Reference of the paper: Nahmias, E., Coates, D. J., & Kvaran, T. (2007). Free will, moral responsibility, and mechanism: Experiments on folk intuitions. Midwest studies in Philosophy, 31(1), 214-242.
Replication team: Jonathan Phillips & Fiery Cushman

Which study in the paper do you replicate?

Experiment 1

If it is not the first study, please explain your choice:

NA

In this study, what is the main result you will focus on during replication? Please give all relevant statistical details present in the paper:

The primary effect is the impact of the description of mechanism of determinism (neuropsychological vs. psychological) on judgments of moral responsibility, free will, blame, praise and whether the agent's decisions are 'up to' them. This effect of mechanism was found in judgments about a non-real world called 'Erta' and the real world.

The original effect of this manipulation on each of these five judgments was reported in footnote 13 as follows: "Up To: $F(1, 301) = 36.45$, $p < .001$; FW: $F(1, 298) = 36.375$, $p < .001$; MR: $F(1, 295) = 52.81$, $p < .001$; Blame: $F(1, 294) = 44.973$, $p < .001$; and Praise: $F(1, 297) = 18.84$, $p < .001$."

What is the corresponding hypothesis?

These results support the two corresponding hypotheses stated on page 232:

(1) most people will judge that determinism is not threatening to FW and MR if determinism is described in nonmechanistic (psychological) terms.
(2) significantly more people will judge determinism to be threatening to FW and MR if determinism is described in mechanistic (neuroscientific) terms.

What is the corresponding effect size?

Free will: d = 0.69
Moral responsibility: d = 0.83
Decision up to agent: d = 0.69
Blame: d = 0.77
Praise: d = 0.5

Was the original effect size:

- Not explicitly reported in the original paper, but inferable from other information present in the original paper

What is the corresponding confidence interval (if applicable)?

> *Free will*: 95% CI = [0.46, 0.93]
> *Moral responsibility*: 95% CI = [0.6, 1.07]
> *Decision up to agent*:95% CI = [0.46, 0.93]
> *Blame*: 95% CI = [0.54, 1]
> *Praise*: 95% CI =[0.27, 0.73]

Was the original confidence interval:

- Inferable from information present in the original paper.

From which population was the sample used in the original study drawn? (Which country, language, students/non-students, etc.)

> According to page 225 of the manuscript: “For this study, 1,124 Georgia State University undergraduates participated for extra credit in a critical thinking course. Participants were contacted through e-mail correspondence and randomly assigned to one of the eight versions of the experimental task. Participants were 41.2 percent male and 57.5 percent female, with 1.3 percent declining to give gender information.”

Was the nature of the original population:

- Explicitly reported in the original paper

What was the original sample size (*N*):

> 1124 across all 8 conditions; 307 across the 4 conditions being replicated (after exclusion)

Was the original sample size:

- Explicitly reported in the original paper

Does the study involve a selection procedure (e.g. comprehension checks)?

- YES

If YES, describe it briefly:

There were two different comprehension checks used in the study. The first was a true-false comprehension question that was completed after participants first read the vignette. The second was a similar true-false question that was completed after the depend

measures to ensure that they had kept the relevant details about the vignette in mind while answering the key questions. Participants were excluded from the analyses if they answered either of these questions incorrectly.

Were all the steps of the selection procedure (including, e.g., comprehension checks):

- As explicitly stated on page 226 in the paper: “We excluded from analysis (and from results presented in Table 1) all participants who did not complete the survey in full and all participants who responded incorrectly to either of the two manipulation checks (twenty-two percent of all participants).”

Overall, would you say that the original paper contained all the information necessary to properly conduct the replication:

- NO

If NO, explain what information was lacking:

> Many of the specific measures and materials were not included in the original manuscript. Some details about the procedure completed by participants was also not included. This information, however, was provided by the first author on request.

Power analysis and required sample size:
(Please, describe briefly the power analysis you conducted to determine the minimum required sample size. If the original effect is a null effect, just describe the required sample size you obtained by doubling the original sample size.)

> A power analysis using the pwr.t.test function in the pwr package in R suggested that to achieve .95 power with a significance level of .05, we would need a sample size (per group) of ~56 for the *free will measure,* ~39 for the *moral responsibility measure*, ~56 for the ‘up to the agent’ measure, ~45 for the *blame measure*, and ~105 for the *praise measure*. To ensure adequate power, we decided to collect a sample somewhat larger than the one necessary for replicating the weakest effect (praise), resulting in the decision to recruit 250 participants.

Projected sample size:
(Please, describe the actual sample size you plan to use in the replication.)

> 250 participants