Norwegian University of Life Sciences (NMBU)

Endowment Effects and Loss Aversion in the Risky Investment Game

Stein T. Holden and Mesfin Tilahun

Norwegian University of Life Sciences
Centre for Land Tenure Studies

Centre for Land Tenure Studies Working Paper 01/20
ISBN: 978-82-7490-284-8

Endowment Effects and Loss Aversion in the Risky Investment Game

Stein T. Holden and Mesfin Tilahun

School of Economics and Business, Norwegian University of Life Sciences, P. O Box 5003, 1433 Ås, Norway
Email: stein.holden@nmbu.no

Abstract

The risky investment game of Gneezy and Potters (1997) has been a popular tool used to estimate risk tolerance and myopic loss aversion. We have assessed whether a simple one-shot version of this game that is attractive as a simple tool to elicit risk tolerance among respondents with limited education, can lead to biased estimates of risk aversion due to endowment effects. We use a field experiment with a pool of young business group members with limited education to test for the potential bias associated with the initial endowment allocation. We find a highly significant endowment effect which may explain low investment levels and exaggerated measures of risk aversion where this game has been used to estimate risk aversion. We develop and test a more balanced version of the risk investment game and demonstrate that it gives less bias due to endowment effects than the standard design and a full risk design that creates an endowment effect in the opposite direction, indicating that loss aversion may not be the primary cause of the endowment effect.

Keywords: Endowment effect; loss aversion; gender difference; risky investment game; field experiment; Ethiopia.
JEL Codes: C93; D91.

---

1. Introduction

Endowment effects have been used to explain the Willingness to Accept (WTA) – Willingness to Pay (WTP) gap and the exchange asymmetries often found in settings where there are supposed to be minimal or no transaction costs (Kahneman et al., 1990; Knetsch et al., 1989; Horowitz and McConnell, 2002). The term "endowment effect" was first used by Thaler (1980) and he related this effect to the fact that losses are weighted more heavily than gains and associated this with prospect theory and loss aversion. Other reasons for the "stickiness" of endowments include "status quo bias" (Samuelson and Zeckhauser, 1988; Kahneman et al., 1991) and "anchoring effects" (Epley and Gilovich, 2001; Simonson and Drolet, 2004). Plott and Zeiler (2005, 2007) have demonstrated that with a set of "priming" tools it is possible to eliminate such WTA-WTP gaps and exchange asymmetries but such priming is an exception in the real world, and this may imply that there exist some fundamental behavioral characteristics that contribute to the widespread occurrence of such exchange asymmetries. A recent review puts emphasis on expectations affecting reference points in combination with loss aversion as explanations for the persistent phenomenon (Marzilli Ericson and Fuster, 2014). Marzilli Ericson and Fuster (2011) disentangle endowment effects from ownership and show that expectations affect reference points and may thereby trigger endowment effects. Our study provides new insights on the presence of endowment effects associated with monetary endowments and risky prospects rather than commodities when expected returns should be the same across treatments.

Many of the standard experiments in behavioral economics start by providing an initial endowment in form of money. It is generally accepted that such an endowment can have a wealth effect while endowment effects are associated with ownership of commodities. Here we investigate whether a monetary endowment can create endowment effects that are not simply wealth effects. Experiments that typically provide an initial monetary endowment include the dictator game, the ultimatum game, the trust game, the public goods game and the risky investment game of Gneezy and Potters (1997). The monetary nature of these endowments may be the reason for endowment effects being ignored by assumption. However, if endowment effects are caused by loss aversion, one should perhaps not just assume away endowment effects for money.

In the dictator game economists were initially surprised that respondents did not hold back more money and this may have taken the attention away from possible endowment effects in the dictator and ultimatum game as well. Later studies have shown that it makes a difference whether the endowment is windfall money or earned money where studies suggest that the endowment effect is stronger for earned money (Cherry et al., 2002). A similar effect has also been found for the public goods game (Muehlbacher and Kirchler, 2009).

We focus on the risky investment game of Gneezy and Potters (1997). More specifically, we use a simple one-shot version of the game that was first used by Gneezy et al.(2009). The respondents are provided an initial endowment X, of which they are free to invest any amount $0 \leq x \leq X$ and they have a 50 % probability of winning $3x$ and of losing the amount x invested.

A rational respondent behaving according to expected utility theory should be risk-neutral and invest the whole endowment in the typical small amount gambles used in such experiments (Rabin, 2000; Rabin and Thaler, 2001). However, already Binswanger (1980, 1981) revealed that experimental decisions over risky prospects are not integrated with the wealth of respondents. Most people therefore appear to be risk averse also in small gambles and particularly so if losses are included among the possible outcomes. Narrow bracketing with a combination of risk aversion and loss aversion could potentially explain why respondents exposed to the risky investment game prefer not to invest the full initial endowment received in this game.

Charness et al. (2013) noted in a review of the risky investment game that a weakness is that it does not separate risk-neutral and risk-loving respondents as both these types would invest the whole amount. The beauty of the game is, however, that it is so simple and easy to apply in the field and variants of it have for that reason become a popular tool that has been used for the study of a variety of issues such as gender differences in risk preferences (Charness and Gneezy, 2012), myopic loss aversion among students and professional traders (Gneezy and Potters, 1997; Haigh and List, 2005), and the correlation between risk-taking, testosterone levels, and facial masculinity (Apicella et al., 2008). It has also been tested and found useful in a developing country setting where respondents have limited education and numeracy skills and more complex elicitation methods such as the Holt and Laury (2002) Multiple Price List approach may be associated with more cognitive problems and inconsistent responses (Charness and Viceisza, 2016). This is also the type of environment that our research is focusing on. We hope our study can contribute to the refinement of simple tools for risk preference elicitation in a development context with non-WEIRD samples.

We combine a pilot field experiment and a large sample field experiment in rural Ethiopia to investigate the existence of endowment effects in the risky investment game.Two alternative treatments to the basic design were introduced, one providing a full risk initial lottery endowment (Treatment 2), and one balanced binary treatment (Treatment 3) that should eliminate or reduce the potential endowment effects associated with the first two treatments. We find highly significant and substantial endowment effects in our study.

One of the distinct findings with the risk investment game is that it has been associated with stronger gender differences in risk preferences than some alternative approaches to eliciting risk preferences (Charness and Gneezy, 2012; Charness et al., 2013). In our assessment of potential endowment effects in the game, we assess whether these can be an explanation for these strong gender differences. We assess this by comparing the gender differences for each of the three treatments.

The paper is organized as follows. Part 2 of the paper outlines the experimental design. Part 3 describes the sample characteristics. Part 4 presents the results and part 5 discusses the findings and draws some tentative conclusions and suggestions for further work.

