Is gamification a curse or blessing for the design of riskelicitation methods in the field? Experimental evidence fromCambodian smallholder farmers

Selina Bruns*^, Daniel Hermann*, and Oliver Mußhoff*
***University of Goettingen**
^Corresponding author: Selina.bruns@uni-goettingen.de

Selected Paper prepared for presentation at the 2022 Agricultural & Applied Economics Association Annual Meeting, Anaheim, CA; July 31-August 2

Is gamification a curse or blessing for the design of risk elicitation methods in the field? Experimental evidence from Cambodian smallholder farmers

Selina Bruns* Daniel Hermann Oliver Musshoff

Department of Agricultural Economics and Rural Development, University of Goettingen

*Corresponding author, email: selina.bruns@uni-goettingen.de

May 2022

[AAEA Working Paper]

Abstract

Risk attitude is a key element of the vicious circle of poverty. Reliable instruments to elicit risk attitudes of the rural poor in developing countries are therefore essential to contribute in combating poverty. Many sophisticated risk elicitation methods require numeric skills, which the rural poor often lack due to low education levels. Thus, results might be inaccurate due to low levels of understanding of the task. To contribute in identifying proper measures, this paper empirically analyzes two risk elicitation experiments, the well-established Holt and Laury task (HL-task), a novel Wheel-task, and one self-assessment question. The Wheel-task is a novel elicitation method designed by the authors to be particularly easy to understand and entertaining for the participant to motivate participation. It follows the wheel of fortune model, where a spin on the wheel determines ones payout, while always having the option to choose between spinning the wheel and a safe payout. The data was collected among 247 smallholder farmers from rural Cambodia. We find i. high inconsistencies for the HL-task and on average risk averse choice patterns, ii. relatively low inconsistency levels and on average risk seeking behavior for the Wheel-task and iii. slightly risk seeking self-assessments of participants. While the Wheel-task results in less noisy outcome than the HL-task, we conclude that reasons like enthusiasm for gambling might drive the results in the Wheel-task.

Introduction

According to the most recent statistics, around 767 million people live in poverty. The vast majority of these global poor reside in rural areas, depending to some degree on agriculture as their source of livelihood - in particular as smallholder farmers (World Bank, 2018). These smallholders operate under particularly risky conditions, e.g. due to unstable environmental, market, and household conditions (World Bank, 2016, 2020). Furthermore, the risk involved in economic decision making is extraordinary for poor small farm managers, as investments with respect to e.g. crop cultivation and input use are crucial for the livelihood of the household. However, to interrupt the downward spiral of poverty, individuals need to take the risk of investing into human, physical or social capital, which can be a difficult endeavor without any safety net. Individuals that are risk averse to the extent that they refuse to invest into new technologies will be less able to cope with shocks, and hence, might end up in chronic poverty (Mosley and Verschoor, 2005). Thus, risk attitude analysis is a crucial step in contributing to combating poverty, as supporting custom-fit policy interventions can be derived.

Critical to this analysis is having a reliable instrument to elicit risk attitude. In recent years, various experimental methods, such as playing lotteries, investment games, or rather intuitive games, have been tested in the lab or the field to understand how individuals decide under uncertainty (Charness et al., 2013). Nevertheless, the methods being used to elicit risk attitudes often involve complex and abstract tasks. For example, praised as the "gold standard", the HL-task -a very precise, yet complex binary lottery (Charness et al., 2013)- is often the first choice, even when measuring risk attitudes among low-literacy subjects (Verschoor et al., 2016). Literature suggest that complex tasks are not understood by the poor - who often lack basic schooling - thus yielding flawed data (Brick et al., 2012).

Therefore, two essential questions emerge. First, which method(s) should a researcher choose when aiming to elicit risk preferences of low-literacy subjects. As this is by far not a new question, many researchers have approached it by executing a battery of complex and simple (incentivized and non-incentivized) methods in the field, conducting a within-subject analysis (Charness et al., 2013; He et al., 2018; Crosetto and Filippin, 2016). They typically find low consistency among the answers given in the different experimental tasks. However, when contrasting the within-subject results with real-life behavior, studies do find hints for a general rule of thumb. In synthesis, simple-incentivized methods, or those with better illustration, have been shown to be easier to understand by low-literacy individuals (Charness et al., 2013, 2020; Verschoor et al., 2016; Dave et al., 2010) compared to complex methods. Also simple survey measures (such as Dohmen et al. (2011)) appear to be a fair indicator of risk attitude (Dohmen et al., 2011; Charness and Viceisza, 2016). However, in simple-incentivized methods it is more difficult, if not impossible, to detect confusion (Charness and Viceisza, 2016) as well as to disentangle risk neutral from risk seeking behavior (Crosetto and Filippin, 2016). Thus, while simple methods might be better understood, they tend to lack the level of precision needed (Charness et al., 2013).

Second, if complex, yet finer, tasks might overtax low-literacy subjects and simpler, yet courser, methods might be more applicable but less precise - then which elements should be included in the design of a successful risk elicitation method for low-literacy subjects? Many researchers have identified and approached this question, mainly by either developing easier tasks (e.g. Lejuez et al. (2002); Crosetto and Filippin (2013)) or by employing better illustration of complex tasks (Ihli et al., 2013). For example, Lejuez et al. (2002) developed a simple, non-numeric method, which incorporates elements of gamification, the Baloon Analogue Risk Task (BART). They put forward the balloon-task, where the participant can pump air into a balloon and receives a certain amount per pump. If the balloon pops, she looses everything. She can always decide to stop pumping and go home with the money earned. Incorporating gamification, i.e. using game mechanics in a non-gaming situation (Deterding et al., 2011), can drive motivation of the participant, as well as her engagement and her participation (Sailer, 2016). However, different game mechanisms can have different effects, also depending on the context. Thus, while incorporating gamification might hold considerable benefits, there is little to no evidence how game mechanisms incorporated in risk elicitation methods effect low-literacy subjects.

