

Replication of Phillips and Cushman 2017

“Morality constrains the default representation of what is possible”

PNAS 114(18), 4649-4654.

<http://www.pnas.org/content/114/18/4649>

The original paper includes several studies. We randomly chose study 1a. In this between-subject experiment, participants are randomized to one of two conditions: they are either forced to make their judgments very quickly (≤ 1500 ms) or are asked to reflectively deliberate on the possibility of each event (≥ 1500 ms), while in both cases making judgments of whether various events are possible or impossible. The main focus is on the comparison of the judgments of the possibility of immoral events between the high and low time pressure conditions. Participants judge immoral events as more impossible when they are forced to make judgments quickly compared to when they are asked to reflectively deliberate.

Hypothesis to replicate and bet on: Participants judge immoral events as more impossible when they are not able to deliberate compared to after deliberating. To evaluate this hypothesis, the authors perform a z -test ($z = 5.423$, $p < 0.001$); p. 4650. This particular test was chosen since it was the key result in study 1a.

Criteria for replication: The criteria for replication are an effect in the same direction as the original study and a p -value < 0.05 in a two-sided z -test.

Power analysis: The original study had 460 participants after exclusions. The standardized effect size (Cohen's d) was $d = 0.506$. To have 90% power to detect 67% of the original effect size, a sample size of $n = 370$ is required. Since we require the replication sample size to be at least as large as in the original study, the replication experiment will use a sample size of $n = 460$ and the power will thus be $>90\%$ to detect 67% of the original effect size.

Sample: The original paper mentions no restrictions on who could participate but that all participants were recruited through *TurkPrime*. We will do the same. The original study also did two kinds of exclusions: One at the participant level (by creating an average response time for each participant across all trials and excluding all responses above 6 seconds from that average; excluding participants in the reflective condition who had an average response time that was less than 1 second and excluding participants in the speeded condition who had an average response time below 800ms; this led to an exclusion of 38 participants); and one at the trial level (excluding all trials with response times less than 500ms and excluding all reflective trials with response times less than 1.5 seconds; this led to an exclusion of 89 observations). We will use the same exclusion criteria. We will make sure that participants can only participate once from the same account in this specific study, and we will only recruit participants with a HIT approval rate of 95% or above. We will also check all IP addresses via <https://www.ipqualityscore.com/>; and we will remove any participants where one or more of the following is true: fraud score ≥ 85 ; TOR = True; VPN = True; Bot = True; abuse velocity = high. The replication sample size is the sample size after any exclusions of participants.

Materials: We will use the same material as in the original study, kindly provided by the original authors. In particular, the experiment will be conducted using the original *Testable* survey.

Procedure: We will closely follow the procedure of the original experiment. The following summary of the experimental procedure is therefore largely based on the description of the experiment in the article (p. 4650) and the section “Materials and Methods” (pp. 4653-4654).

Participants will first be shown a Captcha, and will thereafter provide informed consent. After this we will include an attention check that participants will need to pass to continue to the study. This attention check is in addition to any other potential attention check(s) used in the original study. Participants will be asked to make judgments in six contexts. After a description of each context, they will be shown a series of candidate events, one at a time, and asked to press a key to indicate whether they think the event is possible or impossible. The between-participant manipulation is the time to make this judgment: participants will be forced to make the judgment either very quickly (≤ 1500 ms) or will be asked to reflectively deliberate on the possibility of each event (≥ 1500 ms).

Here is an example of a context from the original study: “Josh is on the way to the airport to catch a flight for a hunting safari in Africa. He leaves with plenty of time to make it there, but his car breaks down on the highway. Now Josh is sitting in his car near a busy intersection and knows he needs to get to the airport soon if he is going to catch his flight.”

We will include 144 different events, which were designed by the original authors to fall into five categories: 48 ordinary events that do not violate any norms, 24 events that violate statistical norms, 24 events that violate physical laws, 24 events that violate moral rules, and 24 events that violate norms of rationality. Examples of events in all categories are the following: “ii) Is it possible or impossible for Josh to (a) hail a taxi at the intersection, (b) fix his car by banging on it, (c) teleport himself to the airport, (d) sneak onto public transportation, or (e) sell his car for a ride to the airport?” Our replication focuses on immoral events but we will include all events.

Analysis: The analysis will be performed as in the original paper. In particular, a z -test will be used to determine if participants’ judgments of the possibility of immoral events are affected by deliberation. To conduct this test, we will analyze possibility judgments for the immoral events using a generalized linear mixed effects model with speed vs. deliberation as fixed effects and random intercepts for each participant and each of the six scenarios (as shown below).

```
summary(glmer(responses ~ condition3 + (1|turkID) + (1|condition2), data=d1a[d1a$condition1=="immoral",], family = "binomial"))
```

Subject payments: We are standardizing payments across all replications so that studies have a certain show-up fee depending on the expected length of the study, with an hourly wage from the show-up fee of \$8 and a minimum payment of \$1 (for studies with incentive payment we use the same incentive payment as in the original study; and this payment is paid in addition to the show-up fee). If we have problems recruiting, we will increase the show-up fee.