2. Experimental design

We combined a pilot field experiment with a large sample field experiment. The respondents are members of youth business groups located in rural areas in northern Ethiopia. The experiments were implemented for one group at a time with up to 12 group members who played the games and were simultaneously interviewed by 12 experimental enumerators. The baseline treatment was based on the one-shot version of the game first used by Gneezy et al.(2009). Respondents are given an endowment X from which they can invest a share x/X which is tripled by the researchers $(3x/X)$ and which can be won with a 50 percent probability, or otherwise lost such that the respondent only retains $X - x$. The lucky winners obtain $X - x + 3x = X + 2x$.

The second treatment gave the respondents a lottery prospect of $3X$ with a 50 percent chance of winning. The respondents were then offered to sell all or part of the lottery prospect and would then get a payment of one-third of the lottery winning value they would sell. If they sell y out of $3X$, they will get $y/3$ as payment. Losers of the game will get $y/3$ and winners will get $3X - y + y/3$.

Against the H0 hypothesis that there is no significant difference in the amount invested in Treatments 1 and 2, which implies that there are no endowment or anchoring effects linked to the initial endowment, we test the following hypotheses:

Hypothesis 1a: *An endowment effect implies a bias towards the first endowment allocated and Treatment 2 leads to a larger investment in the risky option than Treatment 1.*

Hypothesis 1b: *A larger share of the respondents invest the full amount in Treatment 2 than in Treatment 1.*

The third treatment aimed at striking a balance between the first two treatments that both may lead to a bias towards the initial riskless or risky base options. This treatment was implemented as a set of binary choices to elicit the optimal portfolio or balance between a safe and a risky option with the same expected returns as in treatments one and two. The first binary choice is between getting X with certainty and $3X$ with a 50% probability. The preferred choice in this first binary choice is then offered in the second binary choice where the alternative choice is $X/2$ for sure and a 50% chance of winning the tripled second half of the full lottery amount (Expected value: $0.5 * 3X/2$). Further binary choices are provided till and optimal mix of safe and lottery amounts are identified. Details of the experimental protocols for the three treatments are provided in Appendix 2. Treatment 3 allows us to test the following hypotheses:

Hypothesis 2a: *Treatment 3 gives an average investment level that is larger than for Treatment 1 and lower than for Treatment 2.*

Hypothesis 2b: *The share investing the full amount in Treatment 3 is larger than for Treatment 1 and smaller than for Treatment 2*

Hypothesis 2c: *The average investment levels are the same in Treatments 2 and 3 but these are higher than in Treatment 1.*

Hypothesis 2d: *The share investing the full amount in Treatments 2 and 3 are the same but these are higher than in Treatment 1.*

Hypotheses 2c and 2d are based on the theory that the endowment effect is driven by loss aversion and an initial endowment in form of a risky prospect that maximizes the potential loss should therefore not induce an endowment effect.

The standard game has been shown to give significant gender differences with women investing significantly less than men in most earlier studies (Charness and Gneezy, 2012). It is still a mystery why this game tends to give stronger gender differences than other games used to investigate gender differences in risk preferences (Filippin and Crosetto, 2016). We wonder whether this could be associated with an endowment effect bias that may be stronger for women. We therefore want to test the following hypothesis:

Hypothesis 3: *The gender difference is stronger in Treatment 1 than in*

Treatments 2 and 3 as it is driven by the endowment effect that is stronger in Treatment 1 and stronger for women than men.

If Hypothesis 3 holds, the gender difference in investment level should be lower in Treatment 3 than in Treatment 1. If the endowment effect (in Treatment 1) is primarily driven by loss aversion, Treatment 2 should not lead to a higher investment level than Treatment 3. Treatments 2 and 3 can reveal whether the gender difference is independent of the endowment effect and loss aversion. If the endowment effect explains why women invest less in Treatment 1, and the endowment effect is a pure anchoring effect, then women should invest more than men in Treatment 2. If the gender difference in Treatment 1 is caused primarily by a gender difference in loss aversion that materializes into a stronger gender difference in the endowment effect, the gender difference should be reduced in Treatments 2 and 3. If the endowment effect is an anchoring effect that is independent from loss aversion, but still gender-specific and stronger for women, women should invest more than men in Treatment 2.

The English versions of the research protocols are included in Appendix 2. These were translated to the local language Tigrinya which was the language used in the field. The enumerators were trained with both versions and we ensured that the translations were accurate and that the enumerators understood the questions correctly and used the same exact wording in the local language for all the questions and explanations.

3. Sampling and implementation

The respondents in the experiment were sampled from rural youth business groups in northern Ethiopia. The group members were resource-poor rural youth and young adults that due to their poverty had been found eligible to join youth business groups in their home communities (*tabias*) based on their land poverty, residence, and demonstrated interest in developing a rural livelihood in their home community. The average age was 31 years and with a standard deviation of 10 years, giving more age variation than the typical student samples used in laboratory experiments. The mean level of education was five years but it varied from no education to 12 years of completed education. Still, financial and business skills are important for them to succeed in their business activities. Women constituted close to one third of the group members.

Treatment 1 was used in a baseline survey in the study area in 2016 for a sample of 1138 youth business group members in 119 business groups in five districts in the Tigray region of Ethiopia.

The initial endowment of 30 ETB used as the safe amount was equivalent to a daily rural wage rate in agriculture in the study areas in 2016. This amount was split in two 10 ETB and two 5 ETB notes which allowed investment levels of 0, 5, 10, 15, 20, 25 and 30 ETB. We wanted for practical reasons to avoid the splitting into a finer sub-division which would require the use of coins. This was also the reason for multiplying the invested amount with three rather than the 2.5 factor used in the initial Gneezy and Potters study and several other studies.

Local schools were used as field labs. One youth group was interviewed at a time with 12 enumerators doing the experiments and interviews of 12 members simultaneously. Three classrooms were used, locating an experimental enumerator and a group member in each corner of a classroom. This prevented communication between group members during the games. It also implied that the enumerators never interviewed or did experiments with more than one group member per group, thereby ensuring orthogonality between groups and enumerators, to control for and minimize potential enumerator bias in the estimation.

The low share of respondents investing the full amount in the 2016 experiment led the authors to worry that the design could lead to bias and reveal respondents as less risk tolerant than they really are. With new funding from a new project, a follow-up survey was planned in 2019. To test the hypothesis of an endowment effect, treatment 2 was implemented as a pilot test in one of the districts together with treatment 3 which should strike a balance between treatment 1 and 2 which both could create a bias towards their respective safe and risky initial endowments.

A large share of the sample in this pilot study has also participated in the 2016 experiment, thereby combining a within-subject and between-subject design. Treatments 2 and 3 were randomized at group level for a sample of youth business groups and group members (N=243 for Treatment 2 and N=304 for treatment 3) in the pilot district.

