Testing the Elicitation Procedure of the Minimum Acceptable Probability

Maria Polipciuc* Martin Strobel*

November 9, 2021

Abstract

*Maastricht University. Email: m.polipciuc@maastrichtuniversity.nl. We thank Elias Tsakas and participants in the BEELab proposal meeting for valuable comments.

1 Introduction

Individuals have often been found to prefer exposure to a randomly generated risk than to an equiprobable risk generated by an opponent in a strategic situation. In the context of trust games, this strategic risk premium has been dubbed *betrayal aversion* (Bohnet and Zeckhauser, 2004). Many papers find that betrayal aversion is an important determinant of trust CITE.

Betrayal aversion is identified as the difference in first mover behavior in two games: a binary trust game CITATION and an equivalent game where the decision at the second node is made by a random device. First movers typically have to indicate their *minimum acceptable probability* (MAP), which is a strategy-method elicitation procedure CITE which has many features in common with the Becker–DeGroot–Marschak mechanism. That is, prior to knowing how many second movers (devices) chose the favorable outcome at the second node, first movers have to state their conditional threshold probability of the good outcome at the second node for preferring to send money to the second mover over the outside option (which keeps the two players' equal initial endowments unchanged). Then, should that threshold be reached or exceeded, first movers send money to second movers, and the outcome is decided by their matched second mover's choice (their matched device's choice). Should the threshold not be met, the outside option is implemented.

A recent paper has shown theoretically that the elicitation procedure of minimum acceptable probabilities used in most papers leaves the door open to potential confounds such as "ambiguity attitudes, complexity, different beliefs, and dynamic optimization" (Li et al., 2020) if players are not rational expected utility maximizers.

Moreover, a couple of empirical papers which use more stringent identification procedures for betrayal aversion by controlling for beliefs in the two games do not find betrayal aversion (Fetchenhauer and Dunning, 2012; Polipciuc and Strobel, 2020), or find it to play a role for trusting only when beliefs are far more optimistic than is generally the case (Engelmann et al., 2021).

In this note, we use an online experiment to measure how much of what has been called betrayal aversion is due to distributional dependence, regardless of the source of risk being random or strategic. We remove the strategic component and show participants complete distributions over probabilities of the good (and bad) outcome of a lottery, and ask them to state their cutoff probability of the favorable outcome for preferring the lottery to a safe payoff. Should the numerical example in (Appendix A in Li et al., 2020) be applicable to our setting, we expect participants to more readily accept the lottery when it comes from a distribution with a higher expected value. We find the opposite to be true: the higher the expected value of the probability of the favorable outcome, the higher the minimum acceptable probability required by participants to accept the lottery.

While this is at odds with our expectations, it ties in with findings from the empirical literature on distributional dependence of willingness to pay (WTP). The MAPs from which one identifies betrayal aversion are elicited through a variant of the Becker–Degroot–Marschak (BDM) (Becker et al., 1964) mechanism, which is an often used mechanism for eliciting valuations. In the Becker–DeGroot–Marschak mechanism, a potential buyer states the maximum price for which she is willing to buy a good. A price is drawn, and if it is lower than or equal to the price she stated, she buys the good. If the price is higher, she keeps her endowment and does not buy the good. There a couple of differences between the 'standard' BDM

and MAP elicitation: (1) the auctioned good is a lottery, (2) instead of giving a maximum price for which they prefer the good to a safe payment, participants are asked to state a minimum probability of the favorable outcome of the lottery for which they prefer the lottery to a safe payment and (3) the underlying distribution of the probability of the favorable outcome is not (implicitly) uniform, as is the case in most studies using the distribution of potential prices for the good.

Theoretical literature has pointed out that the BDM mechanism is not incentive compatible if players are not rational expected utility maximizers (Karni and Safra, 1987; Horowitz, 2006). This is because individuals face uncertainty regarding the price of the good and additional uncertainty about whether they will buy the good or not. If their utility function is influenced by these uncertainties, changing the price distribution of the good might influence their MAP.

Several empirical papers find this to be the case: generally, the higher the expected price of the good, the higher the WTP (for a short review of this literature, see Tymula et al., 2016). Here I plan to say something about possible things which may be causing the result. For this, I have to understand if two papers cited by Tymula et al. (2016)—Kőszegi and Rabin (2006) and Wenner (2015)—are relevant for our study, and what they imply for our findings.

The paper is structured as follows. Section 2 describes the experimental design and procedures. Section 3 sets forth the hypothesis. Section 4 presents the results. Section 5 explains how our results inform the existing literature.

