Replication of

Expectations as Endowments: Evidence on Reference-Dependent Preferences from Exchange and Valuation Experiments

by Ericson, K.M.M./Fuster, A. (2011)
in: The Quarterly Journal of Economics, 126(4), pp. 1879–1907.

Replication Authors:
Adam Altmejd, Anna Dreber, Emma Heikensten and Magnus Johannesson

---

In a valuation experiment, Marzilli Ericson and Fuster randomize subjects to a high or low probability of obtaining an item (a mug) and elicit their willingness-to-accept it. They find that the high probability treatment increases the valuation of the item. The paper includes two experiments where other things are also tested. Experiment 2 was chosen as it was the last experiment.

Hypothesis to bet on:

The willingness to accept (WTA) for a mug is higher for a high probability of receiving the mug for free compared to a low probability of receiving the mug for free (a comparison of the mean WTA between the treatment MH (80% chance of receiving the mug for free at the end of the experiment) and the treatment ML (10% chance of receiving the mug for free at the end of the experiment) in Experiment 2).

Power Analysis

The original p-value is 0.03 (independent samples t-test of the difference between the MH and the ML treatments, the measure used is the $ln(WTA)$ for the mug minus the $ln(WTA)$ for a pen also elicited in the experiment, observations with 0 WTA for any of the two goods are dropped (footnote 25), p. 1896)): *"However, when we consider the subject level difference between $ln(WTA_{mug})$ and $ln(WTA_{pen})$, we find a mean of 0.33 for subjects in the* MH *treatment and 0.01 for subjects in the* ML *treatment. This difference is statistically significant ($p = 0.03$, t-test)."*

The original sample size is 112 participants (52 in the MH treatment and 60 in the ML treatment). To achieve 90% power the required sample size is 250 participants.

Sample

The sample for replication consists of 250 students from Harvard University (their Harvard IDs will be checked). Apart from having participated in the original experiment, there are no exclusion criteria.

Materials

We use the material of the original experiment. That is, we will be working in a computer lab with carrels, with a mug on each. Other material that will be brought into the lab, except mugs, are coins that are labeled with 1 and 8 on each sides, index cards, ten sided dies (with numbers 0 to 9) and pens. Mugs and pens are supposed to be as similar as possible to those used in the original experiment, i.e., a university travel mug and a silver metal university pen.

Procedure

We follow the procedure of the original article, with only slight but unavoidable deviations as outlined below. The following summary of the experimental procedure is therefore based on the section "III.B. Procedures" (pp. 1894–1895) in the original study.

The experiment is performed in a lab environment. Each subject is placed at a carrel with a computer and a mug. The experimenter first flips a coin with sides labeled 1 and 8 for each subject. Then the subject is given an index card with this number written on it.

The next step for the subjects is to start reading the instructions on the computer screen in front of them. The subjects are told that the experimenter roles a ten sided die (one time for each subject) at the end of the experiment. If the die comes up with a number strictly lower than the number on the index card, the subject will receive the mug for free. The subjects are also told that if the die comes up 9 they will be able to choose between the mug and a randomly determined amount between \$0 and \$10.

This implies that subjects who have index cards with 1 (8) have a 10% (80%) chance of walking away with the mug given that the die comes up 0 (0-7). The subjects with a 1 (8) on their index cards are in the treatment ML (MH) for low (high) probability of getting the mug. For all subjects there is a 10% chance to be able to choose between the mug and money.

Then the experimenter reads important parts of the instructions out loud. Subjects start to fill out a personality questionnaire in two parts. In the middle they get reminded that they may get the mug for free or have the possibility to choose between the mug and money. When the questionnaire is done the subjects are asked to make choices between different monetary amounts from \$0 to \$9.57 with increments of \$0.33 or to keep the mug. They are told that if the die comes up 9 one of these rows will be chosen randomly otherwise their decisions will not be revealed to the experimenter. Further instructions are then given and the subjects are told that if the die comes up 8 they will be able to choose between a pen and a randomly determined dollar amount. The subjects are then given the pen to inspect it and are then asked to make choices between different monetary amounts, from \$0 to \$9.57 with increments of \$0.33, or to keep the pen.

To summarize the game: All subjects have a mug on their desk and are at one point given the pen to inspect it. All subjects have a 10% chance of getting the choice between mug and money (if the die comes up 9) and a 10% chance of getting the choice between a pen and money (if the die comes up 8). There are two treatments and who is in which treatment is randomly determined by a coin flip. In treatment MH the subjects have an 80% chance of getting the mug (if the die comes up 0-7) and in ML they have a 10% chance of getting the mug (if the die comes up 0).

