Replication of

Slow to Anger and Fast to Forgive: Cooperation in an Uncertain World

by Fudenberg, D./Rand, D.G./Dreber, A. (2012)
in: The American Economic Review, 102(2), 720–749.

Replication Authors:
Colin Camerer, Taisuke Imai, Gidi Nave

Fudenberg et al. vary the benefit to cost ratio b/c (the rewards to cooperation) in an indefinitely repeated prisoner's dilemma with noise (meaning that there is a specific probability that an intended move is changed to the opposite move where the other player can only observe the outcome). Cooperation is found to be higher in the treatment where there are cooperative equilibria ($b/c = 2$) compared to in the treatment where there are no cooperative equilibria ($b/c = 1.5$).

Hypothesis to bet on:

Cooperation in a repeated prisoner's dilemma with noise (a specific probability that an intended move is changed to the opposite move) is higher when there are cooperative equilibria (a comparison in the levels of overall cooperation between the $b/c = 1.5$ and the $b/c = 2$ treatment).

Power Analysis

The original p-value is 0.001 (logistic regression over all decisions with a $b/c = 1.5$ dummy variable (treatment) clustered on both subject and interaction pair; pp. 732–733, footnote 31): *"We see that there is markedly less cooperation when $b/c = 1.5$, both in the first round (1.5 versus 2, $p = 0.016$; 1.5 versus 2.5, $p = 0.003$; 1.5 versus 4, $p = 0.004$) and overall (1.5 versus 2, $p = 0.001$; 1.5 versus 2.5, $p < 0.001$; 1.5 versus 4, $p < 0.001$)."*

The original sample size is 124 participants (72 in the $b/c = 1.5$ treatment and 52 in the $b/c = 2$ treatment). To achieve 90% power the required sample size is 120 participants.

Sample

The sample for replication consists of 120 students drawn from the Claremont university subject pool. There are no exclusion criteria.

Materials

We use the material of the original experiment (programmed in z-Tree) along with the original instructions, both available at the journal's webpage.

Procedure

We will 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. 723–725) in the original study.

The infinitely repeated game is induced by having a known constant probability that the interaction will continue between two players following each round. We let the continuation probability be $\delta = 7/8$. With probability $1 - \delta$, the interaction ends and subjects are informed that they have been rematched with a new partner. There is also a known constant error probability that an intended move is changed to the opposite move. Our main conditions use $E = 1/8$; there are also control conditions with $E = 1/16$ and $E = 0$ (the replication will only use $E = 1/8$.) Subjects are informed when their own move was changed (i.e., when they made an error), but not when the other player's move was changed; they are only notified of the other player's actual move, not the other's intended move. Subjects are informed of all of the above in the experimental instructions.

To rematch subjects after the end of each repeated game, we use the turnpike protocol as in Dal Bo (2005). Subjects are divided into two equal-sized groups, A and B. A-subjects only interact with B-subjects and vice versa, so that no subject ever plays twice with another subject, or with a subject who plays with a subject they had played with, so that subjects cannot influence the play of subjects they interacted with in the future. Subjects are informed about this setup.

To implement random game lengths, we pregenerate a sequence of integers $t_1, t_2, \ldots,$ according to the specified geometric distribution to use in all sessions, such that in each session every first interaction lasts t_1 rounds, every second interaction lasts t_2 rounds etc. (The replication will use the same interaction game lengths as in the original paper.)

To get the same average number of interactions (11 in $b/c = 1.5$ and 11.5 in $b/c = 2$, respectively) per subject we will include 22 subjects in each session, and run 3 sessions of each treatment (i.e. in total 6 sessions and 132 subjects). Subjects will be randomly allocated to each treatment.[1]

After all rounds have been played, subjects will be privately paid in cash based on the sum of their earnings using the same show-up fee ($10) and incentives as in the original study (average earnings were $22 per subject in the original study, including earnings of the dictator game). The replication will not include the dictator game and an additional temporal discounting task that were conducted following the prisoner's dilemma game.