Based on the outcome of this pilot study and the comparison with the baseline treatment, it was decided to scale up the binary treatment to the full sample (N=2184).

4. Estimation strategy

The sample from 2016 received the baseline treatment (Treatment 1) while Treatments 2 and 3 were implemented in in one pilot district in 2019, and Treatment 3 was then scaled up to a large sample of youth business group members in 2019. Treatment 2 was a pilot treatment implemented for a random set of groups and members in one district. As a first robustness check we assess whether the responses in this district were systematically different from in other districts for the baseline treatment in 2016. We also run separate regression models for this pilot district with the 2016 and 2019 samples jointly and we run full sample regressions with district fixed effects to control for possible location effects that may be correlated with the treatment effects. Our design confounds year with the baseline treatments and there is a risk that the youth have changed their behavior in the baseline treatment over this three year period. We scrutinized this by including individual characteristics (sex, age, birth rank and education) and inspect whether the gain in age over the three years could have changed their responses. Another potential source of bias could be the enumerators used in the experiments. While they were doing only one interview per group each, we had a change in enumerators from 2016 to 2019 based on the quality of their work and availability (selection of the best available ones for the 2019 survey and dropping some poor performers). The inclusion of enumerator fixed effects controls for such possible selection bias. We had five enumerators that participated in both years and as an additional robustness check we run a separate model for the sample from the enumerators that were involved in both years.

The share invested from the maximum safe amount ($X = 30ETB$) is used as the dependent variable. This implies that $r = \frac{x}{X}$ and $0 \leq r \leq 1$.

We use Wilcoxon rank sum tests, also called Mann-Whitney tests (Wilcoxon, 1945; Mann and Whitney, 1947) to compare the distributions of this risk-share (r) variable across treatments. We also assessed the shares of the samples for each treatment with $r = 1$. We used Chi-square tests to compare the frequency of full investments across the treatment samples.

To further test the treatment effects and to control for other variables, we estimated linear panel data models with variants of the following specification:

$$r_{gi} = r_1 + \alpha_2 Fullrisk_g + \alpha_3 Binary_g + \alpha_{4d} D_d + \alpha_{5e} E_d + \alpha_{gs} s_{gi} + g_g + \epsilon_{gi} \quad (1)$$

where subscript g represents group, subscript i represents individual, r_1 represents the estimated share invested in the baseline treatment, α_2 captures the treatment effect for Treatment 2 as the mark-up share invested in the risky lottery, α_3 represents the treatment effect for Treatment 3 as the mark-up share invested, D_d represents a vector of district dummy variables, E_d represents a vector of enumerator dummy variables, s_{gi} represents a set of individual characteristics (sex, age, birth rank, education), g_g represents group random effects, and ϵ_{gi} represents the error term.

The following alternative specifications were estimated to test the robustness of the results:

- 1) A parsimonious model that only included the treatment dummies and the district and enumerator fixed effects on the full sample
- 2) A full sample model with additional individual controls,
- 3) A model for the sample using the same enumerators in 2016 and 2019, with additional controls
- 4) A model for the pilot district combining 2016 and 2019 data with group random effects
- 5) As d) but with group fixed effects
- 6) As d) but with individual fixed effects.

A substantial share of the groups and individuals in the pilot district was the same in the 2016 and 2019 samples. This facilitated the use of group fixed effects and individual fixed effects as additional controls for unobserved heterogeneity. These are exploited in Table 4 as further robustness checks that allow control for unobservable time-invariant within-group as well as within-subject characteristics as well.

The linear panel data models yield coefficients that are marginal effects and are convenient for that reason. Since our dependent variable is a share with values from zero to one, we also estimated fractional probit models that take this into account. We have not included the results from these models, however, because they gave marginal effects that were very close to those from the linear panel data models.

5. Results

Figure 1 shows the full sample investment distribution for all three treatments. The figure illustrates highly significant differences in distributions across the three treatments. Figure 2 shows the investment distribution for Treatment 1, comparing the pilot district (Degua Tembien) distribution with that of the full sample. Degua Tembien was the district where the pilot test of Treatments 2 and 3 took place in 2019. It can be seen that the response distribution in the pilot district is very similar to that in the full sample. Figure 3 shows the distribution of investments in the pilot district in 2019 for Treatments 2 (Full Risk) and 3 (Binary) (243 versus 304 respondents). We see that a substantially larger share invested the full amount in Treatment 2 than in Treatment 3.

Table 1 presents average shares invested out of the maximum safe amount that can be obtained for the three treatments in the full sample and in the pilot district. Table 2 assesses the statistical significance of the treatments using Wilcoxon ranksum/Mann-Whitney tests for the shares invested. Table 2 also includes tests for the gender differences for the different treatments to assess whether these gender differences exist in our sample and are sensitive to the alternative treatments.

Table 3 presents the results from linear panel data models with youth group random effects and with standard errors corrected for clustering at the youth group level. Models (1) and (2) are for the full sample, and Model (2) includes additional individual controls. Models (1) and (2) also include district fixed effects and enumerator fixed effects. Model (3) includes the sample for which the same enumerators were used in 2016 and 2019 as an extra robustness check for potential enumerator selection bias. We see that the treatment effects remain significant and robust. As found in other studies, we also find a highly significant gender effect with men investing about 5 percentage points more of the endowment than women do on average.

Table 4 presents models for the pilot district, combining the 2016 and 2019 data and imposing alternative controls for unobserved heterogeneity. Model (4) includes group random effects. Model (5) includes group fixed effects and Model (6) includes individual fixed effects. All the models included enumerator fixed effects.

Result 1: *Treatment 2 results in significantly higher average investment level than Treatment 1.*

Result 2: *Treatment 2 gives a much larger share of respondents that invest the full amount than Treatment 1.*

Results 1 and 2 imply that we cannot reject Hypotheses 1a and 1b.

Elaboration: These results can be seen by visual inspection of the distributions and the confidence intervals for each investment level in Figure 1 and

Tables 1-3. Figure 1 shows that when an initial endowment of 30 ETB is provided (Treatment 1), a much larger share of the respondents invested only 5 or 10 ETB and a much smaller share invested the full amount of 30 ETB. The finding indicates that we cannot reject Hypothesis 1a that there is an endowment effect causing the respondents to invest less on average in Treatment 1 than in Treatment 2. Furthermore, we cannot reject Hypothesis 1b that Treatement 1 is associated with a much smaller share investing the full amount (10.1% of the sample) than Treatment 2 (37.4% of the sample, see Table 2).