After the subjects have made their decision and the die has been rolled (for each subject individually), subjects will be privately paid in cash using the same show-up fee (\$10) as in the original study.

Analysis

The analysis will be performed exactly as in the original article. That is, the mug/money choices for the different amounts of money give a measure of each subject's willingness-to-accept (WTA_{mug}) for the mug. Similarly, the pen/money choices give the WTA for the pen (WTA_{pen}). For the analysis we estimate the difference between $ln(WTA_{mug})$ and $ln(WTA_{pen})$ for each subject ($diff_i$). As in the original paper we test if this variable ($diff_i$) is significantly different between the two treatments using a t-test.

In the original article the mean of $diff_i$ was 0.33 for subjects in the MH treatment and 0.01 for subjects in the ML treatment. This difference was statistically significant with a p-value corresponding to 0.03 when estimated with a t-test. Since Mrzilli Ericson and Fuster used $ln(WTA)$ subjects that indicated a \$0 WTA

for either the mug or the pen were dropped. There were three such subjects in ML and five in MH (two of which indicate a \$0 WTA for both items). We will do the test in exactly the same way and therefore also exclude subjects with a zero WTA for the mug or the pen.

Differences from Original Study

The replication procedure is the same as that of the original study, with some unavoidable deviations. This replication will be performed at the Harvard Decision Science Laboratory at Harvard University in Cambridge MA, USA, in 2015, with students at Harvard. The original data was gathered at the Harvard Decision Science Laboratory at Harvard University in Cambridge MA, USA, in 2010, with undergraduate students and graduate students at Harvard. The experiment will be in English as in the original study.

The paper contains two experiments: for the replication the focus is only on Experiment 2 as it is the last experiment in the paper.

Replication Results

In total 262 subjects participated, which is slightly more than the planned 250 subjects. In the original study the mean subject level difference between $ln(WTA_{mug})$ and $ln(WTA_{pen})$ is 0.33 in the high probability (MH) treatment and 0.01 in the ML treatment; i.e. a treatment difference of 0.32 ($p = 0.03$). In the replication the mean subject level difference between $ln(WTA_{mug})$ and $ln(WTA_{pen})$ is 0.132 in the MH treatment and -0.088 in the ML treatment. This difference of 0.22 is not significant using the same test (a two-sided t-test with unequal variances) as used in the original study ($p = 0.055$). The original result is thus not replicated. The relative effect size in the replication is 68.75% (0.22/0.32).

The difference ($WTA_{mug} - WTA_{pen}$) in dollar terms is \$0.36 for the MH treatment and $-$\$0.28 for the ML treatment ($p = 0.051$). The average WTA of the mug is \$4.31 in the MH treatment and \$3.36 in the ML treatment ($p = 0.005$). The average $ln(WTA)$ of the mug is 1.22 in the MH treatment and 0.86 in the ML treatment ($p = 0.007$). The average WTA of the pen is \$3.95 in the MH treatment and \$3.64 in the ML treatment ($p = 0.3475$). The results of the original study and the replication are summarized in Table 1 below.

Below in Table 2 we also reproduce the regression results in Table II of the original study for the original study and the replication. In these regression equations the MH treatment variable is significant, and when the $ln(WTA_{pen})$ is controlled for the MH treatment coefficient is also of similar magnitude to the original study. In the numbers in the tables below it should be noted that when the log of the mug or the pen is used, 0 valuations are dropped. The number of 0 WTA valuations in the MH treatment is 5 for the mug and 3 for the pen. The number of 0 WTA valuations in the ML treatment is 8 for the mug and 1 for the pen. Since there is some overlap, a total of 14 observations are dropped because of this when the difference between $ln(WTA_{mug})$ and $ln(WTA_{pen})$ is used.

Unplanned Protocol Deviations

Due to difficulties in recruiting, the show-up fee was raised from \$10 to \$20. A few students from the Harvard Extension School attended early sessions, but one of the original authors who participated as an observer in the second session pointed out that Harvard Extension School students did not participate in the original experiment. Therefore Harvard Extension School students were excluded from participating in any of the remaining sessions. As Extension School students could not be identified exactly in the data set from the already completed sessions, it was also decided after discussion with the original authors to remove observations for participants with a birth date lower than 1980 in these sessions (5 in total). In some cases subjects also did not follow the instructions and valued the pen without first seeing it (9 in total). These were also dropped

from the sample. One subject did not wait for the coin flip, but self selected into one of the treatments. After discussions with the original authors it was decided to exclude these observations as well. These decisions were taken before looking at the data.