In this study we investigate weather incorporating an element of gamification in a precise risk-elicitation measure can be beneficial for the risk analysis among low-literacy subjects. To do so, we develop a novel task: The Wheel-task. It is designed to be particularly easy to understand and entertaining for the participant to motivate attentive participation. It follows the wheel of fortune model, where a spin on the wheel determines ones payout, while always having the option to choose between spinning the wheel and a safe payout. In particular, the game is not played in a digital version (in contrast to other popular simple tasks such as Lejuez et al. (2002); Crosetto and Filippin (2013), but with a physical wheel, which is useful for rural settings; and we designed it so that we can calculate the constant relative risk aversion (CRRA) values, which makes it comparable to the HL-task and thus rather precise. The data was collected among 247 smallholder farmers from rural Cambodia. Clustered as a least developed country (United Nations, 2018), Cambodia is among the poorest countries in the region and thus presents itself as a highly fitting study region for our investigation.

Notably, we do not investigate what measure is better in predicting risk attitude but whether risk preferences measured between incentivized experimental risk measures and survey measures are consistent. In particular, we undertake a horse-race between a complex incentivized (HL-task) (Holt and Laury, 2002) and our simple, entertaining incentivized (Wheel-task) experimental risk measure and contrast the results with the general risk question (DO) (Dohmen et al., 2011).

The results of our within-subject comparison suggests little consistency among the aforementioned measures. We observe high inconsistencies in the HL-task and a rather risk averse choice pattern, lower inconsistencies in the Wheel-task with a rather risk seeking choice pattern, and a rather risk neutral choice pattern in the survey question. We therefore test a possible explanations for these differences across tasks, namely that risk seeking choice pattern in the Wheel-task might be due to a utility generated from the gamification element in the task. We find that participants in the Wheel-task do derive utility simply from playing it, thus risk attitude might be measured incorrectly when too much gamification is involved. Our contribution to literature is twofold: 1) We add to the ongoing debate on which measure(s) to use when investigating risk attitudes of low-numeracy subjects 2) Test gamification in a new setting.

The paper is organized as follows. In Section 2 we provide a short literature review. In Section 3 we present the experiments in detail as well as the study group and region. In Section 4 we show and discuss the results to end with concluding remarks in Section 5.

Literature Review

This paper extends the categorization of risk elicitation methods put forward by (Charness and Viceisza, 2016) by dividing the simple/incentivized dimension into numeric and non-numeric (gamified) instruments (see Table 1)

Table 1: Taxonomy of risk elicitation methods

	Complex	Simple	
		Numeric	Non-numeric (gamified)
Incentivized	Holt and Laury (2002): *"HL-task"*, Tanaka et al. (2010)	Binswanger (1980), Gneezy and Potters (1997), Eckel and Grossman (2002)	Slovic (1966), Lejuez et al. (2002), Crosetto and Filippin (2013), *"Wheel-task"*
Non-Incentivied	Weber et al. (2002)	Dohmen et al. (2011): *"DO"*	

Incentivized approaches include some kind of monetary or in-kind stimulus, included into the design of the game to incentivize the player to play at her best – to yield true results. These approaches can be complex, which are for example multiple price lists (MPL) among which the player has to choose the preferred option. Binswanger (1980) represents a well-known pioneer in this context, using a lottery-choice approach among farmers in rural India. Incentivized methods can also be simple (numeric), which can be e.g. in form of a simple investment game (Eckel and Grossman, 2002). Simple (non-numeric, gamified) methods are those where the player is not confronted with numbers and probabilities but rather with more intuitive tasks e.g. BART (Lejuez et al., 2002). Non-incentivized approaches are mainly embedded into surveys. Most commonly, individuals are asked to state their general (Dohmen et al., 2011) or domain specific (Weber et al., 2002) preferences. The examples seen in Table 1 have been chosen on the basis of popularity in the field. Especially the HL-task, among others, makes up the frontline method that has been used in the context of a developing country (Verschoor et al., 2016).

Selecting one or multiple methods that are suitable for applying in a rural setting of a developing country is however a rather complex endeavor. Due to the particular environment and circumstances of subjects, researchers have to carefully assess the tradeoff between the preciseness of a methods per design and the ability for subjects to actually understand the task. MPL's, and especially the HL-task, is being praised as the gold standard when it comes to eliciting risk attitudes. While there is proof for its predictive power and many advantages attached to the method, it is important to ask whether this is a valid universal claim, as most of the most popular methods shown above were developed and tested among high-literacy subjects from western universities (Henrich et al., 2010). Abstract lottery choices tend to be relatively complex and thus demand a certain skill-set of the participants, especially cognitive skills such as abstracting, calculating, and retaining facts in memory. However, as many participants in poor rural areas of developing countries tend to have low education and numeracy levels, and do not deal with intellectual content on a daily basis, complex incentivized tasks can, in general, cause great confusion. This can lead to reduced accuracy in estimated results for complex incentivized tasks Charness and Viceisza (2016).

One prominent indicator for reduced accuracy in MPL's is the inconsistency rate, i.e. switching more than once between the lotteries, which tends to be rather high in the HL-task conducted among low-literacy subjects. While the inconsistency rates are in-between 10-15% in developed country settings (Charness and Viceisza, 2016), they tend to be crucially higher in developing countries. Brick et al. (2012) come up with an inconsistency rate of 41% in South Africa, Doerr et al. (2011) find 39% in Ethiopia, Jacobson and Petrie (2009) find 50% in Rwanda, and Galarza (2009) finds that 50% of the farmers in Peru switch at least two times. Even though Anderson and Mellor (2009) suggest that inconsistencies might indicate indifference between the alternatives, it is more likely that it indicates the lack of understanding of the task. Brick et al. (2012) suggest that MPL methods "although widely used in both student samples and in the field, might not be the most appropriate elicitation tool within a developing country context" (p.141). This opinion is shared by many other researchers as well ((Charness and Viceisza, 2016; Holden, 2014; Nielsen et al., 2013; Dave et al., 2010).

However, Verschoor et al. (2016) point out the crucial importance of how a task is displayed to low-literacy subjects by suggesting that the HL-task can actually be a suitable instrument, depending on how the lottery is designed and implemented. Careful application of the HL-task has, e.g., been done by (Ihli et al., 2013) among smallholder farmers in Uganda, yielding considerably lower inconsistency levels relative to comparable field research. This suggests that lab designs of more complex tasks may be rephrased for use in the field, often by translating probabilities into easily comprehensible illustrations such as different colored balls for different probabilities (Sagemüller and Mußhoff, 2020; Liu, 2013; Tanaka et al., 2010) or a coin flip (Brick et al., 2012), embedding the gamble into real-life scenarios and using easy demonstrations with cards and fingers (Holden, 2014). Also, researchers commonly apply 'salient probabilities' of 50/50 chances in gambles (Brick et al., 2012) or one-fourth, one-half, three-fourth and one in the HL-task (Harrison et al., 2010), rather than the more complex and more frequently changing probabilities of lab-designs.