Table 1: Mean shares invested out of the maximum safe amount for alternative treatments and samples

	— Full sample —			— Pilot district —			Same enumerators		
Treatment	Mean	St.Err	N	Mean	St.Err	N	Mean	St.Err	N
T1:Safe Base	0.443	0.007	1138	0.425	0.015	249	0.460	0.011	487
T2:Full Risk	0.691	0.021	243	0.691	0.021	243	0.609	0.035	102
T3:Binary	0.565	0.007	2184	0.611	0.019	330	0.560	0.011	898
Total	0.535	0.005	3565	0.578	0.012	822	0.530	0.008	1487

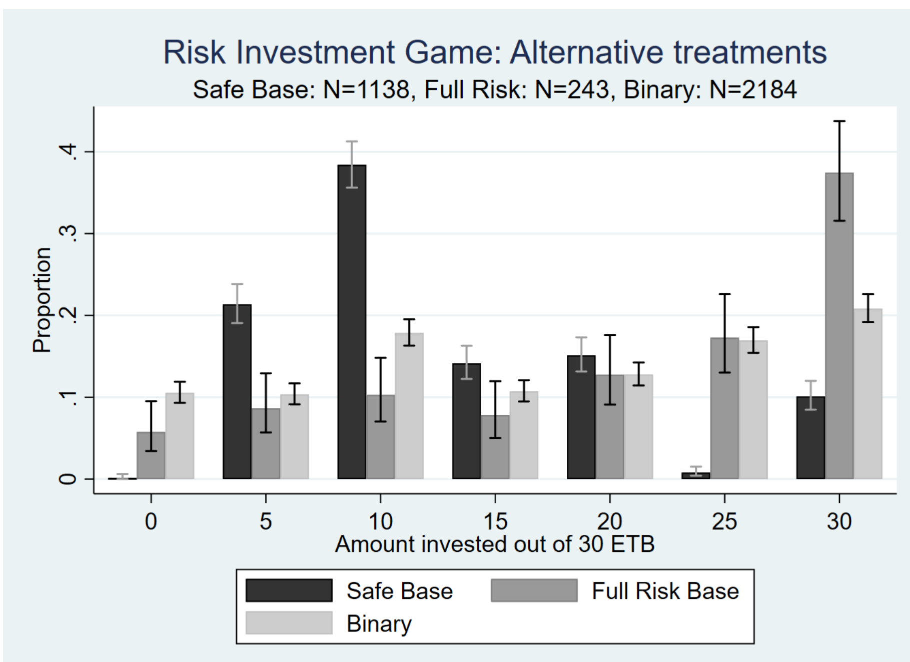

Figure 1: Distribution of investments in Treatments 1, 2 and 3 (full sample)

Result 3: *Treatment 3 resulted in significantly higher average investment level than Treatment 1 and lower average investment level than Treatment 2.*

Table 2: Treatment effects and gender differences

Treatment	Variable	Full sample Mean	Full sample St.err.	Full sample N	Males Mean	Males St.err.	Males N	Females Mean	Females St.err.	Females N	Wilcoxon p-value	Chi-sq. p-value
T1	Average share invested	0.443	0.007	1138	0.463	0.009	779	0.399	0.012	359	0.0002	
Safe Base	Share invest Full amount	0.101	0.009	1138	0.121	0.012	779	0.058	0.012	359		0.001
T2	Average share invested	0.691	0.021	243	0.704	0.026	158	0.669	0.035	85	0.31	
Full Risk	Share invest Full amount	0.374	0.031	243	0.405	0.039	158	0.318	0.051	85		0.179
T3	Average share invested	0.565	0.007	2184	0.576	0.009	1510	0.540	0.013	674	0.014	
Binary	Share invest Full amount	0.208	0.009	2184	0.225	0.011	1510	0.171	0.015	674		0.004
All	Average share invested	0.535	0.005	3565	0.548	0.007	2447	0.504	0.009	1118	0.0002	
	Share invest Full amount	0.185	0.007	3565	0.204	0.008	2447	0.146	0.011	1118		0.000

	Tests for shares invested	Full sample z-score	Full sample P-value	Degua Tembien z-score	Degua Tembien P-value	Same enumerator z-score	Same enumerator P-value
T1 vs. T2	Wilcoxon-Mann-Whitney	-10.965	0.0000	-9.078	0.0000	-3.993	0.0001
T2 vs. T3	Wilcoxon-Mann-Whitney	5.744	0.0000	2.770	0.0056	1.425	0.1542
T1 vs, T3	Wilcoxon-Mann-Whitney	-10.487	0.0000	-6.448	0.0000	-5.321	0.0000

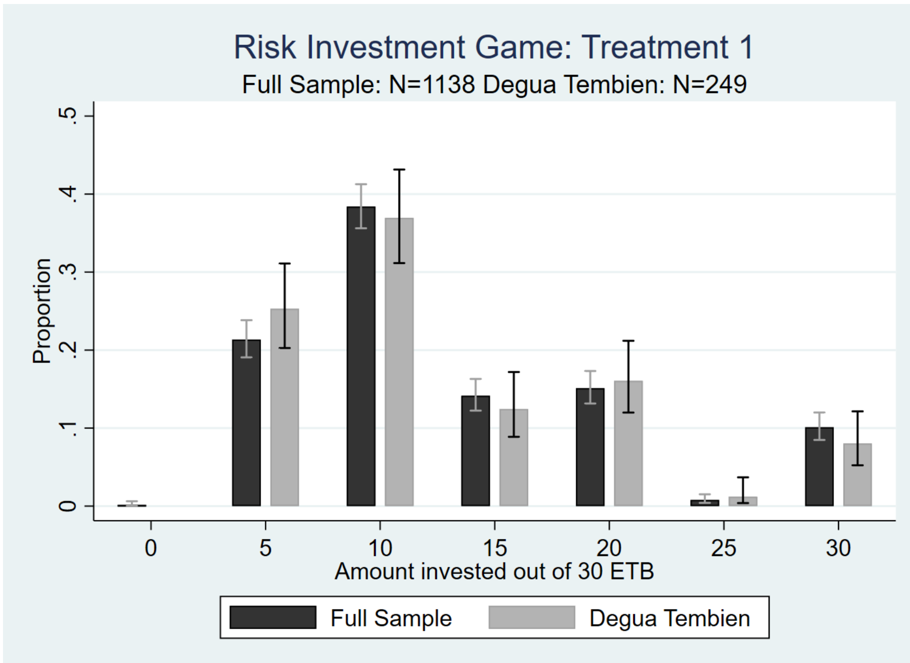

Figure 2: Robustness check (Treatment 1): Pilot district vs full sample

Result 4: *Treatment 3 gave a share investing the full amount that is significantly larger than that for Treatment 1 and significantly smaller than that of Treatment 2.*

Results 3 and 4 imply that we cannot reject Hypotheses 2a and 2b.

