find an inconsistency rate of 39 percent in Ethiopia. An exception is de Brauw and Eozonou (2011) who find an inconsistency rate of 14 percent in Mozambique. Jacobson and Petrie (2009) and Engle-Warnick et al. (2011) use an instrument that is different from HL (but can also be used to assess inconsistencies) and find a rate of 52 percent in Rwanda and Peru respectively.[6]

[Table 2 about here]

The original HL results indicate that the median and modal number of safe choices is 5.0 in the low-payoff condition. Higher payoffs lead to more risk-averse behavior; the median and modal number of safe choices in the 20x treatment is 6.0. In the 50x and 90x treatments with real payoffs, the medians are respectively 7.0 and 7.5, and the modes are 7.0 and 9.0.[7] Only

[5] A notable exception is Harrison, List and Towe (2007) whose source data collected on numismatists show an inconsistency rate of 73% if one classifies those who always choose B as consistent and 83% if as inconsistent.
[6] The list of studies included in this table is unlikely to be exhaustive. We consider it a lower bound for the following reasons:

- Several studies utilizing complex risk measures (including Holt-Laury) do not explicitly report inconsistency rates.
- Some studies do not allow for inconsistencies since they stop (and deem the subject consistent) once s/he switches for the first time.
- Studies with high inconsistency rates are less likely to get reported in the literature.

[7] Given the high per-capita cost of the 20x, 50x, and 90x treatments, they also asked for hypothetical choices; these choices showed considerably less risk aversion than those with real payoffs.

eight percent of the choices in the low-payoff treatment and six percent of those in the 20x real-payoff treatment indicate risk-seeking behavior.

Gneezy and Potters (1997) use a simple investment task. Each person is endowed with 100 units. Any number of these could be invested in a risky asset that has a one-third chance of success, and a payoff of 3.5 times the investment is successful; whatever is not invested is kept. Charness and Gneezy (2010) adapt this task to avoid probability-weighting issues (the inverse S-shaped curve; see Wu and Gonzalez 1996), by making the chance of success one-half and the success payoff 2.5 the investment made. Under either payoff calibration, any risk-neutral or risk-seeking person should invest all 100 units.

This mechanism typically gives a full range of values, and it is possible to directly compute the constant relative risk aversion (CRRA) coefficient from the investment choice (one can only determine a range with the HL mechanism). Furthermore, it seems rather easy to understand. One disadvantage is that one cannot distinguish between risk-neutral and risk-seeking preferences; however risk-seeking preferences appear to be rare. Charness and Gneezy (2010) elicit the risk preferences of students at UCSB with financial incentives. Table 1 of their paper shows that the average investment for 136 males and 64 females was 75.82 percent and 60.25 percent, respectively, giving an overall average of 70.84.

A number of other studies have used the GP mechanism. A summary of the results is provided in Table 3 (reproduced from Table 4 of Charness and Gneezy 2012, with the Charness and Villeval 2009 added in the last row).

[Table 3 about here]

We see considerable variation in the investment rate, with villagers more risk-averse than people in Western societies. The overall investment levels range from 44.67 to 65.42 among student populations and from 23.08 to 50.03 among villagers. In general, the purchasing power of the stakes involved is higher for the villagers, perhaps explaining the greater degree of risk aversion (as HL find with higher stakes). Nevertheless, we see that females are more risk averse than males in 10 of these 11 studies around the world, with virtually no difference in the other.

Dohmen et al. (2011) utilize a general risk question in the German Socio-Economic Panel (SOEP), which simply requests that respondents give an assessment of their general willingness to take risks on a 0-10 scale. This approach differs from the others in at least two important ways: (1) there are no financial incentives provided, and (2) this question is not specific to financial risk-taking, which may well be better suited to the purposes of development economists. It is clear that there are many forms of risk-taking (for example, physical risk, financial risk, emotional risk), so that this question may pick up traits orthogonal to the issue of concern; however, they also find that this measure matches up well with the results of an experimental validation. In any event, Figure 1 below (from their Figure 1) shows a good spread of risk attitudes in their data, with a strong peak at 5 (the average value in the range).

[Figure 1 about here]

Our experiments were conducted in rural Senegal (in the regions of Thies and Diourbel) in December 2010. Since the experiments were conducted as part of a larger project on linking farmers to markets, we had access to a sample of farmer groups that are part of the federation of non-governmental organizations of Senegal (FONGS), which represents such groups at the national level. The subjects in our experiments did not participate in this larger project, but were drawn from farmer groups that were in the same general population as the random sample of groups that was selected to be eligible for the larger scale RCTs. The participants were members of village-level farmer groups and were recruited by means of such groups.

We conducted four sessions across two days. On one day, two sessions—one HL and one GP—were held in a village in Diourbel. Each session was conducted with members of distinct farmer groups (between subjects). On the next day, two identical (between-subject) sessions were held in a village in Thies. In all sessions, we elicited the WTR question as part of a pre-survey, with the option to revise the response after the HL or GP task was completed. A typical session lasted between 2.5 and 3.0 hours. The average payoffs were FCFA 5,070.79 (approximately US$11), about twice the average daily wage for a comparable sample of households. There were 45 participants in the HL sessions and 46 participants in the GP sessions. Since all responded to the WTR question, these data comprised 91 observations.

Implementing such experiments in the field is always a challenging matter. We provide highly-detailed information about the implementation process in Appendix A, but three main aspects of our experiment protocol are noteworthy. (1) The experiments were conducted by a main experimenter in English and translated live into Wolof (the main national language) by a translator.[8] (2) We framed the HL and GP experiments in terms of "seeds" and "yields", since most subjects in rural areas can relate to concepts of risk in agricultural terms.[9] (3) For the WTR question, it turned out—learned while practicing with the translator prior to the experiments—that the term 'risk' was somewhat undefined in terms of the national language. Ultimately, in consultation with the translator, the experimenters agreed on describing risk as a situation that could sometimes lead to a good event (high payoff/gain) or a bad event (low payoff/loss).[10]

---

[8] For further discussion on the potential merits of this approach vis-à-vis other approaches, see Viceisza (2012).
[9] It is not uncommon to partially frame field experiments of this type, in the hopes of achieving a higher rate of comprehension. See table 2 for additional discussion.
[10] It may well be that this added noise to the WTR data. To the extent that it may have, others seeking to use this instrument in a developing country context should keep this in mind.

In our version of the HL formulation, a non-risky seed gave a payoff of FCFA 1,000 (800) in rainy (dry) weather, while a risky seed gave a payoff of FCFA 2,000 (100) in rainy (dry) weather; the probabilities were systematically varied. A screenshot is shown in Figure 2.