With a core focus on an appealing design, simple (non-numeric, gamified) methods such as the balloon task stand out for their simplicity. They are highly intuitive to grasp and no education level is needed in order accomplish them. However, they are seldom applied in the field, which might be due to their low level of preciseness, e.g., the narrow range of variables generated from it (Lejuez et al., 2002). Moreover, recent measures that include gamification elements are digital tasks, which can be challenging to conduct in rural areas of developing countries due to e.g. unstable energy supply. Furthermore, while there is evidence that the risk attitude measured through the BART correlates with real world risk behavior, the evidence tends to focus on small samples form high-literacy subjects (Lejuez et al., 2002, 2003).

Non-incentivized methods, be it domain specific or general, are praised in literature for being cheap and easy to conduct and easy to understand by participants (Menkhoff and Sakha, 2017), but the precision of the method is seen ambivalent. Menapace et al. (2016) are concerned that results might not capture true attitudes as the general DO question is not embedded into any real setting and therefore individuals cannot associate it with their past behavior. Also, hypothetical survey questions might lack in precision as they are not incentive compatible and suffer from framing effects and potentially biased answers (Dohmen et al., 2011). Despite their carefully laid out concerns regarding hypothetical survey questions, Dohmen et al. (2011) further compare and find that nevertheless the best comprehensive risk attitude indicator appears to be the general question. Many researchers, such as Charness and Viceisza (2016), find it to be an accurate alternative to an MPL task.

A general rule of thumb, regarding the choice of a suitable method, found in literature is 'simpler is better' (Charness and Viceisza, 2016; Crosetto and Filippin, 2013; Dave et al., 2010). Simpler and more intuitive methods appear to trump more complex methods, as the ability for low-numeracy, low-educated individuals to actually understand them yields more qualitative data. Using simpler designs makes it more likely for researchers to find a closer approximation of real risk attitudes (Dave et al., 2010). The challenge thus is to identify a simple, appealing, and yet precise risk measurement tool.

Thus, in this study focused on picking representatives for the categories in Table 1 for our comparison. We compare the HL-task (complex-incentivized) with our novel Wheel-task (simple-incentivized). Furthermore, we use the DO (non-incentivized) to contrast our experimental methods, as literature suggests it to be a fair risk-attitude indicator for low-literacy subjects.

Materials and Methods

Experiment 1: HL-task

Being praised as the gold standard to elicit risk preferences throughout literature (Anderson and Mellor, 2009), the binary lottery developed by Charles A. Holt and Susan K. Laury in 2002 seems to be the first choice instrument for researchers interested in understanding risk attitudes (source). For measurement, subjects are presented with paired lottery-choices among which they need to choose upon over ten rounds. As seen in Table 2 the payoffs in option A and option B stay unchanged throughout the ten rounds. As seen in Table 7A the payoffs of option A range in a narrow margin (2 USD-1.60 USD) while those of option B are more variable (3.85 USD- 0.10 USD). What changes during the lottery are the probabilities with which the payoffs occur. Thus, within the ten rounds the player is expected to switch from option A to option B. The switching point is crucial as it determines the risk attitude of the subject. A risk neutral individual is expected to switch in round five, as the difference in expected payoffs from option A to option B turns negative (-0.18 USD). Even an extreme risk averse person would show a switching point, namely in round ten, as the probability for receiving the high payoff of 3.85 USD is certain. A risk loving person would switch even before round five, with the most extreme starting off with option B in the first place and thus being the only exception of not switching at all (Holt and Laury, 2002). The player is clustered to have an inconsistent choice pattern if she i) switched more than once between the lotteries and/or ii) chose lottery A in row 10.

In our experiment, we aimed at capturing the complexity, while still using some tools to ensure proper explanation. Thus, we provided the enumerators with 20 cards. The cards had different colors, e.g. 10 green cards and 10 blue cards. The enumerators were extensively trained to use the cards to explain the logic of the HL-task and to later demonstrate the probabilities in every round (see Figure A.1 for the illustration instruction).

Experiment 2: The Wheel-task

The Wheel of Fortune task is designed by the authors to serve as an investigation of weather incorporating gamification in a physical risk elicitation method can be beneficial for the risk attitude measurement in the field. Basically, the participant has to choose between playing a game, namely spinning the physical wheel of fortune, and a safe amount of money. The game is played over several rounds and in the end, one round will be drawn by lot and will actually be played.

To isolate the potential utility stemming from the game, we design the experiment in two versions, played in a between-subject design to avoid carry-over effects. In both versions, the participant al-

Table 2: HL-task (N=247)

	Lottery A			Lottery B			CRRA^{B,C}
	chance of:		EV^{A} Lottery A	chance of:		EV^{A} Lottery B	Range of constant
	2 USD	1.60 USD		3.85 USD	0.10 USD		relative risk aversion
	odds in percent:		in USD^{B}	odds in percent:		in USD^{B}	
1	10	90	1.64	10	90	0.475	$-\infty \leq r \leq 1.71$
2	20	80	1.68	20	80	0.85	$1.71 < r \leq 0.95$
3	30	70	1.72	30	70	1.23	$0.95 < r \leq 0.49$
4	40	60	1.76	40	60	1.60	$0.49 < r \leq 0.15$
5	50	50	1.80	50	50	1.98	$0.15 < r \leq 0.15$
6	60	40	1.84	60	40	2.35	$0.15 < r \leq 0.41$
7	70	30	1.88	70	30	2.73	$0.41 < r \leq 0.68$
8	80	20	1.92	80	20	3.10	$0.68 < r \leq 0.97$
9	90	10	1.96	90	10	3.48	$0.97 < r \leq 1.37$
10	100	0	2.00	100	0	3.85	$1.37 < r \leq 2.21$

^{A}expected value; ^{B}column was not shown; ^{C}a power utility function of the form U(x) =[x(1r)/(1-r)] is assumed (Holt and Laury, 2002)

ways chooses between 100 cents (1 USD) safe or a gamble. This gamble always has the same payout (for x1 you got 0.05 USD and for x2 you got 2.50 USD), except for round 6 where the participant can either win 1 USD or 2.50 USD[1]. What changes, in both versions, are the probabilities. The gamble starts from a very high probability for the low payout (90 percent for H2(x1)) and then the probability for the higher payout increases to 90 percent in round five (six) and 95 percent in round six (seven). The game is designed with mostly fixed payouts to make the illustration rather easy. Also, we incorporate salient probabilities which can be easily understood. Lastly, a participant's choice pattern is clustered as inconsistent if she switches more than once between lottery A and lottery B.