Elaboration: We see from Figures 1 and 3 that the share keeping the full lottery is reduced significantly when the more balanced binary treatment is used as compared to Treatment 2. This may also be interpreted as evidence of an endowment or anchoring effect towards what is initially provided. Treatment 2 may, therefore, give estimates of excessive risk tolerance among respondents. Treatment 3 may strike a balance and be less biased due to the endowment effects and thus give the basis for less biased measures of risk aversion.

The Wilcoxon–Mann-Whitney ranksum test results for a comparison of the investment levels across treatments for the full, the pilot district, and the same enumerators samples are presented at the bottom of Table 2. The differences between Treatment 1 versus Treatments 2 and 3 were highly significant in all samples. The differences between Treatments 2 and 3 were highly significant for the full and the pilot district samples. However, for the sample utilizing only the same enumerators in 2016 and 2019 the p-value was only 0.15. The insignificant test result is associated with the small sample (N=102) that received Treatment 2 and that used the same enumerators in 2016 and 2019. The lack of significance is thus primarily due to weak statistical power (high standard error) as can be

Table 3: Full sample and same enumerator models with controls

VARIABLES	(1) Full sample	(2) Full sample	(3) Same enumerators
Full Risk	0.206***	0.210***	0.161***
	(0.029)	(0.029)	(0.039)
Binary	0.096***	0.101***	0.111***
	(0.018)	(0.018)	(0.019)
Male dummy		0.046***	0.054***
		(0.012)	(0.018)
Age		-0.000	-0.002**
		(0.001)	(0.001)
Birth rank		0.005**	0.005
		(0.002)	(0.004)
Education (years)		0.006***	0.002
		(0.001)	(0.002)
Constant	0.376***	0.315***	0.396***
	(0.024)	(0.033)	(0.046)
Observations	3,565	3,565	1,487
Number of youth groups	308	308	305

All models with district FE and enumerator FE
Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 4: Robustness checks for pilot district (Degua Tembien) sample

VARIABLES Panel controls	(4) riskshare Group RE	(5) riskshare Group FE	(6) riskshare Individual FE
Full Risk	0.185***	0.197***	0.174**
	(0.040)	(0.047)	(0.067)
Binary	0.105***	0.102**	0.114*
	(0.039)	(0.047)	(0.067)
Constant	0.429***	0.430***	0.410***
	(0.045)	(0.044)	(0.030)
Observations	822	822	822
R-squared		0.141	0.292
Number of groups	53	53	
Number of individuals			593

All models with enumerator FE. Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

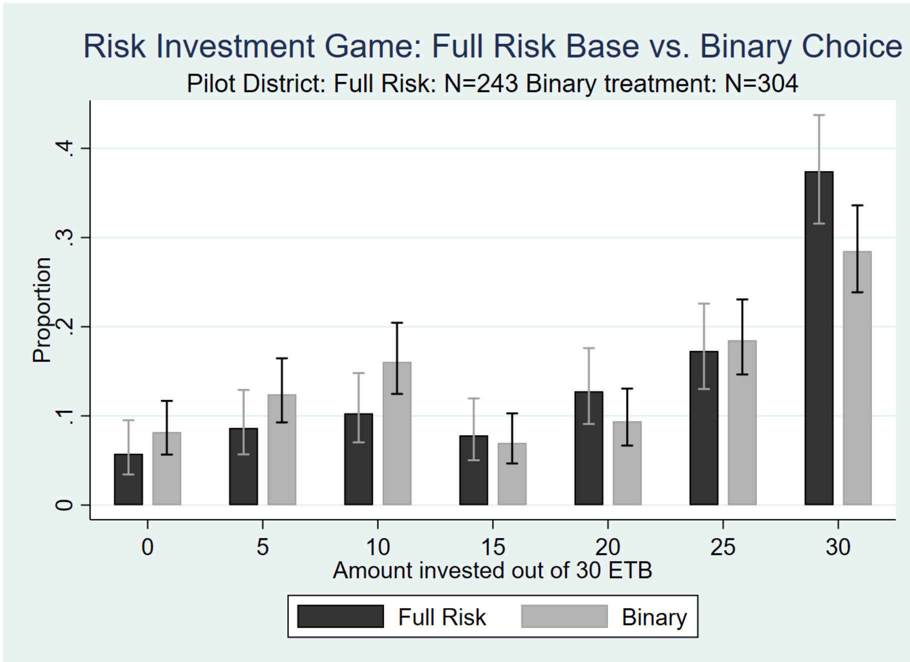

Figure 3: Robustness check Treatment 2 (Full Risk) and Treatment 3 (Binary) in pilot district

seen from Table 1.

Table 3 demonstrates that the treatment effects are robust to the inclusion of additional controls. The individual control variables were also assessed for their systematic variation across treatments, see Appendix Table A1. As Treatment 1 was implemented in 2016 it is not surprising to find a significant age difference between Treatment 1 versus Treatments 2 and 3. Age had, however, very limited effect on the investment levels as can be seen in Table 3. Age is insignificant in Model (2) and significant at 5 percent level in Model (3) but with a very low coefficient. Five years higher age is associated with a 1 percentage point lower investment share. The difference in age cannot therefore explain the large differences in investment levels between Treatment 1 versus Treatments 2 and 3. The age effect even points in opposite direction of the change in mean investment levels in 2016 compared to 2019, when the group members have become three years older.

To assess whether the gender differences in the game can be associated with gender-differentiated endowment effects, we assessed the gender differences within each treatment. We compared the means as well as the likelihood that the whole amount was invested across the treatments. The gender differences are presented in Table 2 with Wilcoxon-Mann-Whitney ranksum tests for differences in means and with Chi-square tests for the difference in the probability that the whole amount is invested.

Result 5: *The gender difference in average investment level remains robust and goes in the same direction across the three treatments.*

Elaboration: Result 5 implies that we can reject Hypothesis 3. The strong gender difference in average investment levels is not explained by the endowment effect associated with the initial endowment provided in the game. And, the endowment effect is not driven by loss aversion as Treatment 2, which endows respondents with a risky prospect with high potential loss, also creates an endowment effect. Most of the gender difference in the responses is due to other things than a gender-differentiated endowment effect. For Treatment 2 there was no significant gender difference but the reason for this is related to the weak statistical power due to the smaller sample. Table 3 also demonstrates a highly significant gender difference in investment level in the game. The effect is not reduced when the regression is run for the reduced sample (Model (3) in Table 3) which used the same enumerators in 2016 and 2019. Using t-tests for the gender difference in mean investment levels across treatments revealed no significant differences in the gender differences (detailed results available upon request).

