Replication of Kouchaki and Gino, 2016
"Memories of unethical actions become obfuscated over time"
PNAS, 2016, 113 (22), 6166-6171.
https://www.pnas.org/content/113/22/6166

The original paper includes several studies. We randomly chose study 7b. In this between-subject experiment, participants are recruited for a study that is run at two points in time ("time 1" and "time 2"), with three days in between. Participants play a die-throwing game at time 1. Participants are randomized into a treatment in which cheating in the die-throwing task is possible or a treatment in which cheating is impossible. At time 2 (three days later), participants engage in a Boggle task where they are asked to identify as many four-letter words as possible from adjacent letters (including corners) in a three-by-three letter grid within two minutes. Participants in the likely-cheating condition are more likely to cheat at time 2 (i.e. in the Boggle task).

Hypothesis to replicate and bet on: The possibility to cheat in a die-throwing game results in more dishonesty in a Boggle game after three days. The hypothesis is evaluated using a χ^2-test between the percentages of cheating participants in the likely cheating treatment vs. the no-cheating treatment ($\chi^2 = 10.48$, $p = 0.001$, $n = 258$, Cramer's $V = 0.20$); p. 6170.

Criteria for replication: The criteria for replication are an effect in the same direction as in the original study and a p-value < 0.05 in a two-sided χ^2-test.

Power analysis: The original study had 258 participants who completed the follow-up online survey (91% response rate). The standardized effect size (Cohen's d) was $d = 0.403$. To have 90% power to detect 67% of the original effect size, a sample size of $n = 583$ is required.

Sample: As in the original study, we will restrict our HITs to US participants and only include participants who take both parts of the study in the data analysis. We will make sure that participants can only participate once (in both parts) 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, we will use the original *Qualtrics* survey. The first part of the survey used in the original experiment is no longer usable, thus the corresponding part will be re-programmed by the replication team.

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 (pp. 6169–6170) and the section "SI Study 7b" (pp. 5–6) in the Supplementary Information.

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. The experiment is a two-part online study that will require participants to engage

in a series of tasks and answer questions at two different points in time. At time 1, participants will play a die-throwing game. In this game, participants will be asked to throw a virtual six-sided die 20 times to earn points that will be translated to real dollars and added to participants' final payment. The instructions of the game will remind participants that each pair of numbers on opposite sides of the die adds up to 7: 1 vs. 6, 2 vs. 5, and 3 vs. 4. The instructions will refer to the visible side that is facing up as "U" and the opposite, invisible side that is facing down as "D." In each round, the number of points that a participant scores will depend on the throw of the die as well as on the side that has been chosen in that round. Each round will consist of one throw. For instance, if a participant chooses "D" in his/her mind and the die outcome turns up to be "4," the participant will earn 3 points for that throw, whereas if a participant chooses "U" in his/her mind, he/she will earn 4 points. Across the 20 rounds participants can earn a maximum of 100 points. Each point is worth 10 cents, so participants can make a maximum of $10.

Participants will be randomly assigned to one of two conditions: likely-cheating vs. no-cheating. In the likely-cheating condition, participants will be asked to choose mentally between U and D before every throw, and after each throw they will indicate the side they had chosen before the throw. In the no-cheating condition, participants will also be asked to choose mentally between U and D before every throw, but in this case they will be asked to report their choice before throwing the virtual die.

Three days later, at time 2, participants will be asked to answer questions about their memory of the die-throwing task with the Autobiographical Memory Questionnaire (AMQ; Rubin et al., 2003). The AMQ measures people's autobiographical memory with eight items (e.g., "*As I think about the coin-toss task that night, I can actually remember it*").

Next, participants will be given an opportunity to cheat: In particular, they will engage in the Boggle task in which participants have to identify as many four-letter words as could be constructed from adjacent letters (including corners) in a three-by-three letter grid within a 2-minute timeframe. For each correct word, participants will be given a $0.50 bonus. Participants will be instructed that letters cannot be used twice, that all letters need to be adjacent and that proper nouns (e.g., Paul, Ural) do not count. Participants will also be informed that the survey will advance automatically after two minutes. After reading the instructions, participants will advance to the timed Boggle task.

Once participants finish the Boggle task, or when the time has expired, they will be asked how many words they identified correctly. Following a page break, participants will be presented with the original letter matrix and asked to type in the words they had identified so that they can be paid their correct bonus. Participants will be informed that they have only one minute to enter the words they have identified, to discourage identifying new words at this stage; the instructions will state that this timeframe should be sufficient, because they will simply be entering the words they have already identified.

The measure of over-reporting will be calculated by subtracting the number of words participants entered from the number they reported when they first saw the nine-letter box. The hypothesis to be replicated compares the fraction of participants who over-reported the number of correctly identified words between the likely-cheating and the no-cheating condition.

Analysis: The analysis will be performed as in the original paper. That is, we will compare the fraction of people who cheat in the Boggle task (i.e., the fraction of people who over-report

their performance) between participants in the likely-cheating condition and the no-cheating condition using a χ^2-test of equal proportions.

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.