Replication of

Reference Points and Effort Provision

by Abeler, J./Falk, A./Goette, L./Huffman, D. (2011)
in: The American Economic Review, 101(2), pp. 470–492.

Replication Authors:
Felix Holzmeister, Jürgen Huber, Michael Kirchler and Michael Razen

---

Abeler et al. find that in a real effort task, effort provision and earnings are higher if the expectations-based reference point is high (HI treatment) compared to low (LO treatment).

Hypothesis to bet on:

Subjects exert more effort (leading to higher earnings) in a real effort task if the expectations-based reference point is increased (a comparison of the average accumulated earnings in the real effort task between the LO treatment and the HI treatment).

Power Analysis

The original p-value is reported as $p < 0.05$; we re-estimated the regression based on the posted data to get the exact p-value which is 0.046 based on a t-test of the coefficient (regression coefficient for "1 if HI treatment" dummy variable in the first regression in Table 1, an OLS model with the accumulated earnings in the real effort task as the dependent variable, p. 478): *"The treatment difference in effort provision is significant in an OLS regression where we compare effort in HI to effort in LO. We regress the accumulated earnings at which a subject stopped on a treatment dummy (see Table 1, column 1)."*

The original sample size is 120 participants (60 in treatment HI and 60 in treatment LO). To achieve 90% power the required sample size is 317 participants.

Sample

The sample for replication consists of 318 students at the University of Innsbruck in Austria, 159 subjects per each of the two treatments. As in the original study, we will exclude students with economics as a major and solely recruit subjects who had participated in no or only a few previous experiments.

Materials

We use the material of the original experiment (programmed in z-Tree) along with the original German instructions. The program is available on the journal's webpage, while the German instructions have been kindly provided by the authors.

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 "I. Design" (pp. 473–475) in the original study.

The work task in this experiment consists of counting the number of zeros in tables containing randomly ordered zeros and ones. The experiment involves two stages. Prior to the

first stage, subjects read the instructions (for the first stage only) and answer control questions. During the first stage, subjects have four minutes to count as many tables as possible. Thereby, the first stage serves to familiarize subjects with the task, permitting subjects to have a good understanding of how difficult the task is and how much they could earn in a given time before they know about the amount of the fixed payment (which is revealed only prior to the second stage).

After the first stage, subjects read the instructions for the second (and main) stage. Contrary to the first stage, subjects now can decide themselves how much and for how long they want to work. As soon as they push the "stop"-button on the screen, they get immediately paid and can leave. We employ the same procedure as in the original study and let subjects arrive for the experiment one at a time. Individual starting times are at least 20 minutes apart. Upon arrival, subjects are guided to one of three essentially identical, neutral rooms. Instructions and payments are also administered in their rooms.

Moreover, subjects do not earn their accumulated piece rate from the second stage for sure as they either receive a fixed rate or the accumulated piece rate earnings with equal probability. Which payment subjects receive is determined only after they have decided when to stop working. The two treatments HI and LO only differ in the amount of the fixed payment (€7 in the HI treatment, €3 in the LO treatment) which determines the rational expectation-based references. Subjects are randomly assigned to one of the two treatments.

After completing both stages, subjects will be privately paid in cash based on the same incentives and using the same show-up fee (€5) as in the original study (average earnings were €8.70 per subject in the original study).

Analysis

The analysis will be performed exactly as in the original article. That is, differences in effort provision between the two treatments HI and LO are estimated by an OLS regression with the accumulated earnings at which a subject stopped as the dependent and a treatment dummy as independent variable.

Differences from Original Study

The replication procedure is identical to that of the original study, with the only unavoidable deviation that the replication will be performed at the University of Innsbruck, Austria, in 2015, on students from the University of Innsbruck, whereas the original data was gathered at the University of Bonn, Germany, in 2007, on students from the University of Bonn. However, as in the original study, only subjects who did not have economics as a major and who had participated in no or only a few previous experiments are invited. The experiment will be conducted in German as in the original study.

While the original study also looked at several control treatments, the focus of the replication is only on the earnings difference between the HI and the LO treatment.

Replication Results

In the replication experiment, the average accumulated earnings in the real effort task are 7.01 in the HI- and 6.34 in the LO treatment (average accumulated earnings were 9.22 (HI) and 7.37 (LO) in the original study). Estimating the difference in effort provision between the two treatments with an ordinary least squares regression with the accumulated earnings at which a subject stopped as the dependent variable and a treatment dummy as independent variable yields a coefficient of 0.667 with a corresponding p-value of 0.160 (see Table 1). Accordingly, the null hypothesis of no difference in effort provision given a higher expectation-based reference point cannot be rejected. For the original study, the coefficient of the treatment dummy is 1.850 such that the relative effect size of the replication experiment equals 36.04% (0.667/1.850).

The first 96 observations of the replication experiments were conducted before the summer break in June and July 2015. The remaining 222 observations were gathered in September and October 2015. Since the weather was very hot and sunny in Austria this summer, we compare the two samples to rule out any concerns about implications of temperature differences. Neither regressing the average accumulated earnings in the real effort task on the treatment dummy only for the 96 observations gathered in June/July ($coef. = -0.275$, $p = 0.760$) nor the 222 observations gathered in September/October ($coef. = 1.074$, $p = 0.053$) results in an successful replication. Similarly, controlling for the room temperature does not change the replication results, neither in the entire sample ($coef. = -0.204$, $p = 0.362$) nor in one of the subsamples (June/Juli: $coef. = 0.060$, $p = 0.851$; September/October: $coef. = -0.225$, $p = 0.533$), although the average room temperature was 0.77°C higher in June/July ($p < 0.001$). Additionally including an interaction term "temperature x treatment" shows that the treatment effect does not vary with the room temperature significantly ($coef. = -0.504$, $p = 0.259$). Similarly, including a dummy variable for the observations gathered in September/October ($coef. = 0.436$, $p = 0.548$) and an interaction term "treatment x Sept/Oct" ($coef. = 1.349$, $p = 0.189$) depicts that the treatment effect does not significantly vary between the experiments conducted in June/July and September/October.

Unplanned Protocol Deviations

The replication experiments were conducted exactly in the way as described above without any deviations from the protocol.

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 36.04% and the p-value of the hypothesis test is 0.160.

Table 1: Treatment difference in effort provision (HI compared to LO treatment) in the original and replication experiments

	Original Study	***Replication Study***
1 if HI treatment	**1.850****	**0.667**
	(0.917)	(0.474)
Constant	7.370***	6.342***
	(0.648)	(0.335)
Observations	120	318
Adjusted R^2	0.03	0.00

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