[Figure 2 about here]

In our version of the GP task, non-risky seeds pay FCFA 100 per kilo regardless of the weather. Risky seeds paid FCFA 300 per kilo if the weather was good (rainy) and nothing if the weather was bad (dry). So, one receives 1,000 units if one purchases only non-risky seeds and corresponding amounts for purchases of other numbers of risky seeds in the event of either rainy or dry weather (each 50 percent likely). We adapted the Charness and Gneezy (2010) payoff structure, making the successful payoff thrice the investment instead of 2.5 times the investment, as we felt that respondents would be less likely to be confused. We framed the decision in terms of how many "risky seeds" one wished to purchase. A screenshot is shown in Figure 3.

[Figure 3 about here]

One might be concerned that the physical procedure by which we elicit risk preferences is unusual, as there were no face-to-face interviews. Yet our approach is similar to that maintained in typical laboratory experiments (including HL) in Western societies where subjects are instructed at the classroom/session level in order to have a common understanding of the instructions and procedures. Some other studies in rural parts of developing countries have used similar techniques, as can be seen in Table 2. As the table shows, it is not clear that individual-level instruction necessarily leads to higher consistency. In addition, interviewing each person individually is enormously costly in terms of time.

One might also be concerned about possible bias and experimenter demand effects. Regarding possible bias, it would seem that the primary source would be an individual making a mistake in filling out the form, since comprehension (without coaching) should be the same in both cases. While we cannot entirely rule this out, our sense is that this was rather rare, particularly since we had assistants present to alert the experimenter to any need for clarification. Regarding experimenter demand effects, our own experience conducting experiments and surveys in the field suggests that—if anything—individual-level elicitation, which is typically done by having decentralized enumerators explaining the games/questions and eliciting responses, is more conducive to experimenter demand effects. We avoid this by having one of the authors be the main experimenter (paired with a translator), providing instructions at the session level and having individuals record their own responses, with assistance as necessary.[11]

3. EXPERIMENTAL RESULTS

Holt-Laury Task

Figure 4 compares the original HL low-real-payoff data with the Senegal data. The horizontal axis displays the decision (that is, 1 through 10) and the vertical axis displays the probability with which option A is chosen (across all respondents). The median and modal number of safe choices is 5.0, which is the same as the numbers in the original HL data for the low-payoff treatment and is of course smaller than these statistics in their higher-real-payoff treatments.

However, it is clear that the patterns differ between the original HL data and our data, as the shape of the distribution is dramatically different. If the Senegal participants on aggregate understand the task, one might expect the curve to look as it does in HL; however, this is not the case. Most notably, the Senegal curve shows relative insensitivity to changes in probability (and

[11] Enumerators' instructions and incentives are typically to ensure that subjects 'understand' the task prior to giving their responses, but there is a fine line between providing proper instructions and coaching to the extent that the responses 'look as they should'.

is not even monotonic). Furthermore, even in the last decision (where option B pays a greater amount with certainty), 40 percent of the subjects still choose option A. This is somewhat troubling. In fact, there is a great deal of within-subject inconsistency in the Senegal data.

[Figure 4 about here]

In order to analyze this inconsistency, we classify people into the following types:

1. Respondents who first chose option A and at some point switched to option B. We see these subjects as those who truly understood and consider them to be consistent.
2. Respondents who always chose option B. We see these subjects as consistent, but in principle we cannot rule out the possibility that they misunderstood.
3. Respondents who always chose option A. While we see these subjects as consistent, we also think they did not quite understand, since they should have switched to option B in decision 10. So, we classify them as "inconsistent" in subsequent analysis.
4. Respondents who switch at least twice.

Under this classification scheme (and if we disregard the possibility of mechanical entry errors), 11 of 45 participants (24.4 percent) appear to have possibly understood the task (type 1 or type 2), while another 11 participants always chose A (type 3). Finally, more than half of the participants (51.1 percent) switched columns at least twice. There is also a reasonable inference that people of type 2 did not really understand the task, since it takes a very pronounced risk-seeking preference to choose option B in the first row, since the expected payoff from option A is 820 and the expected payoff from option B is 290. Thus, in some sense, at most only the type 1 individuals (5 of 45 people) understood the task. This is obviously not a very high level of comprehension At most, 48.9 percent of the participants understood the task.

Overall, our impression is that our formulation of the HL task was not well understood by the participants. While there may be a better way to present this task to people in this type of environment, it seems that using a relatively sophisticated mechanism is not effective in this rural environment. Furthermore, this seems to be supported by some of the findings in Table 2.

Gneezy-Potters Task

Figure 5 shows the distribution of risky *seeds* chosen in the GP task conducted in Senegal. The average number of risky seeds chosen in Senegal is 4.78, or 47.8 percent. Recall that in past studies the range among villagers for this elicitation mechanism was roughly 25–50 percent, so this is actually in the upper part of this range.

[Figure 5 about here]

It is clear that there is a big spike at 5 risky seeds, as nearly one-third (15 of 46) purchased this number of seeds. Overall, the fitted kernel density looks fairly close to a normal distribution. We can compare these data to the two other studies involving adults for which we have individual data (Charness and Villeval 2009 and Haigh and List 2005). In the first case, the Kolmogorov-Smirnov test of cumulative distributions gives $\chi^2_2 = 2.71$, $p = 0.258$, so that there is no significant difference between the two distributions. In the second case, this test gives $\chi^2_2 = 4.56$, $p = 0.102$, so that again there is no significant difference between the two distributions.

Of course, we cannot know the *true* risk preference of the respondents in this study, so we cannot really state with any certainty that this mechanism elicits accurate responses. However, there is nothing in the data to indicate that people failed to understand the task here. It is unclear why our operationalization of the HL task appears to confuse participants relative to that of the GP task, as the graphical representations are very similar. One speculation is that our participants have a much more difficult time with varying probabilities than with varying payoffs. Perhaps there is a sense in which people understand a constant 50 percent probability better than the range of probabilities in the HL task. In any case, our interpretation is that the participants in our study gave more meaningful responses in the GP task than in the HL task.

Non-incentivized survey question

Figure 6 shows the distribution of subjects' levels of willingness to take risk on a scale of 1 to 10.[12] Dohmen et al. (2011) argue that such a measure generates the best all-around predictor of risky behavior. Using their original data, we can compare this distribution to the original (see Figure 2.1 for the original distribution reported in their paper). The histograms are similar in some ways but have two exceptions: the peak at 5 is smaller in the Senegal data and there is a very high peak at 10 in the Senegal data, with 27.4 percent of the observations. This is dramatically different than the roughly 1 percent in the Dohmen et al. data (based on their Figure 1). While we do not have their full data, a Kolmogorov-Smirnov cumulative distribution test across the two data sets gives χ^2_2 = 25.46, p = 0.000 on the assumption that there are 22,000 respondents in the latter study, with a 1 percent rate for 10.[13]

We can look more deeply into those who reported 10 on this question. It turns out that 68 percent of these people are women, with 17 of 49 women (34.7 percent) choosing the highest category. Given that women have been found to be at least financially more risk averse (see, for example, Charness and Gneezy 2012, for strong evidence) in settings around the world, this seems decidedly odd. For a comparison, only 3 of 29 female employees (10.3 percent) in Charness and Villeval (2009) chose to invest in the top range. The test of proportions gives Z = 2.380, p = 0.017 (two-tailed test) for the difference in rates.[14]

[Figure 6 about here]

[12] We report the first WTR responses, as they are not contaminated by the intervening task and data from the HL and GP treatments can be safely pooled. In any event, there is little difference between the first and second WTR measures. Eighty-two people (90.1 percent) did not change their report, 7 people (7.7 percent) decreased their WTR, and 2 people (2.2 percent) increased their WTR.