Analysis

The analysis will be performed exactly as in the original article, using a logistic regression over all individual decisions, with a b/c value dummy as the independent variable, clustered on both subject and interaction pair.

Differences from Original Study

The replication procedure is identical to that of the original study, except for some unavoidable deviations. This replication will be performed at the Claremont CNS lab subject pool. The original data was gathered at the Harvard Decision Science Laboratory at Harvard University in Cambridge MA, USA, in 2009-2010, on students at Boston-area colleges and universities. The experiment will be in English as in the original study.

The original study also looks at first round cooperation, other benefit to cost ratios and error rates, but the focus of the replication is only on the difference between overall cooperation in $b/c = 1.5$ and $b/c = 2$.

The replication will not include the dictator game and an additional temporal discounting task that were conducted following the prisoner's dilemma game.

Replication Results

The total of 128 subjects (80 in $b/c = 1.5$ and 48 in $b/c = 2$) participated in the replication experiments. The effect is tested using a logistic regression over all individual decisions, with a b/c value dummy (which takes 1 when $b/c = 1.5$ and 0 otherwise) as the independent variable, clustered on both subject and interaction pair. As in the original study, there is less

[1] In the original article, there is a dictator game once the repeated prisoner's dilemma games are over. Since participants were not aware of the dictator game when playing the prisoner's dilemma, we will not include the dictator game in the replication.

cooperation when $b/c = 1.5$, both in the first round ($p = 0.015$) and overall ($p < 0.001$). See Table 1, columns (3) and (4).

The original effect size (for overall decisions) is −0.627 (the estimated coefficient on the treatment dummy). The replication effect size is −0.605. Therefore, the relative effect size is 96.49% (−0.605/−0.627).

Unplanned Protocol Deviations

Due to the capacity limit of Claremont CNS lab, 16 subjects participated in the total of 8 sessions (5 sessions for $b/c = 1.5$ treatment and 3 sessions for $b/c = 2$ treatment). Recruiting for the last three sessions was unusually difficult. We doubled the show up fee and it is possible that intensive recruiting may have generated a difference in the volunteers who self-selected to participate. Due to a mistake in selecting the treatment for the last session, the replication sample is unbalanced as the $b/c = 2$ treatment is undersampled. Though, as outlined below, robustness checks with treatment-balanced samples by downsampling yield qualitatively similar results. Apart from that, the replication experiment has been conducted exactly the way as outlined above, without further 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 96.49% and the p-value of the hypothesis test is < 0.001.

One potential concern is the unbalanced nature of our sample (i.e., we undersampled subjects in $b/c = 2$ treatment, because of a mistake in selecting the treatment for the final session). As a robustness check, we estimate the same model using a treatment-balanced sample by downsampling (using only 48 subjects in each treatment, by dropping the data from the last two $b/c = 1.5$ sessions). The result is reported in columns (5) and (6) in Table 1. Again, we observe significantly less cooperation when $b/c = 1.5$, both in the first round ($p = 0.012$) and overall ($p < 0.001$).

Table 1: Logistic regression over first-round/all individual decisions in the original and the replication experiments

	Original Study		***Replication Study***		***(Downsampled)***	
	(1)	(2)	(3)	(4)	(5)	(6)
$b/c = 1.5$ dummy	−0.757**	**−0.627*** **	−0.510**	**−0.605** **	−0.611**	−0.741***
	(0.314)	(0.182)	(0.209)	(0.160)	(0.244)	(0.177)
Constant	0.945***	−0.034	0.188	−0.073	0.188	−0.073
	(0.247)	(0.139)	(0.157)	(0.125)	(0.157)	(0.125)
Observations	1428	11972	1024	8320	768	6240

Note: Standard errors are clustered by subject and group.
*** Significant at the 1 percent level
** Significant at the 5 percent level
* Significant at the 10 percent level