As seen in Figure A.2, we used one physical wooden wheel and created six round slices that were put on the wheel as layers while the enumerator was explaining the probabilities. We illustrated the probabilities for H2(x1) and H2(x2) by coloring the probabilities for H2(x2) in dark blue, using the logic of a pie chart. When explaining the rounds the enumerator thus always showed the one dollar bill as one option and then presented the wheel, adding the correct paper circle on it for the respective probability. The participant could then either vocally communicate or simply point at her choice. Before the experiment started, the enumerator turned the wheel with a random slice to show that it is a fair wheel and once the wheel stopped and a wooden arrow indicated the color chosen (exactly as in any common wheel of fortune game), the enumerator explained how much money the participant would have won and how much money she would otherwise could have won.

Table 3: Wheel-task V1 (N=139)

	Lottery A (safe)			Lottery B (Wheel)					CRRA^{B,C}
	Lottery A		EV^{A}(A)	Lottery B				EV^{A} (B)	Range of constant
	amount	odds		amount	odds	amount	odds		relative risk aversion
	USD	%	USD	USD	%	USD	%	USD	
1	1.00	100	1	0.05	90	2.50	10	0.295	$-\infty \leq r \leq 1.51$
2	1.00	100	1	0.05	75	2.50	25	0.663	$1.51 < r \leq 0.50$
3	1.00	100	1	0.05	50	2.50	50	1.275	$0.50 < r \leq 0.32$
4	1.00	100	1	0.05	25	2.50	75	1.888	$0.32 < r \leq 0.95$
5	1.00	100	1	0.05	10	2.50	90	2.255	$0.95 < r \leq 1.47$
6	1.00	100	1	1.00	5	2.50	95	2.425	$1.47 < r \leq 2.57$

^{A}expected value; ^{B}column was not shown; ^{C}a power utility function of the form U(x) =[x(1r)/(1-r)] is assumed

After playing V1 (our baseline version) with 139 participants, we introduced Version 2 (V2). Here,

[1]The last row was installed after an extensive pilot, in which we learned that participants tend to monotonously choose the gamble or monotonously choose the safe bet. We changed the payout to 1 USD in round six to design a harmonic series. Furthermore, even a highly risk averse player is expected to switch to lottery B (wheel) in the last round. Staying with lottery A monotonously choice throughout all six rounds is a characterization of a irrational player.

we incorporated an additional round to understand if participants were also willing to forgo a safe amount of money for playing the wheel game when, instead of making them potentially only better off (round 6 in V1), they could be worse off. In other words, we introduce a new round which can reveal if there is a utility for the participants solely derived from playing. Therefore, we introduced a new first row in the Wheel-task V2, all other rounds stay the same:

Table 4: Wheel-task V2 (N=107)

	Lottery A (safe)			Lottery B (Wheel)					CRRA^{B,C}
	Lottery A		EV^A(A)	Lottery B				EV^A (B)	Range of constant relative risk aversion
	amount USD	odds %	USD	amount USD	odds %	amount USD	odds %	USD	
1	1.00	100	1	0.05	90	1	10	0.145	$-\infty \leq r \leq 5.06$
2	1.00	100	1	0.05	90	2.50	10	0.295	-5.06 < r $\leq$ 1.51
3	1.00	100	1	0.05	75	2.50	25	0.663	1.51 < r $\leq$ 0.50
4	1.00	100	1	0.05	50	2.50	50	1.275	0.50 < r $\leq$ 0.32
5	1.00	100	1	0.05	25	2.50	75	1.888	0.32 < r $\leq$ 0.95
6	1.00	100	1	0.05	10	2.50	90	2.255	0.95 < r $\leq$ 1.47
7	1.00	100	1	1.00	5	2.50	95	2.425	1.47 < r $\leq$ 2.57

^{A}expected value; ^{B}column was not shown; ^{C}a power utility function of the form U(x) =[x(1r)/(1-r)] is assumed

Here, everything stays the same except for the new first row. As seen in Table 4, H1(x1) is still a safe 1 USD win, however H2 holds the following options: either the participant wins 0.05 USD to 90 percent or she wins 1 USD to 10 percent. The participant now might forgo 0.95 USD if she plays the game.

General risk question

As a final risk assessment we use a non-incentivized method, the general risk question. To measure risk behavior, participants are asked the following question: *How do you see yourself: are you generally a person who is fully prepared to take risks or do you try to avoid taking risks? Please tick a box on a scale, where the value 0 means: 'not at all willing to take risks' and the value 10 means: 'very willing to take risks'.* (Dohmen et al. (2011), p. 525). The scale can vary (source), we used a scale from 1 (not willing at all to take risks) to 10 (very willing to take risks). We embedded the question into our questionnaire. During the training, the enumerators developed an understanding of the concept of risk attitude and a standardized way to explain it. Thus, we ensured that the question was phrased in the same way for each participant and no examples that might bias the participant were used.

Study area and study group

We conducted a field study with 247 smallholder farmers from rural Cambodia. The study builds on a survey that was implemented throughout sixteen villages of the Ratanakiri Province, remotely situated in northeastern Cambodia. The data collection took place from August to October 2018, including an intensive pilot (N=40). With a gross national income per capita of 1,075 USD, Cambodia is clustered as a least developed country (United Nations, 2018) with the province being classified as one of the poorest areas in the country, which is the reason for conducting our data collection and experiments in that exact region. Of the 150,000 citizens, 88 percent live in rural areas and depend predominantly on smallholder agriculture Ritzema et al. (2019); ADB (2014).