[13] Even if we reduce the number of observations to 220 and assume a rate of 2 percent, the Kolmogorov-Smirnov test gives χ^2_2 = 16.74, p = 0.000.

[14] There is no significant difference in male rates (the test of proportions gives Z = 0.55); in fact, this goes in the opposite direction, with 8 of 42 males (19.0 percent) in the Senegal data and 8 of 33 males (24.2 percent) in Charness and Villeval (2009) who invested in the top range.

Overall, we find that the Senegal results differ significantly from those of Dohmen et al. Furthermore, the risky options chosen by so many females in the Senegal study strongly suggests either a lack of comprehension or frivolous responses, since extreme risk-taking behavior has been rare in previous studies. Furthermore, recall (Table 3) that females are consistently more risk averse than males with the GP mechanism, at variance with the pattern observed here.[15]

A comparison of predictive accuracy

It is not necessarily straightforward to assess whether one elicitation device performs better than another. From an empirical standpoint, we can envision a few ways to proceed:

1) Developing a formal theoretical model that provides explicit hypotheses that are testable using data, either for a given device or across devices.
2) For any given device, comparing the patterns in the data for a given sample with patterns for another (comparable) sample.
3) For any give device, assessing whether certain covariates correlate with responses in "expected" ways, where expected may be defined as a result of 1 above or based on previous empirical literature. A typical example of such a covariate would be gender, but it could also be naturally occurring data that we might expect to be correlated with the experimental data; for example, data on actual agricultural risk-taking behavior.

Thus far, we have mainly relied on tests of type 2 and 3. In this section, we proceed along the lines of type 1. Our procedure uses data from one device to get a 'noise' parameter, which is then used to assess whether predicted behavior in the other device is different from actual behavior. We perform the following steps:

1) Consider decision-making in the GP device.
2) Assume that respondents make decisions according to expected utility and that they have square root utility. This enables us to calculate the expected utility □ of the number of risky seeds chosen □ as follows: □(□)=□□+(1−□)10−□, where □=0.5 as specified in the GP instrument.
3) Estimate a logit model of noisy decision-making.[16] That is, estimate the noise parameter □ by regressing □ on □ assuming that the probability of choosing □ follows a logistic specification: □□=□□□(□)Σ□=0,1,…,10□□□(□).

[15] We note that it is true that in our data women and men are observably different on certain characteristics and that gender may serve as a proxy for these.
[16] We also run a probit model and find subsequent results to be robust.

4) Predict decision-making (probabilities of occurrence) for the safe and risky lotteries in the HL device assuming the aforementioned logit specification and applying the estimate of the noise parameter ▯ obtained from GP.
5) Compare these predicted probabilities to the actual probabilities observed for the HL device. If these are statistically significantly different, we can conclude that the device does not seem to perform well since even after having assumed a model of noisy decision-making and accounted for the noise in the other device we still reject that predicted and actual behavior is similar.

We also follow the same procedure, but start with the data from the HL device and work in the reverse direction. That is, we obtain an estimate for the noise parameter using HL data and use this to predict behavior in the GP instrument. We then compare predicted and actual behavior in the GP device to get a sense of its performance.

Based on two-sided t-statistics, we find the following. For the HL mechanism, we reject at the 5% level that predicted and actual behavior are the same. This finding holds for both the safe (option A; *t*-stat of -2.2665, *p*-value of 0.0496) and the risky (option B; *t*-stat of 2.2685, *p*-value of 0.0495) lottery.[17] On the other hand, in the GP device we do not come close to rejecting that predicted behavior is the same as actual behavior (*t*-stat of 0.0337, *p*-value of 0.9737). These findings further support our previous claim that the HL device does not seem to have performed well, while the GP mechanism is substantially more effective. Further robustness checks on available characteristics (such as gender, age, education) suggest that this difference is *not* due to the average subject in the HL task being different from the average subject in the GP task.[18]

[17] When testing HL data, we test for the two lotteries separately. In case of indifference, we partition the mass according to the proportion who chose A or B. More detail on this analysis is available upon request.

[18] More detail on this analysis is available upon request. In short, we basically perform means-comparison tests across HL and GP for a relatively wide range of demographic characteristics (such as gender, age, education, marital status, number of children, and whether the subject is a group leader) and do not find these to be significantly different.

4. DISCUSSION

We test three different risk-preference elicitation mechanisms in rural Senegal. Our primary interest is pragmatic, whatever the relative intellectual merits of these approaches: do we get meaningful responses in this environment with each of these? If one is interested in policy implementation in developing nations and knowledge of risk preferences is useful, it is vital to implement a mechanism that is easily understood and gives non-spurious responses.

Our findings are rather straightforward. The Holt and Laury (2002) mechanism, a mainstay of risk elicitation in experimental economics, does not appear to induces sensible or realistic choices in rural Senegal. The majority of the respondents show inconsistency in their choice behavior and, even after allowing for noisy decision-making observed behavior, seems to be significantly different from predicted behavior. Of the remaining respondents, many make choices that are clearly dominated or that required an extreme degree of risk-seeking preferences to rationalize their choices. Thus, a researcher might wish to be cautious about using a relatively sophisticated mechanism in a rural, developing-country environment.

There is an additional issue regarding the analysis of inconsistency. Hirschauer, Maart, and Musshoff (2012) demonstrates that including inconsistent subjects in a HL (or similar) analysis "will bias the mean as well as the variance of the risk attitudes of the subject group of interest in ways that cannot be determined a priori. The only way to check for, and finally avoid, biases is to drop inconsistent subjects from the analysis." It is unclear how confident one should be in making policy recommendations when one must eliminate most of the data due to inconsistent choices. We also suspect that the differing levels of comprehension may correlate with differences in preferences, giving skewed results and making policy recommendations particularly precarious. So comprehension is a very serious issue, as there seems to be no good way to account for inconsistency, except possibly by excluding a large portion of the data.