To further inspect the robustness of the results, the pilot district sample is used with alternative controls, see models (4) - (6) in Table 4. We utilize the fact that for this district many of the same youth groups and group members were included in the 2016 as well as 2019 samples. This allows us to impose stronger controls for unobserved time-invariant heterogeneity through the use of group fixed effects or individual fixed effects. We see from Table 4 that the treatment effects were robust to these alternative specifications. Treatments 2 and 3 give significantly larger average investment levels than Treatment 1 in all model specifications and the investment levels are 17.4-19.7 percentage points higher for Treatment 2 than Treatment 1 and 10.2-11.4 percentage points higher for Treatment 3 than for Treatment 1.

6. Discussion

We will now discuss our findings in light of findings in other relevant studies. While variants of the risky investment game of Gneezy and Potters (1997) have been a very popular tool both in lab and field experiments, we are not aware of any earlier studies that have investigated the possible endowment effects associated with the initial endowment provided in the game. This may be because the game initially was associated with loss aversion, more specifically myopic loss aversion and repeated versions of the game that are more likely to trigger loss aversion. It is less obvious that the one-shot version of the game invokes loss aversion as it may be perceived to operate in the gains domain only. However, if the initial endowment leads to an immediate shift of the reference point, investing part of the endowment can be perceived to be in the loss domain and thereby invoke loss aversion.

Marzilli Ericson and Fuster (2011) showed that expectations could affect the size of the endowment effects by manipulating the probability of being able to keep a commodity that was initially given to respondents. A higher probability of being able to keep the commodity was associated with a stronger endowment effect. They concluded that endowment effects are real but operate via expectations instead of formal ownership. Unlike their experiment, our alternative treatments do not change the rational expected returns across our treatments. And our experiment did not include any commodity, only monetary prospects, and still we found highly significant "endowment effects" associated with the prospect that was first allocated to the respondents. Such an "endowment effect" was even realized when there was a high probability of loss associated with the initial prospect. This may appear puzzling also in the light of the findings of Marzilli Ericson and Fuster (2011).

What makes it particularly interesting to study endowment effects with the risk investment game is that endowment effects have been attempted explained by loss aversion (Thaler 1980; Kahnemann et al. 1991). We may critically ask why a lottery prospect like Treatment 2 invokes an endowment effect if this endowment effect is explained by loss aversion? The more loss averse should then sell themselves out of the risky lottery prospect and not be triggered to become more likely to keep (more of) the lottery prospect. Our results therefore appear to contradict basic endowment effect theory. This may indicate that the endowment effect, rather than being explained by loss aversion, is explained by other things.

We argued that loss aversion is also relevant for money and could lead to an endowment effect for money for that reason. Our study is in a rural economy where cash is scarce and this could potentially enhance the endowment effect for money and possibly also risk aversion and loss aversion due to narrow bracketing. However, the "endowment effect" we find may rather be a form of starting point bias or attachment to the initial prospect that may have manipulated the subjective reference points to differ from the rational expected returns than be driven by the potential associated loss (in Treatment 2).

Our results show that endowment effects can introduce biases that will cause upward bias in estimates of risk aversion from the one-shot version of the game. Already Charness and Gneezy (2004) found that the game was sensitive to framing effects. Our Treatments 1 and 2 tested for endowment effects by alternatively endowing respondents with a safe amount that can be invested or with a lottery prospect that the respondents may sell themselves out of at an identical price in the two treatments. Endowing the respondent with this lottery prospect treatment dramatically increased the share of the respondents that accepted the full lottery as well as the average share invested in the game.

An implication, however, is that both these treatments result in bias due to endowment effects but in opposite directions. We tested for this with our binary balanced treatment and find highly significant indications of such endowment or anchoring effects in direction of what is initially provided when comparing our three treatments. These effects are robust to the alternative econometric specifications and robustness checks.

Given the tractability of the one-shot game for field experiments, it is of general interest to know whether the one-shot version of the game can be used to generate unbiased estimates of risk aversion or risk tolerance.

While most studies using the game have not used the game to estimate a parameter for the risk aversion (capturing the curvature of the utility function) of respondents, this may be tempting and there are a few studies that also do this (e.g. Crosetto and Filippin 2016; Dasgupta et al. 2019). An implication of our study may be that those studies that started by first providing a safe endowment have overestimated the degree of risk aversion in their samples. It is possible that the endowment framing has caused risk-neutral persons to behave as if they were risk averse in the game. However, if also the one-shot game initiates an immediate change in the reference point and thereby loss aversion the behavior in the game may be due to a combination of a concave utility function in gains and loss aversion associated with losses. Loss aversion in combination with a linear utility function should lead to "bang-bang" solutions. This is far from what we observe for all three treatments as the share of interior solutions was close to 0.9 for Treatment 1, about 0.58 for Treatment 2 and close to 0.7 for Treatment 3. Table 2 demonstrated large differences in the shares that invested the full amount in the alternative treatments. This implies that the endowment effect is caused by other things than loss aversion, such as psychological transaction costs. Alternatively, moderate changes in loss aversion and near linear utility curvature contribute to explain the substantial variation we find in the shares investing the full endowment that we observe in Figure 1 and Table 2.

The fact that the game does not distinguish between risk-loving, risk-neutral and slightly risk averse individuals has been considered as a weakness of the design (Crosetto and Filippin 2016). This weakness is also shared with the Binswanger (1980; 1981) and closely related Eckel and Grossman (2002, 2008) games. These have been popular methods for the estimation of risk preferences of poor people in developing countries (Binswanger and Sillers, 1983; Wik et al., 2004; Yesuf and Bluffstone, 2009). Concerns have been raised recently that these methods have resulted in biased perspectives on the distribution of risk preferences in the populations studied with these approaches and that they have underestimated the share of the populations with near risk-neutral and risk-loving preferences (Vieider, 2018; Vieider et al., 2019).

We implement a very simple calibration exercise to explore the relationship between the choices made in the game after endowment effects are eliminated, assuming this is the case in Treatment 3. We use a Constant Partial Relative Risk Aversion ($CPRRA$) utility function; $U = (1-r)^{(-1)}Y^{(1-r)}$. We assume no or limited asset integration and no or moderate levels of loss aversion. With maximization of expected utility and utility being based only on the outcomes in the game, the $CPRRA - r$ coefficient must be very high to lead to optimal investment levels at or below $10\,ETB$ (38% of the sample investing below 33% of the endowment). Alternatively, invoking loss aversion and a linear utility function leads to bang-bang solutions and a switch from investing $30\,ETB$ to $0\,ETB$ when the loss aversion parameter increases to 2 or higher. However, a combination of a reasonably sized $CPRRA - r$ and a lower level of loss aversion leads to intermediate solutions that also easily can fall in the $1 - 10\,ETB$ range as can be seen in Table 5. We suggest that "weak loss aversion" can be the

Table 5: Alternative utility functions and optimal investment level

$CPRRA - r$	Asset integration	Loss aversion coeff.	Optimal investment amount (ETB)
0.1	No	1	29.9
0.2	No	1	27
0.5	No	1	15
0.7	No	1	11
0.95	No	1	8
0.1	No	1.9	4
0.2	No	1.8	4
0.2	No	1.3	17
0.5	No	1.3	8
0.5	No	1.5	5
0.8	No	1.5	3
0.1	30 ETB base	1.9	8
0.2	30 ETB base	1.9	4
0.2	30 ETB base	1.3	30
0.5	30 ETB base	1.3	16
0.5	30 ETB base	1.5	10
0.8	30 ETB base	1.5	6

result of a weak endowment effect that is invoked for the binary choices. This would be consistent with the distribution we see. However, this is an area for possible future investigation. Partial asset integration in combination with weak risk aversion and weak loss aversion may also yield intermediate outcomes in the full range of outcomes observed in the Treatment 3 and as calibrated in Table 5.

