Does honesty require time?

Two preregistered replications of Experiment 2 of Shalvi, Eldar, and Bereby-Meyer (2012)

Ine Van der Cruyssen[2*], Jonathan D'hondt[1*], Ewout Meijer[3], & Bruno Verschuere[2]

[1] Ghent University

[2] University of Amsterdam

[3] Maastricht University

Corresponding author**:** Bruno Verschuere (b.j.verschuere@uva.nl), Department of Clinical Psychology, University of Amsterdam, Nieuwe Achtergracht 129B, 1018 VZ Amsterdam, The Netherlands

Abstract

(current: 140 words; max 150 words)

Shalvi, Eldar, and Bereby-Meyer (2012) found across two studies (each $N = 72$) that time-pressure increased cheating. These findings suggest that dishonesty comes naturally, while honesty requires overcoming the initial tendency to cheat. While statistically significant, a Bayesian reanalysis indicates that the original results had low evidential strength. In a replication attempt of their Experiment 2, time-pressure did not increase cheating ($N = 428$, $r_{pb} = 0.05$, $BF_{01} = 16.06$). One important deviation from the original procedure, however, was the use of mass testing. In a second - direct - replication with small groups of participants, time pressure also did not increase cheating ($N = 297$, $r_{pb} = 0.03$, $BF_{01} = 9.59$). These findings indicate that the original study may have overestimated the true effect of time pressure on cheating and/or the generality of the effect beyond the original context.

Keywords: Intuition, Cheating, Lying, Honesty, Replication, Moral Decision Making, Time Pressure

Whether it is Lance Armstrong using doping, Diederik Stapel publishing fake data, the Enron executives overstating the company's earnings, or people underestimating their earnings on tax forms, cheating occurs in many aspects of life. Framed within the dual-process framework of Kahneman (2011) – that decision making results from the interplay between a fast, automatic and a slow, reflective mode of thinking –, one can wonder whether the tendency to serve one's best interest by cheating is automatic, and that honesty requires deliberation (Bereby-Meyer & Shalvi, 2015).

A straightforward way to examine how Kahneman's 'thinking fast' versus 'thinking slow' affects moral decision making is to manipulate the allotted time to decide on whether to cheat or behave honestly. Shalvi, Eldar, and Bereby-Meyer (2012) reasoned that under time-pressure, people will be more likely to follow their initial tendency, and dishonestly serve their self-interest. In two experiments, participants privately rolled a die under a cup. Payout was based on the self-reported outcome of the roll, with a higher number reported corresponding to a higher pay. As pay-out was entirely based on self-report, there was the opportunity and the financial incentive to cheat and report a higher than actual outcome. Critically, participants either had to roll and report their die roll outcome within 8s, or without any time limit. In both studies, the average reported die roll outcome was higher - i.e., more cheating - in the time-pressure condition than in the self-paced condition. These findings indeed suggest that dishonesty comes naturally, while honesty requires overcoming the initial tendency to cheat. To promote honesty, the authors therefore recommended to give someone time to think rather than to push for an immediate decision (see https://www.psychologicalscience.org/news/releases/when-do-we-lie-when-were-short-on-time-and-long-on-reasons.html).

The study by Shalvi et al. (2012) was theory-driven, has open materials and data, relied on an established manipulation, included manipulation checks, and is well-cited. There

are, however, also reasons to question that time pressure would increase dishonesty. First, limited cognitive capacity - also assumed to trigger automatic tendencies - led to decreased, rather than increased dishonesty in a variant of the die roll game (Foerster, Pfister, Schmidts, Dignath, & Kunde, 2013; for a critique see Shalvi, Eldar, & Bereby-Meyer, 2013). Second, in a game where participants could decide to send a dishonest message to another participant in order to receive more money themselves, time pressure increased, rather than decreased honest behavior (Capraro, 2017; Capraro, Schulz, & Rand, 2019; for a moderation explanation see Köbis et al., 2019). Third, a meta-analysis of 114 studies showed that lying systematically took longer than truth-telling (Suchotzki, Verschuere, Van Bockstaele, Ben-Shakhar, & Crombez, 2017), leading the authors to conclude that honesty – and not dishonesty – is the automatic tendency. Fourth, using a virtual die, time pressure was found to slightly increase cheating in a multiple dice roll paradigm (https://osf.io/j7pbc/), but not in a single die roll paradigm (https://osf.io/6fuhm/). To what extent these reasons cast doubt on the validity of the finding that time pressure increases dishonesty, or whether they can be explained by procedural differences remains unknown. Therefore, and because of the low diagnostic value of the original study (BF = 1.15; see Table 1), we set up an attempt to replicate Experiment 2 of Shalvi et al. (2012).

