Replication of

On the Selection of Arbitrators

by de Clippel, G./Eliaz, K./Knight, B. (2014)
in: The American Economic Review, 104(11), 3434–3458.

Replication Authors:

Teck Hua Ho, Hang Wu

From the perspective of implementation theory, de Clippel et al. compare the performance of two different mechanisms for the selection of arbitrators. They compare the veto-rank mechanism (VR) and shortlisting mechanism (SL), and find that the latter performs better in terms of efficiency (average aggregate payoff). The main focus of the paper is on the performance of VR vs. SL for two different preference profiles, Pf2 and Pf4. We randomly picked one of the two preference profiles for the replication and ended up with Pf2.

> ### Hypothesis to bet on:
>
> Efficiency (average aggregate payoff) is higher with the social choice mechanism Shortlisting (SL) than with the Veto-Rank (VR) mechanism for preference profile Pf2.

Power Analysis

The original p-value is 0.001 (p. 3449): *"Moreover, while the differences are not statistically significant for Pf3, the differences are statistically significant with $p = 0.001$ for Pf2 and $p = 0.055$ for Pf4. This provides further evidence that SL outperforms VR."*

The original sample size is 158 participants (70 in the VR treatment and 88 in the SL treatment). To achieve 90% power the required sample size is 153 participants.

Sample

The sample for replication consists of 156 students (78 participants per treatment, matched in pairs) at the National University of Singapore (NUS). As in the original article, we recruit only undergraduate students from NUS.

Materials

We use the material of the original experiment (programmed on the web) along with the original instructions, available at the journal's webpage.

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 "A. Design" (pp. 3443–3444) in the original study.

In each treatment, an even number of subjects is presented with a set of five alternatives labeled a, b, c, d, e, and they are randomly matched to play one of the mechanisms (VR or SL depending on the treatment). Each treatment consists of 40 rounds divided into four blocks of 10 rounds each. In each round subjects are randomly re-matched. In each of the 4

blocks, subjects have the same preference relation over the five options, but these preferences change from one block to another (i.e. in total there are four distinct preference profiles). Preferences are induced by assigning each of the options (a, b, c, d, e) a dollar value in the set $1.00, $0.75, $0.50, $0.25, $0.00. The first profile Pf1 consists of completely opposed rankings. The second profile, Pf2, represents partial conflict of interest involving only the top two options. The third profile, Pf3, displays a similar partial conflict of interest at the top, but this time with the addition of a focal compromise. The fourth profile, Pf4, captures cases where the veto-rank mechanism admits (undominated) Nash equilibria whose outcome does not belong to the veto-rank SCR. In each session the four induced preference profiles appear in a different order. The four orders are: Pf1-Pf2-Pf3-Pf4; Pf4-Pf3-Pf2-Pf1; Pf1-Pf3-Pf2-Pf4; Pf4-Pf2-Pf3-Pf1.

In the SL treatment the subject who is chosen to be Player 1 moves first and selects a short-list of 3 alternatives (out of: a, b, c, d, e). The second player selects one alternative from this shortlist and this alternative is implemented. In the VR treatment both players simultaneously remove two alternatives from the list of 5 alternatives, and rank the remaining three alternatives. The alternative with the minimum sum of ranks among alternatives which have not been vetoed is selected and implemented.

After the subjects read the instructions they are presented with a short quiz, testing their understanding of the game. When the subjects finished answering the quiz, they are presented with the correct answers.

The number of subjects per session was 17.5 on average for the VR sessions and 22 subjects per session for the SL treatment. We therefore run sessions with about 20 subjects per session and run four sessions of each treatment (in total 8 sessions and 156 subjects). Subjects are randomly allocated to each treatment.

After all rounds have been played, subjects will be privately paid in cash based on the sum of their earnings across all 40 rounds and using the same show-up fee ($10) as in the original study, average per round earnings in the Pf2 block for the VR session was about $1.67 and that for SL session was about $1.72.

Analysis

The analysis will be performed exactly as in the original article. We use the mean-comparison test which compares the aggregated average payoffs between the VR session and the SL session.

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 the National University of Singapore, in 2015, on students from the National University of Singapore, while the original data was gathered at NYU in New York, USA, in 2010 and 2011, on undergraduate students from NYU. The experiment will be in English as in the original study.

The original study tests several preference profiles: though, all preference profiles in the experiment are replicated, the analysis focuses only on profile Pf2.

Replication Results

In the replication experiment, the average per round aggregate payoff of 1.705 in the SL treatment is 0.167 higher than in the VR treatment. The difference is statistically significant with $p < 0.001$ (mean-comparison t-test, $t = 7.714$).

Table 1 summarizes the average per-round aggregate payoffs under the two social choice mechanisms as well as the payoff differences for both the original and the replication study. For Pf2 the outperformance of SL over VR in terms of efficiency is more pronounced in the replication study than the original study. As a measure of effect size we use the difference in average per-round aggregate payoffs between the two treatments. Accordingly, the relative effect size of the replication experiment is 321.12% (0.167/0.052).

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 321.12% and the p-value of the hypothesis test < 0.001.

Table 1: Comparison of average per-round aggregate payoffs under the two mechanisms SL and VR for Pf2 in the original and the replication experiments

	Original Study	*Replication Study*
VR mechanism	1.669	1.538
SL mechanism	1.721	1.705
Difference	**0.052*****	**0.167*****

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