The simpler mechanism, involving equally likely alternatives and a fixed rate of return, led to results that were fairly similar to results among adult employees (more comparable to our participant pool than students) at a French firm. In any case, none of the Senegal data from this mechanism look particularly unusual. This is further supported by the fact that when we allow for noisy decision-making we cannot reject that observed and predicted behavior are similar. This is the case even though the GP mechanism has been put into a price-list format not too dissimilar from that of HL; we conjecture that this difference corresponds to people having considerably more difficulties with varying probabilities than with varying amounts of income.

Finally, the simplest device of just asking people about how prepared they are to take risks delivers some results that differ from the patterns in other studies of risk elicitation. This could be because formulating the notion of *risk* is complex in this environment. There are far more claims of full risk tolerance made in the Senegal data than in the German survey data, and women are much more likely to indicate this tolerance. This latter observation does not mesh well with the strong evidence that women are more financially risk-averse than men; perhaps this general question picks up attitudes orthogonal to financial risk. In addition, it is entirely natural to wonder whether the lack of any sort of incentives played a role in generating our results.

While we do not presume to know the 'true' risk preferences of the people in rural Senegal, we feel it is useful to test for the effectiveness of different sorts of risk-elicitation mechanisms, since measuring risk preferences is critical for designing effective programs. Our study is an early attempt to gather data on this practical and important methodological question.

REFERENCES

Ashraf, B. N. 2009. "Spousal Control and Intra-Household Decision Making: An Experimental Study in the Philippines." *The American Economic Review* 99(4): 1245–1277.

Attanasio, O., A. Barr, J. Cárdenas, G. Genicot, and C. Meghir. 2011. Risk Pooling, Risk Preferences, and Social Networks. UCL Working Paper, London UK.

Bellemare, C., M. Krause, S. Kröger, and C. Zhang. 2005. "Myopic Loss Aversion: Information Feedback vs. Investment Flexibility." *Economics Letters* 87: 291–439.

Castillo, M., and M. Carter. 2011. "Trustworthiness and Social Capital in South Africa: Analysis of Actual Living Standards Data and Artefactual Field Experiments." *Economic Development and Cultural Change* 59(4): 695–722.

Charness, G., and G. Genicot. 2009. "Informal Risk Sharing in an Infinite-Horizon Experiment." *Economic Journal* 119(537): 796–825.

Charness, G., and U. Gneezy. 2010. "Portfolio Choice and Risk Attitudes: An Experiment." *Economic Inquiry* 48(1): 133–146.

________. 2012. "Strong Evidence for Gender Differences in Experimental Investment." *Journal of Economic Behavior and Organization* 83(1): 50-58.

Charness, G., and M. Villeval. 2009. "Cooperation, Competition, and Risk Attitudes: An Intergenerational Field and Laboratory Experiment." *The American Economic Review* 99(3): 956–978.

Dave, C., C. Eckel, C. Johnson, and C. Rojas. 2010. "Eliciting Risk Preferences: When is Simple Better?" *Journal of Risk and Uncertainty* 41(3): 219–243.

de Brauw, A., and P. Eozonou. 2011. *Measuring Risk Attitudes among Mozambican Farmers.* IFPRI Harvest Plus Working Paper 6, Washington DC.

Doerr, U., T. Mahmoud, and U. Schmidt. 2011. *Overconfidence and Risk Management of Ethiopian Farmers.* Working Paper.

Dohmen, T., A. Falk, D. Huffman, U. Sunde, J. Schupp, and G. Wagner. 2011. "Individual Risk Attitudes: Measurement, Determinants, and Behavioral Consequences." *Journal of the European Economic Association* 9(3): 522–550.

Dreber, A., and M. Hoffman. 2007. *2D:4D and Risk Aversion: Evidence that the Gender Gap in Preferences is Partly Biological*. Stockholm School of Economics Working Paper, Stockholm.

Engle-Warnick, J., J. Escobal, and S. Laszlo. 2011. "Ambiguity Aversion and Portfolio Choice in Small-Scale Peruvian Farming." *The B.E. Journal of Economic Analysis & Policy* 11(1): 68.

Ertac, S., and M. Gurdal. 2010. *Deciding to Decide: Gender, Leadership, and Risk-Taking in Groups.* TÜSÝAD-Koç University Economic Research Forum Working Papers 1028, Istanbul.

Fellner, G., and M. Sutter. 2009. "Causes, consequences, and cures of myopic loss aversion - An experimental investigation." *Economic Journal* 119: 47-60.

Galarza, F. 2009. *Choices under Risk in Rural Peru.* Munich Personal RePec Archive Working Paper 17708, Munich.

Giné, X., P. Jakiela, D. Karlan, and J. Morduch. 2010. "Microfinance Games." *American Economic Journal: Applied Economics* 2(3): 60–95.

Gneezy, U., and J. Potters. 1997. "An Experiment on Risk Taking and Evaluation Periods." *Quarterly Journal of Economics* 112: 631–645.

Gneezy, U., K. Leonard, and J. List. 2009. "Gender Differences in Competition: Evidence from a Matrilineal and a Patriarchal Society." *Econometrica* 77: 1637–1664.

Gong, B., and C-L. Yang. Forthcoming. "Gender Differences in Risk Attitudes: Field Experiments on the Matrilineal Mosuo and the Patriarchal Yi." *Journal of Economic Behavior and Organization*.

Haigh, M., and J. List. 2005. "Do Professional Traders Exhibit Myopic Loss Aversion? An Experimental Analysis." *Journal of Finance* 60(1): 523–534.

Harrison, G., 2011. "Experimental Methods and the Welfare Evaluation of Policy Lotteries." *European Review of Agricultural Economics* 38(3): 335–360.

Harrison, G., J. List, and C. Towe. 2007. "Naturally Occurring Preferences and Exogenous Laboratory Experiments: A Case Study of Risk Aversion." *Econometrica* 75: 433–458.

Harrison, G., S. Humphrey, and A. Verschoor. 2010. "Choice under Uncertainty: Evidence from Ethiopia, India, and Uganda." *Economic Journal* 120 (March): 80–104.

Heinemann, F., R. Nagel, and P. Ockenfels. 2009. "Measuring strategic uncertainty in coordination games." *Review of Economic Studies* 76(1): 181–221.

Hill, R., E. Maruyama, and A. Viceisza. 2012. *Breaking the Norm: An Empirical Investigation into the Unraveling of Good Behavior*. *Journal of Development Economics* 99 (1): 150-162.

Hirschauer, N., S. Maart, and O. Musshoff (2012), "How to avoid mean and variance bias in Holt-and-Laury lotteries - a note," unpublished manuscript.

Holt, C., and S. Laury. 2002. "Risk Aversion and Incentive Effects." *The American Economic Review* 92 (5): 1644–1655.

Jacobson, S., and R. Petrie. 2009. "Learning from Mistakes: What do Inconsistent Choices over Risk Tell Us?" *Journal of Risk and Uncertainty* 38(2): 143–158.

Lammers, J., M. Lau, and H. Verbon. 2006. Perceived HIV- contamination risk, risk aversion and time preferences (A Laboratory Experiment in South Africa). Conference paper.