Reproducing original results

We first verified the original results by reanalyzing the data provided by the authors. Following their exact analysis strategy, we reproduced the key effect of interest. The two-tailed Mann-Whitney U test showed that participants in the time-pressure condition reported a significantly higher die roll outcome than participants in the self-paced condition. The effect was small to moderate (Experiment 1: *rank-biserial correlation or r_{rb}* = -0.22, 95% CI [-0.43, 0.01], corresponding to Cohen´s *d* = 0.44; Experiment 2: r_{rb} = -0.28, 95% CI [-0.48, -0.05],

corresponding to Cohen´s d = 0.58). We also reproduced ancillary effects, see

https://osf.io/fjca2/.

Table 1. Self-reported die roll outcome in the time pressure and the self-paced conditions of

Shalvi et al (2012) and our replications.

	Average Reported die roll (SD)		Z value Mann-Whitney U test	p value	Rank-biserial correlation (95%CI)	Bayes Factor
Study	Self-Paced	Time Pressure				
Shalvi et al. 2012, Experiment 2	3.42 (1.84)	4.38 (1.50)	2.17	.030	-0.28 (-0.48; -.05)	$BF_{10} =$ 1.15
Preregistered Replication 1 (mass testing)	3.76 (1.70)	3.61 (1.73)	-0.87	.808	+0.05 (-∞; 0.14)	$BF_{01} =$ 16.06
Preregistered Replication 2 (small groups)	3.76 (1.63)	3.67 (1.67)	-0.44	.671	+0.03 (-∞; 0.14)	$BF_{01} =$ 9.59

Note: The null hypothesis (H0) is that there is no difference between conditions. For the original study, the alternative hypothesis (H1) was that time pressure would affect cheating. For PDR1 and PDR2, the one-sided alternative hypothesis (H1) is that time pressure increases cheating. BF_{10} expresses how more likely the data are to occur under H1 than under H0. BF_{01} expresses how more likely the data are to occur under H0 than under H1. The Bayesian analyses were not preregistered.

Preregistered Direct Replication 1 (PDR1)

PDR1 was a preregistered replication of Experiment 2 reported in Shalvi et al. (2012), using a protocol approved by the original authors and a sample size more than five times that of the original study.

The current study deviated in several ways from the original study. The most notable deviation between PDR1 and the original studies is the session size. Sessions consisted of up to 6 participants in the original study, and of 228 and 233 participants in PDR1. The prime reason for the larger session size was that we wanted to make it feasible to test a substantially larger number of participants than the original study within a reasonable time. Note that the

original authors did not consider session size to be a key element of their design, and that a vignette study (see supplementary materials at https://osf.io/h8bjv/) provided no evidence that session size would affect the perceived chance of bonus payment. Secondly, when analyzing the original data, we noticed that - even before excluding participants that did not meet the 8s deadline - the sample sizes of the time pressure and the self-paced condition were unequal. The original authors clarified that this was a result of the randomization procedure: Up to 6 participants subscribed for a session, and all participants within a session were randomized to either the time pressure or the self-paced condition. Such between-session randomization is undesirable as the experimenter is no longer condition-blind, bringing about the risk that the experimenter influences the results (Rosenthal, Persinger, Vikan-Kline, & Fode, 1963). We therefore chose to randomize participants to the time pressure versus self-paced conditions within each session. Furthermore, there were also differences in the precise die rolling procedure (original study: shake cup back and forth on a table; current study: shake cup in hand), the software (original study: Eprime; current study: Qualtrics), test language (original study: Hebrew; current study: English) and the country (original study: Israel; current study: the Netherlands) that we return to in the General Discussion.