The original author who participated as an observer in the second session of the experiment also pointed out a few deviations in the procedures from the original experiment. Subjects in the replication were given the pen inside a black felt bag (the pens were delivered in a black felt bag), while they had no such felt bag in the original experiment. From the third session and onwards this was done in the same way as in the original study. Due to difficulties with the experimental software the subjects were also told to enter "1" as their ID number in the program, rather than assigning them a unique number as was done in the original experiment. From the third session and onwards this was done in the same way as in the original study. Furthermore, in the original study subjects were paid in private outside of the laboratory, while in the replication subjects were paid at their desks inside the lab, with payments not visible to other subjects.

Participants were planned to only be recruited through the Harvard Decision Science Laboratory at Harvard University (using their subject pool). But as recruitment was slow subject were also recruited on campus. But only Harvard students could participate as a valid Harvard ID was required. As recruitment was done using different systems 3 students were able to register and participate in the experiment twice. The data from their second participation was excluded.

After exclusions, the final sample consists of 262 observations instead of the planned 250 observations (that we collected slightly more than the planned number of observations was because we wanted to make sure to reach at least 250 subjects after the last session and therefore over-recruited to have some margin).

Discussion

Given the criteria and procedure outlined above, the hypothesis of interest has not been replicated at a significance level of $\alpha < 5\%$. The relative effect size equals 68.75% and the p-value of the hypothesis test is $p = 0.055$.

However, it is a borderline case. The p-value is slightly above 0.05. If the tests are carried out directly comparing the mean WTA between the two treatments (with or without a log transformation) without deducting the WTA of the pen, the p-value is < 0.05 (but this test was not significant in the original study).

Table 1: The results of the original study and the replication. The results for the comparison targeted for the replication in bold text.

	Original Study			
	MH *Treatment*	ML *Treatment*	*Difference*	*p-value*
WTA_{mug}	4.12	3.74	0.38	0.44
$ln(WTA_{mug})$	1.30	1.11	0.19	0.17
$WTA_{mug} - WTA_{pen}$	0.92	0.02	0.90	0.06
$ln(WTA_{mug}) - ln(WTA_{pen})$	**0.33**	**0.01**	**0.32**	**0.03**
WTA_{pen}	3.20	3.72	−0.52	0.27
	Replication Study			
	MH *Treatment*	ML *Treatment*	*Difference*	*p-value*
WTA_{mug}	4.31	3.36	0.95	0.005
$ln(WTA_{mug})$	1.22	0.86	0.36	0.007
$WTA_{mug} - WTA_{pen}$	0.36	−0.28	0.64	0.051
$ln(WTA_{mug}) - ln(WTA_{pen})$	**0.13**	**−0.09**	**0.22**	**0.055**
WTA_{pen}	3.95	3.64	0.31	0.348

Table 2: Original results and replication results for regression (1) to (5) in Table II in the original study. The dependent variable is $ln(WTA_{mug})$.

	Original Study					***Replication Study***				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
Treatment MH	0.194	0.266**	0.306**	0.308**	0.290**	0.358***	0.266***	0.269**	0.278***	0.292***
	(0.143)	(0.124)	(0.128)	(0.126)	(0.130)	(0.128)	(0.104)	(0.106)	(0.107)	(0.108)
$ln(WTA_{pen})$		0.524***	0.560***	0.568***			0.599***	0.597***	0.611***	
		(0.091)	(0.093)	(0.091)			(0.056)	(0.057)	(0.058)	
Female			−0.199	−0.213*	−0.191			0.197*	0.221**	0.190*
			(0.131)	(0.128)	(0.133)			(0.108)	(0.110)	(0.111)
Age			−0.052*	−0.050*	−0.046*			−0.016	−0.013	−0.020
			(0.027)	(0.026)	(0.027)			(0.020)	(0.020)	(0.020)
Day indicators	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Other controls	No	No	No	No	#	No	No	No	No	#
Observations	106	104	104	104	106	249	248	248	248	249

Note: Dependent variable $ln(WTA_{mug})$. (1), (2), (3), and (5) are OLS regressions; (4) displays predicted marginal effects from a tobit that takes into account censoring at $WTA_{mug} = 9.57$. All regressions contain a constant, and except for (1) also an indicator for $WTA_{pen} = 9.57$ (which indicates that a subject's WTA for the pen may be censored). Standard errors in parentheses. # Cubic in WTA_{pen}.

*** Significant at the 1 percent level
** Significant at the 5 percent level
* Significant at the 10 percent level