Langer, T., and M. Weber. 2004. *Does Binding or Feedback Influence Myopic Loss Aversion? An Experimental Analysis.* University of Muenster Working Paper, Muenster.

Meier, S., and C. Sprenger. 2010. "Present-Biased Preferences and Credit Card Borrowing." *American Economic Journal: Applied Economics* 2(1): 193–210.

Offerman, T., J. Sonnemans, G. van de Kuilen, and P. Wakker. 2009. "A Truth-Serum for Non-Bayesians: Correcting Proper Scoring Rules for Risk Attitudes." *Review of Economic Studies* 76: 1461–1489.

Stockman, C. 2006. *Differences and Similarities in Health Risk Preferences across Two Subject Cohorts.* California University of Pennsylvania Working Paper, California, PA.

Tanaka, T., C. Camerer, and Q. Nguyen. 2010. "Risk and Time Preferences: Linking Experimental and Household Survey Data from Vietnam." *The American Economic Review* 100(1): 557–71.

Viceisza, A., 2012. "Treating the field as a lab: A basic guide to conducting economics experiments for policymaking." *Food Security in Practice Series Technical Guide*. Washington, DC: International Food Policy Research Institute.

Wu, G., and R. Gonzalez. 1996. "Curvature of the Probability Weighting Function." *Management Science* 42: 1676–1690.

APPENDIX A: IMPLEMENTATION DETAILS

Outline

The typical session comprised:

1) Survey 1
2) An introduction of the experimenter, translator, assistant(s), IFPRI, and the project
3) Instructions on the game (more below)
4) The game (sessions 1 and 3 were HL and sessions 2 and 4 were GP)
5) An opportunity to revise certain responses to survey 1
6) The lottery
    a. For HL, this comprised three draws:
        i. The first draw was to determine the row for which the subjects would get paid. In other words, this determined the distribution of the weather.
        ii. The second draw was to determine whether those who chose indifference would be treated as "magasin/storage A" or "magasin/storage B".
        iii. The third draw was to determine the weather, that is, whether it was good or bad.
    b. For GP, this comprised one draw that determined the weather.
7) Survey 2 (demographic survey)
8) Payment

Layout

1) The experiments were conducted in classrooms in the local schools. Boxes were used as dividers to give privacy.
2) The typical layout of the room was as follows:

FRONT OF ROOM (experimenter, translator, and white board)			
Seat 1	Seat 7	Seat 13	Seat 19
Seat 2	Seat 8	Seat 14	Seat 20
Seat 3	Seat 9	Seat 15	Seat 21
Seat 4	Seat 10	Seat 16	Seat 22
Seat 5	Seat 11	Seat 17	Seat 23
Seat 6	Seat 12	Seat 18	Seat 24
BACK OF ROOM (assistant experimenter/cashier)			

Survey 1

1) This survey was administered prior to anything else, that is, as soon as subjects walked into the laboratory.
2) The survey comprised the following questions:
    a. On the following scale of 1 to 10, please indicate how willing you are to take risks.
    b. In your day-to-day life, what do you consider to be a risky decision? Please describe using one or more examples.
    c. How much is 15% of 2,000 FCFA? If you don't know, put an **X**.

Introduction

1) The experimenter introduced himself, the translator and the two assistants. Typically, the main assistant experimenter was not introduced till the end.
2) The experimenter introduced IFPRI and the larger project, typically as follows:
    a. IFPRI is an institute in the United States.
    b. We are conducting a research project on farmer groups, their activities and so on.
    c. We have been holding discussions with farmers across many parts of Senegal. In particular, we have talked to farmers in Diourbel/Thies, but we have not been here before.
    d. For the upcoming task, we will pay you for the decisions that you make. We pay you for two purposes:
        i. Because you came here today and are spending your time with us. This is time in which you could be doing something else, so we pay you for that reason.
        ii. Also, we would like you to take this decision seriously, as you do any other decision in real life.

Instructions and Game

1) Game 1 (HL)
    a. The experimenter handed out the sheet of paper for the HL game.
    b. The experimenter first asked subjects what they thought the pictures on the form represented.
        i. This served as an icebreaker. It basically enabled subjects to start thinking about the material and the decisions they would be presented with during the session.
        ii. In some cases the storage was seen as a school and the good weather was perceived as clouds, but typically subjects soon realized that the task would have something to do with storage and good/bad weather.
    c. After this mini brainstorming, the experimenter explained the following steps:
        i. The brainstorming has shown that the task today has to do with storage and the weather.
        ii. Specifically, suppose there are two types of storage rooms (A/Abdu and B/Bara) that contain two different types of fertilizer/angrais (A/Abdu and B/Bara). We are going to ask you which of these two fertilizers you prefer.
        iii. How are these two fertilizers different? Let's focus on the first row of the first page.
            1. Fertilizer A in magasin Abdu
                a. The fertilizer in magasin Abdu gives FCFA 1,000 as income from production in times of bad weather and FCFA 800 in times of good weather.
                    i. Explain payoff and how it is associated with good/bad weather.
                    ii. Quiz people on how much the payoff is in times of good/bad weather.
            2. Fertilizer B in magasin Bara

a. Now, let's look at the fertilizer in magasin Bara. What is different about it? Well, this fertilizer gives FCFA 2,000 as income from production in times of good weather but FCFA 100 in times of bad weather.
b. So, the difference between the two fertilizers is that the one in magasin Bara pays MORE in times of good weather but LESS in times of bad weather.
    i. Similar to above, explanation and quiz.

3. Recap: So, we've seen that there are two types of fertilizer, the one in magasin Abdu and the one in magasin Bara. We also know that they're affected by the weather.
4. *What do we know about the weather?*
    a. As in real life, sometimes the weather is good and sometimes the weather is bad.
    b. These 10 numbers (1, 2, 3,…, 10) represent 10 years of bad weather.
    c. In the first row, 1 out of 10 years the weather is good and 9 out of 10 years the weather is bad.
        i. The number 1 represents the year that the weather is good.
        ii. The numbers 2, 3,…, 10 represent the years that the weather is bad.
        iii. The numbers in the columns of good/bad weather represent the years that weather can be good/bad.
        iv. So, note that the weather is the same for magasin Abdu and magasin Bara. *What is different is the income from production you get depending on weather being good or bad.*
    d. Questions/quiz for understanding
        i. How many years can the weather be good in row 1?
        ii. How many years can it be bad?
        iii. What is the income from production if the weather is bad?
            1. Depends on whether you buy Abdu or Bara.
        iv. Suppose you buy Abdu and the weather is good, what is your income from production? How about Bara?
        v. How about if the weather is bad?

d. This explains row 1. How are the other rows different from row 1?
    i. Notice that when we go from row 1 to row 2, the only aspect that changes is the number of years that weather can be good/bad. That is, the income from production does NOT change. However, in row 2 the number of years that weather can be good is 2 and the number of years that the weather can be bad is 8.
        1. Typically, the experimenter showed the years with the numbers 1, 2 in the left hand and 3, 4,…, 10 in the right hand.

ii. Now, what happens if we go from row 2 to row 3? Now, weather can be good 3 out of 10 years and bad 7 out of 10 years.

iii. This process was continued up to row 10.