2 Design and procedures

The experiment uses a within-subject design, with each subject being exposed to all treatments sequentially. We use three distributions as treatments. The distributions are ordered in terms of the expected payoff over the entire distribution, as their name suggests: the Good, the Bad, and the Uniform (the Good $>$ the Uniform $>$ the Bad).

Two of the three distributions are meant to emulate treatments in papers on betrayal aversion. The Uniform distribution has equal chances of occurrence for each of the possible lotteries. We assume that this is what participants expect to face in treatments with decisions made by randomization devices, unless specified otherwise. The Bad distribution has an overall chance of a high payoff similar to the share of trustworthy respondents in papers on betrayal aversion (0.2895). The distribution in the Good treatment mirrors the one in the Bad treatment: it has the same variance, and minus the skewness of the Bad distribution. We included this distribution to check if departures from the Uniform distribution in either direction yield effects of similar size (albeit reverse sign).

include table describing distributions

To make the task easy to understand, we represent lotteries via 32 wheels of fortune with 15 sectors each. Dark blue sectors symbolize the high payoff (£4), light blue sectors—the low payoff (£1). The sure payoff participants received if no wheel was spun was £2. In each treatment, participants see the entire distribution of lotteries in that treatment, sorted in ascending order by the probability of the favorable outcome. Figure 1 below shows the distribution for the Good treatment.

Participants are told that one of the wheels will be drawn at random, with all

Table 1: The three distributions: X in 15 chance of high payoff

X	The Good	The Bad	The Uniform
0	1	8	2
1	1	4	2
2	1	4	2
3	1	3	2
4	1	2	2
5	1	1	2
6	1	1	2
7	1	1	2
8	1	1	2
9	1	1	2
10	1	1	2
11	2	1	2
12	3	1	2
13	4	1	2
14	4	1	2
15	8	1	2
Total	32	32	32

wheels having an equal chance to be drawn. They are asked to state a *minimum acceptable frequency* (which we refer to as MAP, even though it does not refer to a probability, but to a frequency, for an easier comparison with papers on betrayal aversion): the lowest number of dark blue sectors in the selected wheel such that they prefer to spin the wheel for their payoff instead of receiving the sure payoff.[1] Specifically, they have to answer: "Which wheels would you like to spin for your bonus?" by inserting an integer between 0 and 15 in the blank space: "I prefer to spin wheels which have at least *[blank]* dark blue sectors."

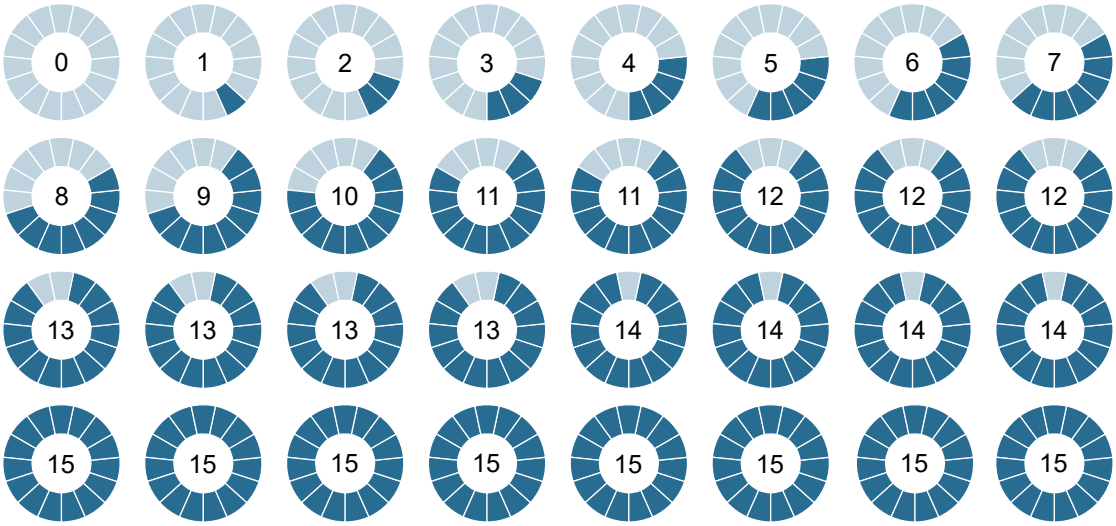

Figure 1: The Good distribution

The experiment was conducted online using Qualtrics. Participants were UK residents registered on a platform for conducting academic studies (Prolific). Since the elicitation of MAPs is rather complex (Quercia, 2016; Polipciuc and Strobel, 2020), we opted for participants who had at least a bachelor's degree. The study was pre-registered at the AEA RCT Registry (https://doi.org/10.1257/rct.7776-

[1]Since there is evidence that participants have an easier time expressing choice using integers than using probabilities (Quercia, 2016), we ask participants to input an integer.