There are some important variations in the probabilities of winning as well as the multiplier for the winning outcome compared to the initial game by Gneezy and Potters (1997) who used it to study the existence of myopic loss aversion among students and traders. These variations imply that caution has to be applied when making comparisons across studies. The initial game included a 1/3 probability of winning and a 2/3 probability of losing and the loss was equal to the amount invested and the amount won was 2.5 times the amount invested.

Our treatments had the same probabilities and multiplier as that of Gneezy et al. (2009), Gong and Yang (2012) and Dasgupta et al. (2019) who used $50-50$ chances of winning and losing and a $3x$ factor for the amount won. We narrow in our comparisons to these studies that have used the same probabilities and mulitpliers as we used in our study.

Gneezy et al. (2009) applied this design of the game in field experiments in a matrilineal society in India and a patrilineal society in Tanzania. They found that on average women in the matrilineal society invested 87% and men 61% of the initial fund and that women in the patrilineal society invested 61% and men 85%. This can be compared to 40% for women and 46% for men in our baseline treatment which indicates much lower average investment levels and a smaller average gender difference. Our large sample makes this gender difference highly significant, however.

Gong and Yang (2012) applied the game in a field experiment in rural China in a matrilineal and a patrilineal society and to compare risk preferences of men and women. Even with such a more favorable expected return function, compared to the initial design, the average share of the initial endowment invested was as low as 4% for the women in the patrilineal group against 37% for men in the same group and against 33% and 54% for women and men in the matrilineal group. Their samples from each group were quite small. The investment levels they found in the matrilineal society are similar to ours but their gender difference is larger. Among the men in the matrilineal society they found a substantial share (about 30%) that invested the full endowment. This compares to the average of 12% for men and 6% for women in our baseline (Treatment 1) sample. However, it increased to 41% and 32% for our Treatment 2 (Full Risk), and to 23% and 17% for the Binary treatment. This demonstrates the endowment effects in Treatments 1 and 2 while Treatment 3 may not suffer from such bias. However, more research is needed to investigate this further.

The study by Dasgupta et al. (2019) used a large sample of 2000 students from India and found that only 6% of the males and 1% of the females invested the full amount. On average males invested 52% of their endowment and females 44%, demonstrating a highly significant gender effect as well as average levels of investment not very different from our baseline treatment.

There are a few studies that have compared the risk investment game with other alternative risk preference elicitation methods. Charness and Viceisza (2016) compare the game with the well-known Holt and Laury (2002) approach and an un-incentivized survey-based approach (Willingness-to-take-risk) in a field experiment in Senegal. While they used the 50% and $3x$ winning factor they used a risky seed framing in their experiment rather than money. This framing may also have had an effect on the responses compared to if they had used money. They found that the respondents have more cognitive problems with the Holt and Laury (2002) approach which gave more inconsistent responses. The risky investment game performed better as it was easier to comprehend and had better predictive power. They compared their study with a study by Dave et al. (2010) that also found that simpler games may be preferable for respondents with limited numeracy skills. We recommend further studies that compare alternative methods and their predictive power in field settings where respondents have limited numeracy skills.

Further research is needed on how alternative framing of the this game and alternative games used to elicit risk preferences affect respondents' reference points. The usefulness of the games depend on their predictive power. The predictive power of a game may depend on the type of real world phenomena and decisions they are used to predict. Perhaps the risky investment game is better at predicting whether to invest and how much to invest decisions while the Holt and Laury game is better for predicting choices among alternative risky prospects? Perhaps endowment effects are part of the explanation, in addition to limited asset integration, for the puzzling high risk aversion in small gambles (Rabin 2000).

Just using the investment share as an indication of risk tolerance may be less problematic as it does not force the measure to be explained by utility curvature alone. Even it may be acceptable to allow it to include endowment effects if this endowment effect is associated with loss aversion. Further studies are needed, however, to test the correlation between this endowment effect and measures of loss aversion obtained through alternative experimental methods in within-subject designs.

7. Conclusion

The one-shot version of the Gneezy and Potters(1997) risky investment game has gained popularity and has been proposed as particularly useful in field settings with respondents with limited numeracy skills (Charness and Viceizca 2016). We have investigated whether the game can lead to biased estimates of risk aversion due to inherent endowment effects. We found strong evidence of such endowment effects associated with the initial endowment allocated in the game. Our results indicate that the endowment effect is not a result of loss aversion as respondents endowed with a risky prospect with high probability of being lost also created an endowment effect. We suggest that the binary version of the game that we used may give less biased measures of risk aversion than the game starting by allocating an endowment that respondents are free to invest some or all of. We recommend more research to further refine the tool and test its predictive power with alternative incentives and framing and to further investigate its correlation with separate measures of loss aversion and risk aversion (utility curvature).

References

Apicella, C. L., Dreber, A., Campbell, B., Gray, P. B., Hoffman, M., and Little, A. C. (2008). Testosterone and financial risk preferences. *Evolution and Human Behavior*, 29(6):384–390.

Binswanger, H. P. (1980). Attitudes toward risk: Experimental measurement in rural india. *American Journal of Agricultural Economics*, 62(3):395–407.

Binswanger, H. P. (1981). Attitudes toward risk: Theoretical implications of an experiment in rural india. *Economic Journal*, 91(364):867–890.

Binswanger, H. P. and Sillers, D. A. (1983). Risk aversion and credit constraints in farmers' decision-making: A reinterpretation. *The Journal of Development Studies*, 20(1):5–21.

Charness, G. and Gneezy, U. (2004). Gender, framing, and investment. Technical report, mimeo.

Charness, G. and Gneezy, U. (2012). Strong evidence for gender differences in risk taking. *Journal of Economic Behavior & Organization*, 83:50–58.

Charness, G., Gneezy, U., and Imas, A. (2013). Experimental methods: Eliciting risk preferences. *Journal of Economic Behavior & 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(2):145–171.

Cherry, T. L., Frykblom, P., and Shogren, J. F. (2002). Hardnose the dictator. *American Economic Review*, 92(4):1218–1221.