1. At this stage, subjects typically smiled indicating their understanding that in row 10 the weather was always good.

e. So, we are going to ask you to make a decision for each of the rows: Abdu or Bara. If you do not know which one to choose, you can choose I for "indifferent."

f. Is this clear?

i. At this point, a row was selected to quiz subjects again. Questions were asked with regard to the probabilities and earnings.

ii. Then, subjects were informed that only one row would be selected for payment. The exact procedures for selecting the row and drawing/simulating the weather were typically explained when the lottery was drawn in order to avoid too much information prior to decisions being made.

iii. Then, decisions were made.

2) Game 2 (GP)

a. The game sheets (appendix) were handed out. Subjects were prompted on the images at the top as an icebreaker exercise. In both sessions they recognized them correctly as the two types of weather.

b. The experimenter asked the subjects to imagine that they are grain farmers and they are to be given 10 kilos of seeds to plant for the new season. They are told that they can take two types of seeds–from Abdu or from Bara. At this point the experimenter emphasized that they must take a total of 10. He did this by giving them examples of the possible combinations of the seeds that they could take.

i. Imagine that you are a grain farmer and you are given 10 kilos of seeds for free for the coming season. You can choose between two different types of seeds to take. You can either take the seeds of Abdu or of Bara. It is important that you realize you can take as many kilos of Abdu and Bara seeds as you want as long as at the end you are taking 10 kilos in total—no more and no less. On your answer sheet you will see a place for you to write how many Abdu seeds and Bara seeds you will take.

c. Next the experimenter explained how the seeds are different. Abdu seed is of higher quality than Bara seed but is more vulnerable to the weather. That is, when there is good weather the Abdu seed produces a harvest that sells for FCFA 300. When the weather is bad the harvest is so bad that it cannot be sold, eaten, or fed to the animals. On the other hand, the Bara seed does not respond to the weather and always gives FCFA 100 francs per kilo.

d. Next the experimenter proceeded to go through the columns for the Abdu seed and explain how different quantities of Abdu seed affect one's income from the harvest given good weather. What was emphasized through examples was that 300 times the number of kilos of the seeds determines the income, which is then provided for the subjects in the column 'xalis…. Abdu' on the side with good weather. This was done to the point where the experimenter felt comfortable with their understanding of the derivation of their income. Next the experimenter explained the bad weather columns for the Abdu seed, which was provided in the column 'xalis … Abdu on' the side with bad weather. This was always zero. Again examples were given until the experimenter felt comfortable with their understanding.

e. The same procedure followed for the Bara seed: first with the good weather 'xalis… Bara' column and then with the bad weather 'xalis…Bara' column. It was emphasized that there was no difference between the columns.
f. The experimenter explained the total income for any given type of weather–by adding the columns of 'xalis… Abdu' and 'xalis… Bara'. It was explained that this number was indicated in the column 'li ngay … xalis'.
g. The experimenter asked subjects specific questions such as:
    i. *If the weather is good and one had 5 kilos of bara seeds how many Abdu seeds does one have? How much money does one make from these Bara seeds? ... from these Abdu seeds? In total?* Then he repeated for other combinations… (1 and 9, 3 and 7… etc. – each time varying the weather)
    ii. *Whenever one subject seemed to dominate by answering correctly in succession, the experimenter asked the translator to explain to him that we would like to hear from other people as well. These examples were repeated until the experimenter felt confident about the understanding of the subjects.*
h. The experimenter explained how weather was unknown at the time of the decision, how this was realistic, and how it was to be determined, from a box, with equal probability of the two types of weather. That is, the weather was to be determined from a box where drawing cards with numbers 1–5 would correspond to good weather, while drawing numbers 6–10 would correspond to bad weather. These numbers were the same as the ones used for the HL task.
i. Decisions were then made.

Lottery
The lotteries were conducted according to the procedures described previously. Typically, we let one of the subjects draw. Papers with numbers 1 through 10 were drawn from a bag.

Survey 2 (demographics)

1) This survey was administered after the main task and comprised the following questions:
    a. Education level
    b. Marital status
    c. Number of children
    d. Primary occupation
    e. How often do you find yourself short of cash?
    f. How much do you agree with the statement "Most people can be trusted"? (1=Strongly Disagree, 2=Disagree, 3=Slightly Disagree, 4=Neither Agree or Disagree, 5=Slightly Agree, 6=Agree, 7=Strongly Agree)
    g. How many cigarettes do you smoke per day?
    h. On a scale of 1 to 10, how patient do you consider yourself?

Payment
After all these steps were done, subjects were called by their seat number—one by one—to get paid in private by the assistant experimenter. They were also paid a fixed fee for showing up.

TABLES AND FIGURES

Table 1—Holt-Laury lottery choices

Option A	Option B	Expected payoff difference
1/10 of $2.00, 9/10 $1.60	1/10 of $3.85, 9/10 of $0.19	$1.17
2/10 of $2.00, 8/10 $1.60	2/10 of $3.85, 8/10 of $0.19	$0.83
3/10 of $2.00, 7/10 $1.60	3/10 of $3.85, 7/10 of $0.19	$0.50
4/10 of $2.00, 6/10 $1.60	4/10 of $3.85, 6/10 of $0.19	$0.16
5/10 of $2.00, 5/10 $1.60	5/10 of $3.85, 5/10 of $0.19	-$0.18
6/10 of $2.00, 4/10 $1.60	6/10 of $3.85, 4/10 of $0.19	-$0.51
7/10 of $2.00, 3/10 $1.60	7/10 of $3.85, 3/10 of $0.19	-$0.85
8/10 of $2.00, 2/10 $1.60	8/10 of $3.85, 2/10 of $0.19	-$1.18
9/10 of $2.00, 1/10 $1.60	9/10 of $3.85, 1/10 of $0.19	-$1.52
10/10 of $2.00, 0/10 $1.60	10/10 of $3.85, 0/10 of $0.19	-$1.85

Source: Holt and Laury (2002)

Table 2—Inconsistency rates reported by selected developing country field studies