Every participant needed to have at least basic skills in speaking and understanding the national language and needed to be a smallholder farmer household head. Since there are no household lists for the region available to researchers, we rely on the expert knowledge of the extension workers from the regional government. Together with the respective village officials, they carefully selected participants to generate a cross-section of the village, paying close attention to randomizing characteristics such as age, gender, education, and income level. Local enumerators privately guided the participants throughout the research session, which comprised an experimental part and a questionnaire. After an approximately three-hour session, the participants received a payout equivalent to a day's wage. The payout consisted of two parts: A fixed amount and the win from the experiments.

As seen in Table 5, the farmers in our sample are on average 39.58 years old. The majority has some basic literacy, as 57% can calculate and schooling took place for 2.87 years, on average. With a share of 59%, there are slightly more females than males in our sample. The farmers scored, on average, 4 out of 12 correct puzzles in the RPM-test. With respect to inconsistent behavior in the two experiments we find that the inconsistency for the HL is with 70% very high and relatively lower for the wheel task (28%)[2]. The average stated risk attitude in the Dohmen question was 5.66, indicating an average risk-neutral attitude, slightly leaning towards risk seeking.

Table 5: Descriptive statistics (N=247)

Statistic	Unit	Mean	St. Dev.	Min	Max
Age	years	39.58	14.51	13.00	76.00
Can calculate	dummy	0.57	0.48	0.00	1.00
Cognitive skills	ordinal[a]	4.06	2.67	0.00	12.00
Education	years	2.87	3.18	0.00	12.00
Ethnicity					
Jarai	dummy	0.48		0.00	1.00
Khmer	dummy	0.26		0.00	1.00
Other	dummy	0.26		0.00	1.00
Gender	dummy[b]	0.59	0.49	0.00	1.00
Spiritual Group					
Buddhist	dummy	0.54		0.00	1.00
Other	dummy	0.36		0.00	1.00
None	dummy	0.10		0.00	1.00
Risk measurement					
Dohmen	ordinal[c]	5.66	2.34	1.00	10.00
Inconsistency HL-task[d]	dummy	0.70	0.46	0.00	1.00
Inconsistency Wheel-task[e]	dummy	0.28	0.45	0.00	1.00

[a] Measured as the number of correct answers given in the RPM ranging from 0-12.
[b] Female=1, male=0
[c] Self-perception risk assessment question ranging from 1=extremely risk averse to 10=extremely risk seeking
[d] Inconsistent choice behavior with 1= i)participant switches more than once between lottery A and B and/or ii) participant chooses Lottery A in row 10; When excluding "ii", the inconsistency level is 64%
[e] Inconsistent choice behavior with 1=participant switches more than once between wheel and safe amount

Results and discussion

Consistency among measures

Figures 1-4 show the choices made as proportion of participants. Figure 1 presents the proportion of participants choosing the safer option (option A) in the HL-task. The majority of participants choose A in the first four rounds, starting off with 80% in round one. While there is a slight peak

[2]Note that both inconsistency rates cannot be directly compared, as the number of rounds differs.

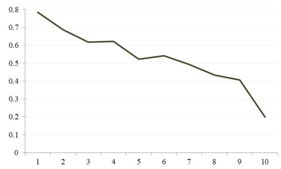


Figure 1: Safe Choices HL-Task

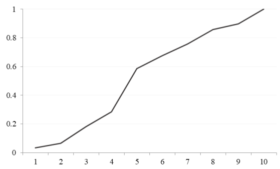


Figure 2: Dohmen Decision

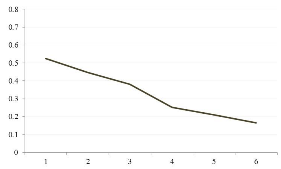


Figure 3: Safe Choices Wheel Task V1

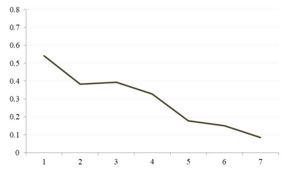


Figure 4: Safe choices Wheel-Task V2

in round six again, the trend of the choice pattern moves towards the riskier option over the last rounds (seven to ten). Figure 2 presents the accumulated choice pattern in the Dohmen task. The answering pattern is relatively stable. Wheel task V1 shows roughly the same pattern (starting safe and moving towards riskier options), whereas choice patterns in Wheel task V2 differ. As seen in Figure 4, most of the participants start off with choosing the safer option, followed by a steep decline (meaning the majority of participants chose the wheel in round two) followed by a slight increase in participants being more interested in the safer option again in round 3. From round four to round seven we see an increasing tendency towards the riskier option, thus in line with Wheel-task V1 and HL-task results. Notably, participants tend to start off riskier in the Wheel-task in contrast to the HL-task, with 80% starting with a safe choice in the HL-task and around 50% in the Wheel-task. We see some shaky results for the risk averse section of the tasks (HL-task, Wheel-task, and DO) and rather steady pattern in the risk seeking section.

Figures 5-7 presents risk attitude, clustered into risk averse (1), risk neutral(2), and risk seeking (3) for all three approaches. Risk attitudes differ between the methods. Using DO (Figure 7), we find results to be rather balanced between risk averse (28.11%), risk neutral (30.12%) and risk seeking (41.77%), with a lean towards risk seeking individuals. The picture looks different when it comes to the risk elicitation experiments.

At a first glance one can see that the extremes, i.e. risk seeking and risk averse, are more pronounced in Figure 6 and Figure 5, respectively, compared the Figure 7. The results from the HL-task show a strong lean towards risk averse individuals (62.65%), while participants are rarely clustered as risk neutral (12.85%) and slightly more often clustered as risk seeking (24.5%). Results of the Wheel-task[3] peak on both ends of the spectrum, i.e. risk neutral behavior is measured rather seldom with 7.69%), with a tendency towards risk seeking behavior (57.89%). In other words, the wheel tasks reveals that roughly 60% of the sample is risk seeking, while the HL-task reveals the opposite, namely that about 60% of the sample is risk averse. Neither of the results fit the answers given in the DO.

Taken together, the categories look different for the three tasks. Through the methods DO and Wheel-task we find that risk seeking behavior is rather strong. However, while DO shows a steady increase (from risk averse to risk seeking), the pattern elicited by the Wheel task looks different. Here we see extremes on both ends, whereas risk seeking is stronger pronounced and very low rates for risk neutral attitudes. While the HL task also reveals extremes on both ends, here the risk neutral player is more heavily pronounced. Thus, we find a tendency towards extremes in the experimentally elicited risk methods.

