Replication of

Imperfect Public Monitoring with Costly Punishment: An Experimental Study

by Ambrus, A. / Greiner, B. (2012)
in: The American Economic Review, 102(7), 3317–3332

Replication Authors:
Teck Hua Ho, Hang Wu

Ambrus and Greiner conduct six treatments combining punishment severity and noise (imperfect monitoring) in sequential Prisoner's Dilemma games. There is a surprising non-obvious U-shaped effect of severity on net earnings (earnings after paying punishment costs). Therefore, we will concentrate only on the basic effect of noise in the regular punishment condition.

> **Hypothesis to bet on:**
>
> When there is imperfect monitoring, allowing punishment reduces net earnings (i.e., earnings after punishment costs; a comparison of the regular punishment with noise treatment and the no punishment with noise treatment).

Power Analysis

The focal effect is the sum of the "regular punishment" coefficient and the "noise * regular punishment" coefficient in the last column in Table 2 (p. 3324), an OLS regression of earnings in each period. Footnote 16 reports that this summed effect is significant at $p < 0.05$ by an F-test (exact p-value $= 0.050$ from authors). As this test is based on a regression including more than the two treatments that will be replicated, the authors provided regression results based on only the two treatments to be replicated. The p-value of the treatment effect in this regression is 0.057 (based on a t-test of the coefficient of the dummy variable for the treatment with regular punishment and noise).

The original sample size is 117 participants in two conditions ($n = 57$ in the no punishment with noise treatment; $n = 60$ in the regular punishment with noise treatment). To achieve 90% power the required sample size is 340 participants.

Sample

In the original sample, 339 subjects participated in 12 sessions, with between 24 and 30 subjects per session. The sample for replication consists of 340 participants from the National University of Singapore. There are no exclusion criteria.

Materials

We use the material of the original experiment (programmed in z-Tree) along with the original instructions which have been made available 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. Experimental Design" (pp. 3320–3321) in the original article.

Upon arrival, participants are seated in front of computers at desks that are separated by dividers. The experiment starts after participants have read the written instructions and completed a short comprehension test at the screen.

At the beginning, participants are randomly and anonymously matched to groups of 3 that stay constant over all 50 rounds. In each round, each of the 3 participants in a group is endowed with 20 tokens and is asked to either contribute all or none of these tokens to a group account. If the endowment is kept, it benefits the participant by 20 points, while if the endowment is contributed, it benefits each of the 3 group members by $0.5 \times 20 = 10$ points. After all group members made their choice(s) simultaneously, they are informed about the outcome of the game. In the noise treatments only a "public record" of each group member's choice is displayed. If a group member did not contribute, then the public record will always indicate "no contribution." If the group member contributed, there is a 10 percent chance that the public record shows "no contribution" rather than "contribution." Participants are fully informed about the structure of the noise.

In the punishment treatments, subjects participate in a second stage in each round. Here they are asked whether they want to assign up to five deduction points to the other two members of their group. In the regular punishment treatments, each assigned deduction point implies a reduction of three points from the punished group member's income. Received punishment is capped at the earnings from the public goods game in the same round, i.e., the punishment deduction cannot exceed the within-round earnings, however, while a punisher always has to pay for assigned punishment points.

At the end of the experiment, participants fill out a short demographic survey. They are then privately paid in cash based on their cumulated experimental earnings plus a show-up fee of AU$5.00 (average earning were AU$28.94 per subject in the original study).

Analysis

The analysis will be performed in the same way as in the original article; but based on a regression including only the two treatments in the replication (rather than all treatments as in the original article; see above). An OLS regression using per round earnings (net of punishment costs and penalties) will be run replicating the result in Table 3, rightmost column. The hypothesis will be tested with a t-test of the regression coefficient of the treatment dummy variable for the treatment with regular punishment and noise.

Differences from Original Study

The replication procedure is identical to that of the original study, with some unavoidable deviations. This replication will be performed at National University of Singapore, in 2015, while the original data was gathered at the Australian School of Business Experimental Research Laboratory at the University of New South Wales in February and March 2010 and 2011. The experiment will be in English as in the original study.

Replication Results

In the replication study, when there is imperfect monitoring, allowing for regular punishment significantly reduces the net profits. This is consistent with the observation of the original study. Table 1 summarizes the results of ordinary least squares regressions of net earnings on the treatment dummy *regular punishment and noise* for the original and replication experiments. The treatment effect of -2.021 in the replication study is statistically significant with a p-value of 0.012 (t-test). Because of a lower standard error in the replication study the p-value is smaller compared to the original study (p-value of 0.057). As compared to the effect size of -2.914 in the original study, the relative effect size of the replication experiments equals 69.35% ($-2.021/-2.914$).

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 been replicated at a significance level of $\alpha < 5\%$. The relative effect size equals 69.35% and the p-value of the hypothesis test is 0.012.

Table 1: Comparison of OLS regression of net earnings on the treatment dummy *regular punishment and noise* in both treatments

	Original Study	*Replication Study*
Intercept	22.367***	23.544***
	(0.720)	(0.371)
Round	-0.014	-0.031***
	(0.018)	(0.010)
Regular punishment and noise	$\mathbf{-2.914^*}$	$\mathbf{-2.021^{**}}$
	(1.487)	(0.795)
Observations	5850	17850
R^2	0.030	0.018

Note: Ordinary least squares regressions; standard errors are provided in parentheses.

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