Study	Method	Protocol	Subjects	Inconsistency	Implications
Lammers, Lau, Verbon (2006)	HL	- Paper-based - Neutral framing - Classroom explanation as part of experiment session - Real stakes	South African students	66.5%[a]	- Not cited due to preliminary nature of the analysis
Galarza (2009)	HL	- Paper-based - Mixed framing with e.g. safe (risky) lottery called sol (luna) - Classroom explanation as part of experiment session - Real stakes	Peruvian farmers	52%[a]	- Some evidence that multiple switching behavior explained by nonlinear probability weighting made in a context of random calculation mistakes - Majority of farmers characterized by prospect theory
De Brauw and Eozenou (2011)	HL	- Paper-based - Loaded framing with lotteries in terms of sweet potato varieties with varying yields - Individual explanation as part of survey - Hypothetical stakes	Mozambican farmers	14%[a]	- About three-fourths of the sample develops risk preferences by rank dependent utility - By making the CRRA assumption, they poorly predict risk preferences among those who are less risk averse
Doerr et al. (2011)	HL	- Paper-based - Neutral framing with lotteries displayed as balls in urns - Individual explanation as part of survey - Real stakes	Ethiopian farmers	39%[b]	- Risk taking of farmers seems to be unrelated to risk preferences but highly correlated to overconfidence measures
Jacobson and Petrie (JP, 2009)	Not HL*	- Paper-based - Neutral framing with lotteries called A and B - Individual explanation as part of survey - Real and hypothetical stakes	Rwandan farmers	52%[c]	- While risk aversion alone does not explain financial decisions, risk aversion and inconsistent choices interact in significant and sensible ways - Mistakes correlate with less than optimal behavior
Engle-Warnick et al. (2011)	Not HL**	- Paper-based - Neutral framing with lotteries displayed as color chips in a bag - Classroom explanation as part of experiment session - Real stakes	Peruvian farmers	52%[d]	- They combine experimental and survey data to find that ambiguity aversion predicts actual technology choices on the farm

Table notes:

[a] These rates are based on subjects who switched more than once in the HL task.

[b] This rate is based on a test of the common ratio effect, which is based on the Allais paradox, paired with the HL task.

[c] This rate is based on the number of people who make at least one mistake.

[d] This rate is based on the number of people who choose the dominated lottery at least twice. If we consider those who choose the dominated lottery at least once, the rate would be 76%.

* The format of the experiment is similar to HL, but with several key differences. HL keep payoffs constant and vary the probabilities of receiving the high and low outcomes. In JP, the probability is always 50–50, and the payoffs are varied. Also, HL present the lotteries all at once to the subjects. JP present lottery pairs sequentially. Finally, JP also present lotteries over losses.

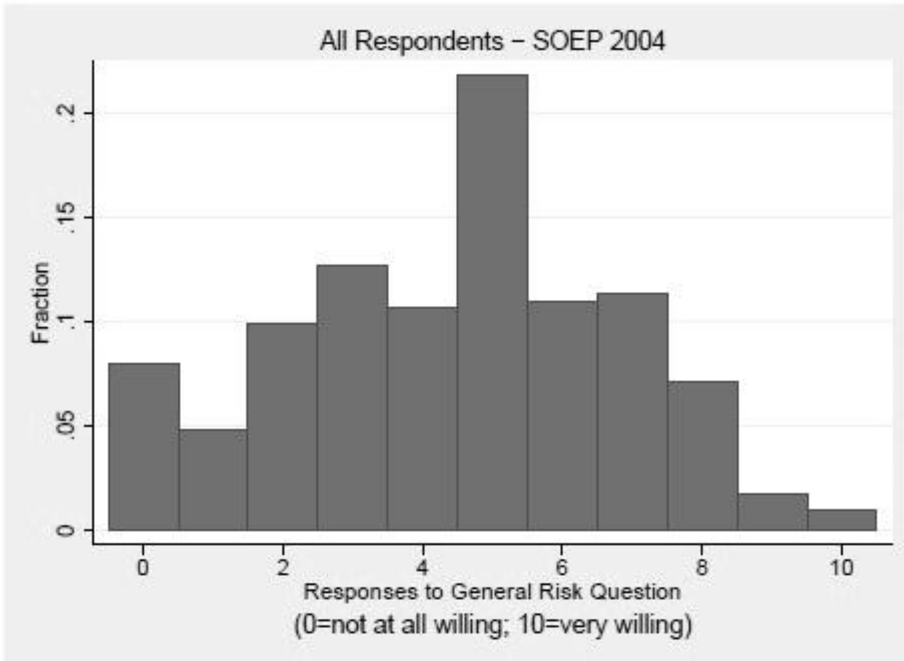
** Their decomposition resembles the instrument in HL. Each row in the decomposition corresponds to a single binary choice between two alternative gambles. Beginning with the first row of choices and moving down, an expected utility maximizer will at some point switch from the left-hand side gamble with lower variance to the right-hand side gamble with a higher variance and slightly higher expected utility.

Table 3—Investment choices in other studies

Study	Participants	Periods	Avg. Male Investment (N)	Avg. Female Investment (N)
Langer and Weber (2004)	Finance students, Mannheim	30	64.62 (93)	58.70 (14)
Haigh and List (2005)	Professional traders, CBOT	9	58.30 (50)	55.59 (8)
Fellner and Sutter (2009)	Undergrads, Jena	18	57.44 (39)	49.04 (79)
Bellemare et al. (2005)	Undergrads, Tilburg	9	45.48 (95)	42.73 (40)
Charness and Genicot (2009)	Undergrads, UCLA	1	59.22 (41)	52.23 (53)
Dreber and Hoffman (2007)	Students, Stockholm	1	69.60 (92)	50.00 (55)
Gneezy, Leonard, and List (2009)	Villagers in Tanzania and India	1	50.00 (157)	50.06 (157)
Ertac and Gurdal (2010) [a]	Undergrads, Turkey	1	72.32 (79)	54.29 (49)
Gong and Yang (forthcoming)	Matrilineal villagers in China	1	53.9 (31)	32.5 (36)
Gong and Yang (forthcoming)	Patrilineal villagers in China	1	37.3 (37)	4.3 (28)
Charness and Villeval (2009)	Workers at French firms	1	62.73 (33)	53.2 (29)

Note: [a] We include only the individual risk decisions where there is a positive expected return from investing in the risky asset. A similar gender difference applies in the other cases.

Figure 1—Non-incentivized survey risk attitudes in Dohmen et al. (2011)



Source: Dohmen et al. (2011). Note: SOEP is German Socio-Economic Panel.