[3] V1 and V2 combined.

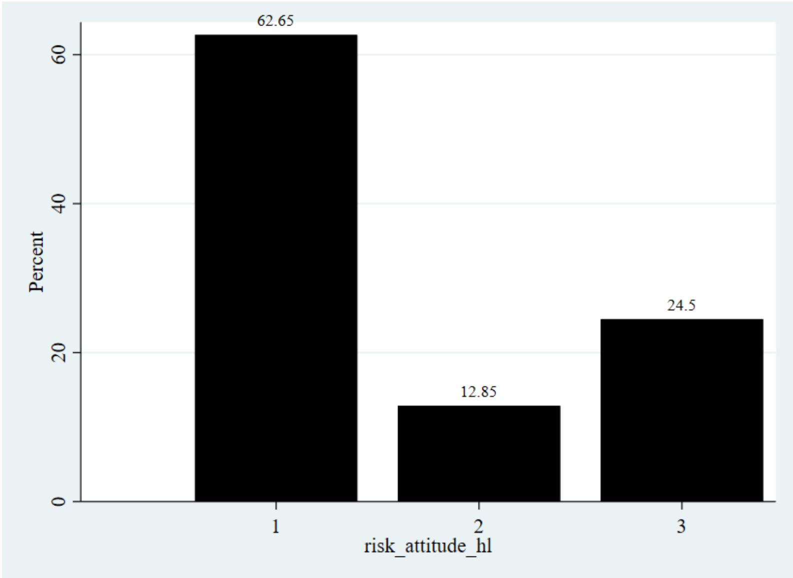


Figure 5: Risk attitudes derived from HL-task (N=247)

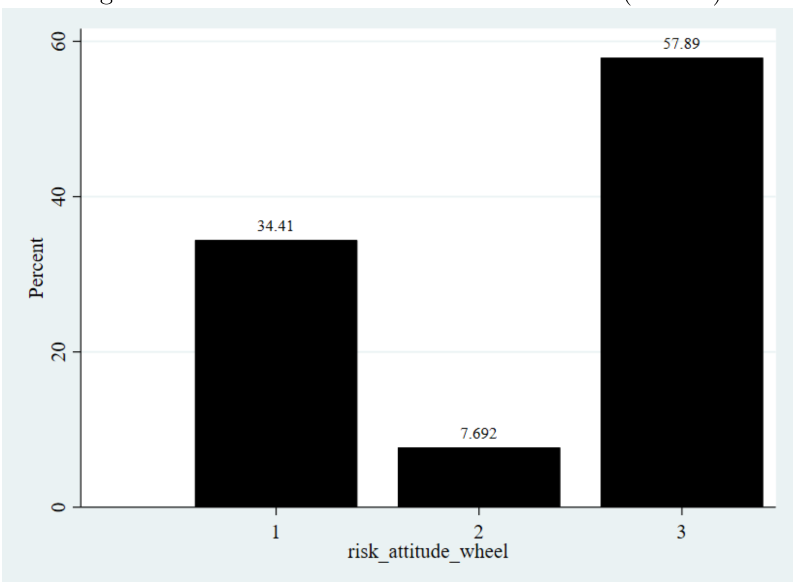


Figure 6: Risk attitudes derived from Wheel-Task (N=247)

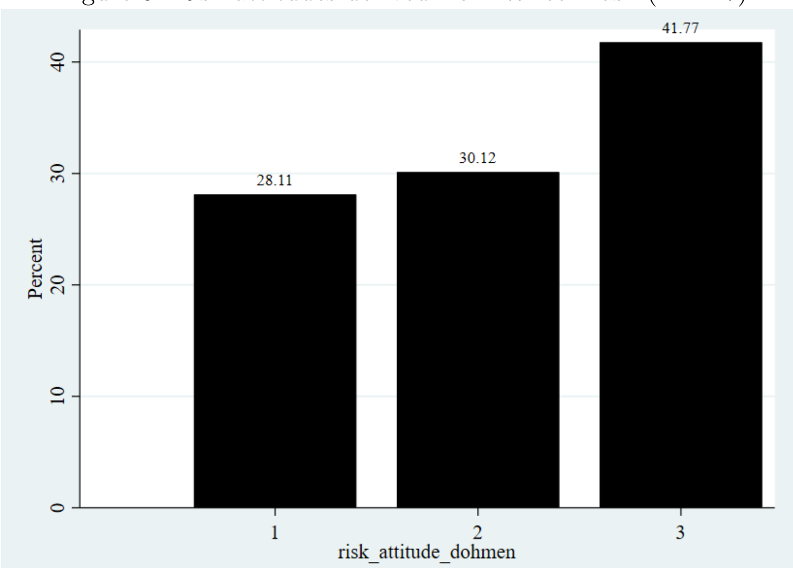


Figure 7: Risk attitudes derived from Dohmen Question (N=247)

Postestimation: Investigating gamification

The finding that inconsistencies in the complex task can partly be explained by cognitive skills might appear - at a first glance - as a strengthening point for the Wheel-task. The Wheel-task is easy, appealing, and incorporates gamification. Also, cognitive skills appear to be rather irrelevant with respect to inconsistent choice patterns. However, as we find a stark tendency towards risk-seeking behavior in the wheel-task, a further investigation is needed.