1.1).

Table 2 in Appendix ... describes the sample. Treatment was assigned in order to balance the number of participants exposed to each of the six possible orderings of treatments. 275 of the 450 participants answered the eliminatory comprehension questions correctly and completed the experiment. Since assignment to treatment happened before participants had gone through the comprehension questions, this leads to slightly different sizes of the subsamples for the six orderings. Those who completed the experiment (did not complete the experiment) spent a median time of 12.4 (5.9) minutes and earned 5 (1) UK pounds.[2]

Table 2: Characteristics of the estimation sample

	Age	Share male	Sample size
Good–Uniform–Bad	30.956	0.333	45
	(8.808)	(0.477)	
Uniform–Bad–Good	33.538	0.346	52
	(9.074)	(0.480)	
Bad–Good–Uniform	37.114	0.523	44
	(11.071)	(0.505)	
Good–Bad–Uniform	33.132	0.491	53
	(9.174)	(0.505)	
Bad–Uniform–Good	32.429	0.333	42
	(9.423)	(0.477)	
Uniform–Good–Bad	33.333	0.205	39
	(10.103)	(0.409)	
Total	33.411	0.378	275
	(9.685)	(0.486)	

Notes: The table shows averages per sequence. Standard deviations in parentheses.

[2]The high median earnings of those who completed the experiment are due to a coding error. Instead of decisions in all three treatments being equally likely to be selected for payment, only the Good and the Uniform treatments were selected, with the probabilities indicated in parentheses. CHECK Participants were informed about the error, which increased the payoffs of all participants who had completed the experiment.

Several papers CITE find that when dealing with complex risks, participants in experiments require an extra premium compared to simple risk aversion. This premium is positively correlated with ambiguity aversion. (ARE EFFECT size similar?)

3 Hypothesis

Let p^* be the true frequency of the high payoff, whose distribution varies between treatments. According to the calculation in Appendix..., we expect the following ordering of MAPs:

Hypothesis 1 *The MAP in the Good treatment (more mass on high values of p^*) is lower than the MAP in the Uniform treatment (a uniform distribution over p^*), which is lower than the MAP in the Bad treatment (more mass on low values of p^*).*

$$MAP_G < MAP_U < MAP_B \tag{1}$$

In the pre-analysis plan, we specified that the alternative hypothesis ($MAP_B < MAP_U < MAP_G$) could be true instead if participants anchor on visual cues of the distributions, such as the mean.

4 Results

First, we present summary statistics for all decisions, by treatment and by decision order. Next, we run non-parametric tests and ordinary least squares regressions to test the hypothesis.

Table 3 presents the average MAP by treatment over all decisions and by decision order. This table already suggests that the hypothesis is not supported by the data, as the average MAP is highest in the Good treatment, followed by the Uniform treatment, followed by the Bad treatment.

Table 3: Descriptive statistics: MAPs by treatment

	All decisions	First decision	Second decision	Third decision
The Good	9.531	9.571	9.458	9.553
	(2.503)	(2.270)	(2.500)	(2.750)
The Uniform	8.844	8.890	8.368	9.227
	(2.382)	(2.392)	(2.119)	(2.539)
The Bad	8.615	8.093	9.124	8.512
	(2.522)	(2.597)	(2.491)	(2.387)
N	825	275	275	275

Notes: The table shows averages per treatment. Each participant made three decisions in randomized order. Standard deviations in parentheses. Possible answers were integers between 0 and 15.

A non-parametric Page's L test confirms this: there is strong evidence that the ordering is the opposite to the one hypothesized ($MAP_B < MAP_U < MAP_G$, p < 0.001).[3]

In Table 4 we present results of ordinary least square regressions of MAPs. When referring to sequence order, we abbreviate treatment using initials e.g. we refer to sequence Good–Uniform–Bad as GUB. Model (1) contains as regressors only dummy variables indicating the treatment. Model (2) adds age and gender as explanatory variables. Model (3) additionally includes risk attitudes. Model (4) also includes dummy variables for the order in which participants were exposed to

[3]Page's L test has the null hypothesis that all possible orderings are equally likely. The alternative hypothesis is that a specified order is the increasing order of alternatives. The Stata command is *pagetrend*.

treatments. In all models, standard errors are clustered at the individual level.