Figure 2—Holt-Laury task screen shot

SESSION ______________ DAY ______________ SEAT ______________

	Storage A				Storage B				Choice (A, B, I)
	good weather	yield	bad weather	yield	good weather	yield	bad weather	yield	
1	1	1000	2 3 4 5 6 7 8 9 10	800	1	2000	2 3 4 5 6 7 8 9 10	100	
2	1 2	1000	3 4 5 6 7 8 9 10	800	1 2	2000	3 4 5 6 7 8 9 10	100	
3	1 2 3	1000	4 5 6 7 8 9 10	800	1 2 3	2000	4 5 6 7 8 9 10	100	
4	1 2 3 4	1000	5 6 7 8 9 10	800	1 2 3 4	2000	5 6 7 8 9 10	100	
5	1 2 3 4 5	1000	6 7 8 9 10	800	1 2 3 4 5	2000	6 7 8 9 10	100	

SESSION ______________ DAY ______________ SEAT ______________

	Storage A				Storage B				Choice (A, B, I)
	good weather	yield	bad weather	yield	good weather	yield	bad weather	yield	
6	1 2 3 4 5 6	1000	7 8 9 10	800	1 2 3 4 5 6	2000	7 8 9 10	100	
7	1 2 3 4 5 6 7	1000	8 9 10	800	1 2 3 4 5 6 7	2000	8 9 10	100	
8	1 2 3 4 5 6 7 8	1000	9 10	800	1 2 3 4 5 6 7 8	2000	9 10	100	
9	1 2 3 4 5 6 7 8 9	1000	10	800	1 2 3 4 5 6 7 8 9	2000	10	100	
10	1 2 3 4 5 6 7 8 9 10	1000		800	1 2 3 4 5 6 7 8 9 10	2000		100	

Figure 3—Gneezy-Potters task screen shot

Day__________ Session __________ Seat __________

_____Seeds Abdu _____ Seeds Bara

Good weather 1 2 3 4 5

Bad Weather 6 7 8 9 10

Kilos of seed Abdu	Kilos of seed Bara	Good weather: Yield Abdu	Good weather: Yield Bara	Good weather: Total Yield	Bad Weather: Yield Abdu	Bad Weather: Yield Bara	Bad Weather: Total Yield
0	10	0	1000	1000	0	1000	1000
1	9	300	900	1200	0	900	900
2	8	600	800	1400	0	800	800
3	7	900	700	1600	0	700	700
4	6	1200	600	1800	0	600	600
5	5	1500	500	2000	0	500	500
6	4	1800	400	2200	0	400	400
7	3	2100	300	2400	0	300	300
8	2	2400	200	2600	0	200	200
9	1	2700	100	2800	0	100	100
10	0	3000	0	3000	0	0	0

Figure 4—Senegal versus original HL data (probability of choosing option A)

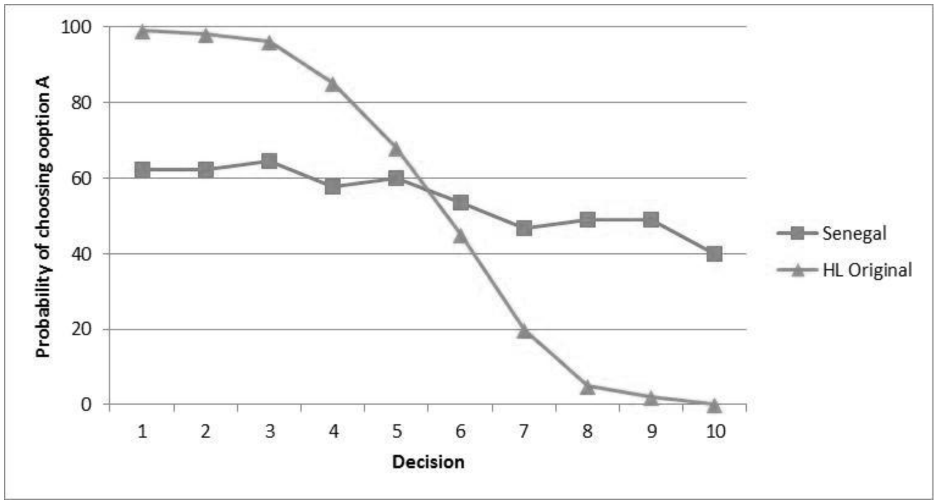


Figure 5—Histogram of Senegal GP data (normal and kernel densities overlaid)

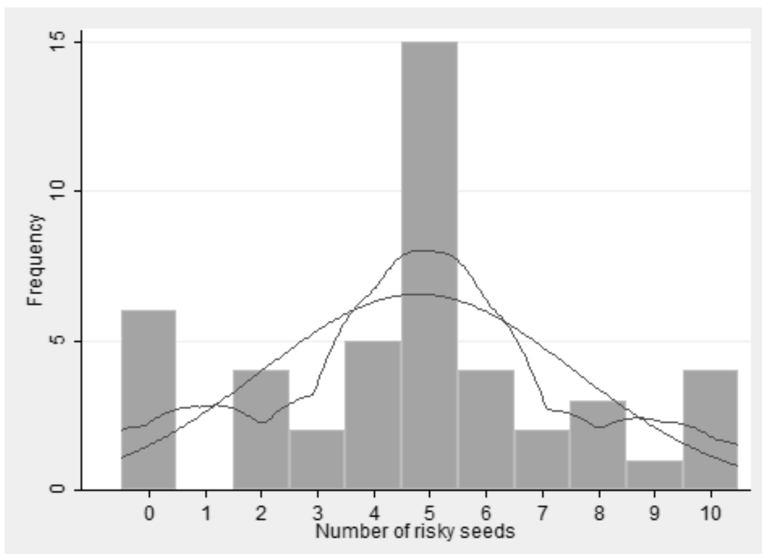
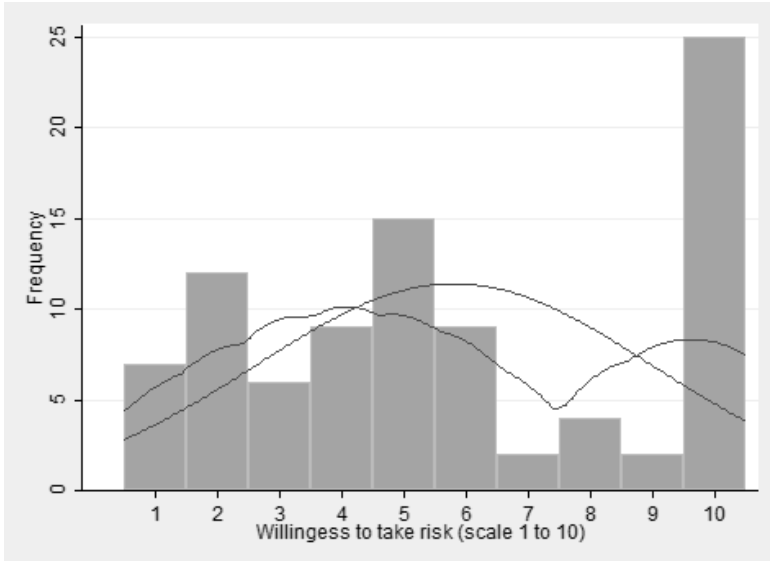


Figure 6—Histogram of Senegal WTR data (normal and kernel densities overlaid)


Frequency
0
5
10
15
20
25
1
2
3
4
5
6
7
8
9
10
Willingess to take risk (scale 1 to 10)