Dasgupta, U., Mani, S., Sharma, S., and Singhal, S. (2019). Can gender differences in distributional preferences explain gender gaps in competition? *Journal of Economic Psychology*, 70:1–11.

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(3):219–243.

Eckel, C. C. and Grossman, P. J. (2002). Sex differences and statistical stereotyping in attitudes toward financial risk. *Evolution and human behavior*, 23(4):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 Behavior & Organization*, 68(1):1–17.

Epley, N. and Gilovich, T. (2001). Putting adjustment back in the anchoring and adjustment heuristic: Differential processing of self-generated and experimenter-provided anchors. *Psychological Science*, 12(5):391–396.

Filippin, A. and Crosetto, P. (2016). A reconsideration of gender differences in risk attitudes. *Management Science*, 62(11):3138–3160.

Gneezy, U., Leonard, K. L., and List, J. A. (2009). Gender differences in competition: Evidence from a matrilineal and a patriarchal society. *Econometrica*, 77(5):1637–1664.

Gneezy, U. and Potters, J. (1997). An experiment on risk taking and evaluation periods. *The Quarterly Journal of Economics*, 112(2):631–645.

Gong, B. and Yang, C.-L. (2012). Gender differences in risk attitudes: Field experiments on the matrilineal mosuo and the patriarchal yi. *Journal of Economic Behavior & Organization*, 83(1):59–65.

Haigh, M. S. and List, J. A. (2005). Do professional traders exhibit myopic loss aversion? an experimental analysis. *The Journal of Finance*, 60(1):523–534.

Holt, C. A. and Laury, S. K. (2002). Risk aversion and incentive effects. *American Economic Review*, 92(5):1644–1655.

Horowitz, J. K. and McConnell, K. E. (2002). A review of wta/wtp studies. *Journal of Environmental Economics and Management*, 44(3):426–447.

Kahneman, D., Knetsch, J. L., and Thaler, R. H. (1990). Experimental tests of the endowment effect and the coase theorem. *Journal of Political Economy*, 98(6):1325–1348.

Kahneman, D., Knetsch, J. L., and Thaler, R. H. (1991). Anomalies: The endowment effect, loss aversion, and status quo bias. *Journal of Economic Perspectives*, 5(1):193–206.

Knetsch, J. L. et al. (1989). The endowment effect and evidence of nonreversible indifference curves. *American Economic Review*, 79(5):1277–1284.

Mann, H. B. and Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*, pages 50–60.

Marzilli Ericson, K. M. and Fuster, A. (2011). Expectations as endowments: Evidence on reference-dependent preferences from exchange and valuation experiments. *The Quarterly Journal of Economics*, 126(4):1879–1907.

Marzilli Ericson, K. M. and Fuster, A. (2014). The endowment effect. *Annual Review of Economics*, 6(1):555–579.

Muehlbacher, S. and Kirchler, E. (2009). Origin of endowments in public good games: The impact of effort on contributions. *Journal of Neuroscience, Psychology, and Economics*, 2(1):59.

Plott, C. R. and Zeiler, K. (2005). The willingness to pay-willingness to accept gap, the" endowment effect," subject misconceptions, and experimental procedures for eliciting valuations. *American Economic Review*, 95(3):530–545.

Plott, C. R. and Zeiler, K. (2007). Exchange asymmetries incorrectly interpreted as evidence of endowment effect theory and prospect theory? *American Economic Review*, 97(4):1449–1466.

Rabin, M. (2000). Risk-aversion for small stakes: A calibration theorem. *Econometrica*, 68:1281–1292.

Rabin, M. and Thaler, R. H. (2001). Anomalies: risk aversion. *Journal of Economic Perspectives*, 15(1):219–232.

Samuelson, W. and Zeckhauser, R. (1988). Status quo bias in decision making. *Journal of Risk and Uncertainty*, 1(1):7–59.

Simonson, I. and Drolet, A. (2004). Anchoring effects on consumers' willingness-to-pay and willingness-to-accept. *Journal of Consumer Research*, 31(3):681–690.

Thaler, R. (1980). Toward a positive theory of consumer choice. *Journal of Economic Behavior & Organization*, 1(1):39–60.

Vieider, F. M. (2018). Violence and risk preference: experimental evidence from afghanistan: comment. *American Economic Review*, 108(8):2366–82.

Vieider, F. M., Martinsson, P., Nam, P. K., and Truong, N. (2019). Risk preferences and development revisited. *Theory and Decision*, 86(1):1–21.

Wik, M., Aragie Kebede, T., Bergland, O., and Holden, S. T. (2004). On the measurement of risk aversion from experimental data. *Applied Economics*, 36(21):2443–2451.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. biom bull 1: 80–83.

Yesuf, M. and Bluffstone, R. A. (2009). Poverty, risk aversion, and path dependence in low-income countries: Experimental evidence from ethiopia. *American Journal of Agricultural Economics*, 91(4):1022–1037.

8. Appendix

Appendix Table A1. Individual characteristics by treatment: t-tests

	T1 Safe Base	T2 Full Risk	T3 Binary	Total	t-tests T1 vs T2	t-tests T1 vs. T3	t-tests T2 vs. T3
Age, years	29.07	32.78	32.24	31.27	-3.710***	-3.170***	0.540
	(9.796)	(9.216)	(9.507)	(9.696)	(0.685)	(0.351)	(0.641)
Birth rank	3.105	3.198	3.37	3.273	-0.093	-0.265***	-0.172
	(2.002)	(1.877)	(2.183)	(2.110)	(0.140)	(0.078)	(0.146)
Education, years	5.345	5.078	4.608	4.875	0.267	0.737***	0.470
	(3.978)	(3.747)	(3.968)	(3.970)	(0.278)	(0.145)	(0.267)
Observations	1138	243	2184	3565	1381	3322	2427

Appendix Table A2. Robustness checks for pilot district: With individual controls

	(1)	(2)	(3)
VARIABLES	riskshare	riskshare	riskshare
Panel controls	Group RE	Group FE	Individual FE
Full Risk	0.188***	0.198***	0.174**
	(0.041)	(0.047)	(0.067)
Binary	0.105***	0.100**	0.114*
	(0.040)	(0.048)	(0.067)
Male, dummy	0.041*	0.035	
	(0.025)	(0.027)	
Age	-0.001	-0.000	
	(0.001)	(0.002)	
Birth rank	0.009	0.008	
	(0.006)	(0.006)	
Education, years	0.006**	0.006*	
	(0.003)	(0.003)	
Constant	0.359***	0.355***	0.410***
	(0.065)	(0.068)	(0.030)
Observations	822	822	822
R-squared		0.149	0.292
Number of groups	53	53	
Number of individuals			593

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1