In all four specifications, participants ask for 0.687 more dark blue sectors on average in the Good treatment compared to the Uniform treatment to be willing to spin the selected wheel ($p < 0.001$ in (4)). They also ask for 0.229 fewer dark blue sectors in the Bad treatment compared to the Uniform treatment ($p = 0.001$ in (4)). More risk loving individuals have higher MAPs ($p = 0.04$ in (4)). The only sequence order which differs significantly from the baseline is GBU, which has significantly higher MAPs than GUB. A closer look shows that this result is due to significantly higher MAPs in GBU relative to GUB in the first decision (not reported). Since at that point the sequence of events and information participants faced in the two treatments was identical, we believe this difference is not a treatment effect. What do you think of this interpretation?

Result 1. *Participants set the lowest requirement to be willing to take a randomly drawn lottery in the Bad treatment, followed by the Uniform treatment, followed by the Good treatment.*

We speculated that such an ordering of MAPs is possible if individuals anchor on visual cues offered by the distributions.

5 Discussion

Table 4: Linear regressions on Minimum Acceptable Frequencies

Dependent variable:	Minimum acceptable frequency			
	(1)	(2)	(3)	(4)
The Good	0.687***	0.687***	0.687***	0.687***
	(0.099)	(0.099)	(0.099)	(0.099)
The Bad	−0.229***	−0.229***	−0.229***	−0.229***
	(0.070)	(0.070)	(0.070)	(0.070)
Age		0.005	0.006	0.006
		(0.013)	(0.013)	(0.014)
Male		−0.047	0.030	−0.029
		(0.286)	(0.284)	(0.283)
Risk attitudes (0–10)			−0.172**	−0.152**
			(0.074)	(0.074)
Sequence				
UBG				0.490
				(0.408)
BGU				−0.008
				(0.434)
GBU				1.173***
				(0.412)
BUG				−0.150
				(0.434)
UGB				0.469
				(0.433)
Constant	8.844***	8.696***	9.520***	9.066***
	(0.144)	(0.460)	(0.593)	(0.627)
N	825	825	825	825

Notes: Standard errors clustered at the individual level in parentheses. The baseline treatment is the Uniform distribution. The baseline sequence is GUB (Good–Uniform–Bad). Risk attitudes are measured on a 0–10 scale, where 0 is very risk averse and 10 is very risk loving.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

References

Becker, G., DeGroot, M., and Marshak, J. (1964). Measuring utility by a single-response sequential method. *Behavioral Science*, 9:226–232.

Bohnet, I. and Zeckhauser, R. (2004). Trust, risk and betrayal. *Journal of Economic Behavior & Organization*, 55(4):467–484.

Engelmann, D., Friedrichsen, J., van Veldhuizen, R., Vorjohann, P., and Winter, J. (2021). Decomposing trust. Technical report.

Fetchenhauer, D. and Dunning, D. (2012). Betrayal aversion versus principled trustfulness—how to explain risk avoidance and risky choices in trust games. *Journal of Economic Behavior & Organization*, 81(2):534–541.

Horowitz, J. K. (2006). The Becker–DeGroot–Marschak mechanism is not necessarily incentive compatible, even for non-random goods. *Economics Letters*, 93(1):6–11.

Karni, E. and Safra, Z. (1987). "Preference reversal" and the observability of preferences by experimental methods. *Econometrica*, 55(3):675–685.

Kőszegi, B. and Rabin, M. (2006). A model of reference-dependent preferences. 121(4):1133–1165.

Li, C., Turmunkh, U., and Wakker, P. P. (2020). Social and strategic ambiguity versus betrayal aversion. *Games and Economic Behavior*, 123:272–287.

Polipciuc, M. and Strobel, M. (2020). Betrayal aversion with and without a motive. Working paper.

Quercia, S. (2016). Eliciting and measuring betrayal aversion using the BDM mechanism. *Journal of the Economic Science Association*, 2(1):48–59.

Tymula, A., Woelbert, E., and Glimcher, P. (2016). Flexible valuations for consumer goods as measured by the Becker–DeGroot–Marschak mechanism. *Journal of Neuroscience, Psychology, and Economics*, 9(2):65–77.

Wenner, L. M. (2015). Expected prices as reference points—theory and experiments. 75:60–79.