WheelByStudyNew.png

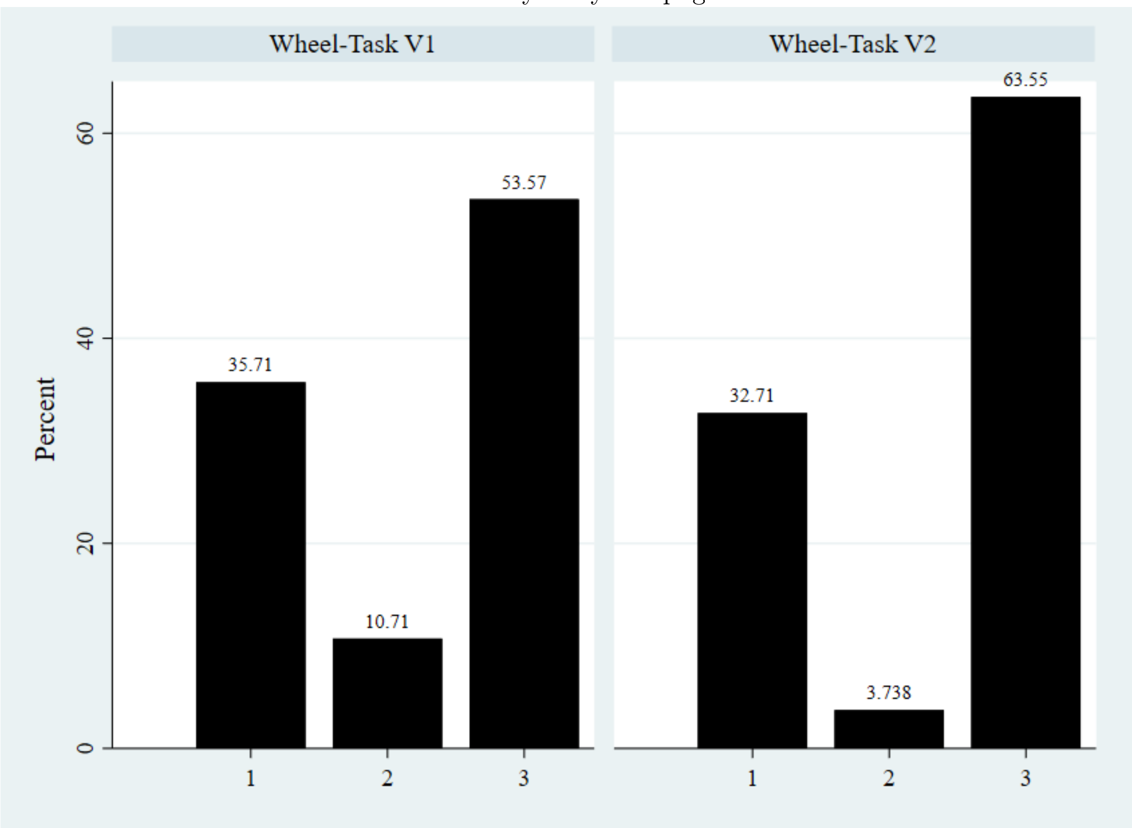


Figure 8: Wheel-Task by study

Figure 8 depicts the risk attitude (as seen in Figures 5-7), but this time, we split the results from the Wheel-task (Figure 6) by versions. While there is again a strong pronunciation on the extremes risk averse (1) and risk seeking (3), V1 and V2 differ. In both versions, the Wheel-task revealed roughly 30% of the sample to be risk averse. However, in V1 10.71% are risk neutral and 53.57% are risk seeking. In V2, only 3.73% are risk neutral and 63.55% are risk seeking. In other words, in V2 plus ten percentage points of the sample are risk seeking in comparison to V1.

We use the first row in V2 of the Wheel-task to zoom into our findings. Out of 107 participants in V2 of the wheel task, 49 (46%) chose the wheel over the safe amount in row 1. Therefore, we assume a large utility derived from playing the wheel. 46% of the sample are willing to forgo 0.85 cents in expected value for the sake of playing the game.

In synthesis, while we find that while the Wheel-task appears as a potentially promising tool for low-literacy subjects in a developing country settings, it over-estimates - most probably by design - risk seeking behavior. This is due to the fact that it does not solely measure risk attitudes (as it is - at its core - a binary lottery), but rather the effect of gamification. The participants appear to partially ignore the odds in the Wheel task, especially in row 1 of V2 and rather want to play the game for the utility derived from playing is larger than the amount at stake.

Conclusion

In this study we investigated weather incorporating an element of gamification in a precise risk-elicitation measure can be beneficial for the risk analysis among low-literacy subjects. To do so, we have analyzed risk perception of 247 smallholder farmers. We collected extensive data on socio-demographics, asked a self-perception risk question and undertook two experiments (HL-task and the novel Wheel-task) in rural Cambodia. In synthesis, our field study reveals that 1) Results from the experimentally elicited risk attitudes and the risk perception question are not concordant and 2) The gamification element in our risk-elicitation method (Wheel-task) most probably led to a (false) overestimation of risk seeking behavior, as participants appeared to decide on the riskier option merely for the sake of playing the game.

The overarching question of this study was if gamification can potentially improve risk attitude measurement among low-literacy subject-pools, i.e. if gamification is a curse or blessing for the design of risk elicitation methods in the field. The results of our study suggest the potential of a measurement error when making the experiment more attractive in form of a engaging game. However, as indicated in the comparatively low inconsistency rates in the Wheel-task, gamification might still increase attention. Future work could investigate who in particular is attracted by gamification and/or could use different within-subject comparisons to elicit a threshold of how much gamification might be useful. Furthermore, different variations of the Wheel-task design could be tested, for example by redesigning the safe amount choice.

References

ADB, A. D. B. (2014). *Cambodia; country Poverty Analysis 2014*. O Books Sun & Moon.

Anderson, L. R. and Mellor, J. M. (2009). Are risk preferences stable? comparing an experimental measure with a validated survey-based measure. *Journal of Risk and Uncertainty*, 39(2):137–160.

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

Brick, K., Visser, M., and Burns, J. (2012). Risk aversion: Experimental evidence from south african fishing communities. *American Journal of Agricultural Economics*, 94(1):133–152.

Charness, G., Garcia, T., Offerman, T., and Villeval, M. C. (2020). Do measures of risk attitude in the laboratory predict behavior under risk in and outside of the laboratory? *Journal of Risk and Uncertainty*, 60(2):99–123.

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.

Crosetto, P. and Filippin, A. (2013). The "bomb" risk elicitation task. *Journal of Risk and Uncertainty*, 47(1):31–65.

Crosetto, P. and Filippin, A. (2016). A theoretical and experimental appraisal of four risk elicitation methods. *Experimental Economics*, 19(3):613–641.

Dave, C., Eckel, C. C., Johnson, C. A., and Rojas, C. (2010). Eliciting risk preferences: When is simple better? *Journal of Risk and Uncertainty*, 41(3):219–243.

Deterding, S., Dixon, D., Khaled, R., and Nacke, L. (2011). From game design elements to gamefulness: defining" gamification". In *Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments*, pages 9–15.

Doerr, U., Toman, O. M., and Schmidt, U. (2011). Overconfidence and risk management of ethiopian farmers. *University of Kiel (Working paper)*.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., and Wagner, G. G. (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the european economic association*, 9(3):522–550.

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.

Galarza, F. B. (2009). Choices under Risk in Rural Peru. Staff Papers 92247, University of Wisconsin-Madison, Department of Agricultural and Applied Economics.

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

Harrison, G. W., Humphrey, S. J., and Verschoor, A. (2010). Choice under uncertainty: evidence from ethiopia, india and uganda. *The Economic Journal*, 120(543):80–104.

He, P., Veronesi, M., and Engel, S. (2018). Consistency of risk preference measures: An artefactual field experiment from rural china. *The Journal of Development Studies*, 54(11):1955–1973.

Henrich, J., Heine, S. J., and Norenzayan, A. (2010). Most people are not weird. *Nature*, 466(7302):29–29.

Holden, S. T. (2014). Risky choices of poor people: Comparing risk preference elicitation approaches in field experiments.

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

Ihli, H. J., Chiputwa, B., Bauermeister, G.-F., and Musshoff, O. (2013). Measuring risk attitudes of smallholder farmers in uganda: How consistent are results of different methods? In *The Second International Agricultural Risk, Finance, and Insurance Conference Paper*.

Jacobson, S. and Petrie, R. (2009). Learning from mistakes: What do inconsistent choices over risk tell us? *Journal of risk and uncertainty*, 38(2):143–158.

Lejuez, C. W., Aklin, W. M., Zvolensky, M. J., and Pedulla, C. M. (2003). Evaluation of the balloon analogue risk task (bart) as a predictor of adolescent real-world risk-taking behaviours. *Journal of adolescence*, 26(4):475–479.

Lejuez, C. W., Read, J. P., Kahler, C. W., Richards, J. B., Ramsey, S. E., Stuart, G. L., Strong, D. R., and Brown, R. A. (2002). Evaluation of a behavioral measure of risk taking: the balloon analogue risk task (bart). *Journal of Experimental Psychology: Applied*, 8(2):75.

Liu, E. M. (2013). Time to change what to sow: Risk preferences and technology adoption decisions of cotton farmers in china. *Review of Economics and Statistics*, 95(4):1386–1403.

Menapace, L., Colson, G., and Raffaelli, R. (2016). A comparison of hypothetical risk attitude elicitation instruments for explaining farmer crop insurance purchases. *European Review of Agricultural Economics*, 43(1):113–135.

Menkhoff, L. and Sakha, S. (2017). Estimating risky behavior with multiple-item risk measures. *Journal of Economic Psychology*, 59:59–86.

Mosley, P. and Verschoor, A. (2005). Risk attitudes and the 'vicious circle of poverty'. *The European journal of development research*, 17(1):59–88.

Nielsen, T., Keil, A., and Zeller, M. (2013). Assessing farmers' risk preferences and their determinants in a marginal upland area of vietnam: a comparison of multiple elicitation techniques. *Agricultural Economics*, 44(3):255–273.

Ritzema, R. S., Douxchamps, S., Fraval, S., Bolliger, A., Hok, L., Phengsavanh, P., Long, C., Hammond, J., and van Wijk, M. (2019). Household-level drivers of dietary diversity in transitioning agricultural systems: evidence from the greater mekong subregion. *Agricultural Systems*, 176:102657.

Sagemüller, F. and Mußhoff, O. (2020). Effects of household shocks on risk preferences and loss aversion: Evidence from upland smallholders of south east asia. *The Journal of Development Studies*, 56(11):2061–2078.

Sailer, M. (2016). *Die wirkung von gamification auf motivation und leistung*. Springer.

Slovic, P. (1966). Risk-taking in children: Age and sex differences. *Child Development*, pages 169–176.

Tanaka, T., Camerer, C. F., and Nguyen, Q. (2010). Risk and time preferences: Linking experimental and household survey data from vietnam. *American economic review*, 100(1):557–71.

United Nations (2018). The least developed country category: 2018 country snapshots.

Verschoor, A., D'Exelle, B., and Perez-Viana, B. (2016). Lab and life: Does risky choice behaviour observed in experiments reflect that in the real world? *Journal of Economic Behavior & Organization*, 128:134–148.

Weber, E. U., Blais, A.-R., and Betz, N. E. (2002). A domain-specific risk-attitude scale: Measuring risk perceptions and risk behaviors. *Journal of behavioral decision making*, 15(4):263–290.

World Bank (2016). Poverty: Overview. *World Bank*.

World Bank (2018). Poverty and shared prosperity 2018: Piecing together the poverty puzzle.

World Bank (2020). World development indicators database. data retrieved online, https://datacatalog.worldbank.org/dataset/world-development-indicators.

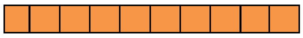
Appendix

HOLT AND LAURY LOTTERY

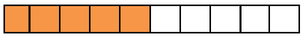
INSTRUCTION

To illustrate the probabilities to you, I will use 10 cards. *Lay down 10 cards*

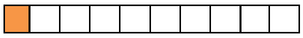
If I say, you get $1 with 100%, I will show you 10/10 cards



If I say, you get $1 with 50%, I will show you 5/10 cards



If I say, you get $1 with 10%, I will show you 1/10 cards



As you can see: More cards mean higher chances! Fewer cards mean lower chances! So, now let's get back to the game. I will start with round one and you tell me which option you like best. And remember, you are playing for real money!

In round one, A, you have the chance to win $2 with 10% *(lay down one card)* or $1.60 with 90% *(lay down nine cards)* and in B, you have the chance to win $3.85 with 10% *(lay down one card)* or $0.10 with 90% chance *(lay down 9 cards)*. Which gamble A or B do chose in this round?

Please note down the results in your EXPERIMENT RESULTS SHEET and continue with the next row. Do so until down with all rows.

Great job! Now, as promised, you can draw a paper out of this box. There is a number between one and ten and will determine which row we will play for real money. *(let him draw the number out of your box A)*

Let's say, the number is 1. Then you check you sheet to see what he has answered in round 1. Let's say he chose option A in round 1. Then you use your Box B and put the right share of two different colored cards in there: here, 1 green and 9 red. Then, with closed eyes the participant draws one card. The respective amount is written down on the EXPERIMENT RESULTS SHEET and will be added at the end.

Figure A.1: HL-task instruction

Figure A.2: Wheel-Task demonstration