Method

The method was preregistered: https://osf.io/jez3g. All materials, data, and analytic scripts are available at https://osf.io/fnh9u/. The study was approved by the ethical committee of the Social and Behavioral Sciences faculty at University of Amsterdam and registered as number 2018-CP-9470. The protocol was carried out in accordance with the provisions of the World Medical Association Declaration of Helsinki.

Participants. As Camerer et al. (2018) showed that the effect size of replications is on average about 50% of the original effect size, we aimed for 90% power to detect an effect of

half the original size (d = 0.29; i.e. 50% of 0.58; note that the preregistration incorrectly mentions 50% of d = 0.66 implying a lower minimum required sample size of N = 366). For a one-sided independent Mann-Whitney U test with alpha = .05, the minimum sample size is N = 428. Anticipating preregistered exclusions (i.e., exclusion of participants who failed to report within the time limit in the time-pressure condition) we tested all attendees of two mass test sessions at the University of Amsterdam. Within each session, students performed a battery of tasks of which our task was the first one. Four hundred sixty-one first-year Psychology students participated. Thirty-three participants of the time-pressure condition were excluded because they did not report their die roll outcome within the time limit. The final sample contained 428 participants (71.73% female, 27.57% male, 0.70% other) with a mean age of 19.77 years (SD = 2.58 years): 198 participants in the time-pressure condition and 230 participants in the self-paced condition.

Procedure. Participants first gave informed consent. They were then randomly assigned to the time-pressure condition (i.e., roll their die and report the outcome within 8s) or the self-paced condition (i.e., roll their die and report the outcome at own pace) using Qualtrics permuted block randomization, which ensures an even distribution of participants across conditions. All participants received a paper cup with a lid and a six-sided die. They were invited to put the die in the cup, close it with the lid, shake the cup once, look through the hole in the lid to see the result of their roll, and report the outcome on the computer, see Figure 1. As a financial incentive for cheating, and per original instructions, participants were informed that *several participants would be randomly selected to receive a monetary reward* according to their reported die roll outcome. More specifically, they learned that their reported number would be multiplied by 2 (1 = €2, 2 = €4, etc.), leading to a bonus of up to €12. After reading the instructions on the computer screen, participants were guided to press a button

that started a timer measuring how long it took participants to roll the die and report their outcome. The instructions were delivered in English (https://osf.io/c2z4f/).

Figure 1. Illustration of the die rolling procedure used in PDR1.

To evaluate whether the participants believed a financial incentive was present and that their die roll was fully anonymous, we collected self-report ratings after the die roll game. Participants were asked to rate the following statements on a 5-point Likert scale (1 = Strongly disagree, 5 = Strongly agree): "Several students will receive a monetary reward for the dice under cup game.", "My dice roll was fully anonymous-only I could know what I rolled." and to indicate on a slider (0-100%) "What is the chance that you will get the reward?". To evaluate whether the participants had read the instructions attentively, they were asked to answer the multiple-choice question "The ratio between the dice roll and the possible reward is..." by choosing between the following options: "the reward (in euro) is equal to the outcome of the dice roll", "the reward (in euro) is two times the outcome of the dice roll", "the reward (in euro) is half of the outcome of the dice roll", or "the reward (in euro) is four times the outcome of the dice roll."

Results

Preregistered analyses.

Effect of time pressure on reported die roll outcome. Participants in the time-pressure condition did not report significantly higher die roll outcomes than participants in the self-paced condition, see Table 1.

Was there cheating? Similar to Shalvi et al. (2012), we evaluated whether there was cheating by comparing the observed distribution in each condition with the expected distribution of a fair roll. We found no evidence for cheating in the time-pressure condition, $X^2(5, N = 198) = 1.21$, $p = .944$, $V = 0.08$[1], nor in the self-paced condition, $X^2(5, N = 230) = 9.06$, $p = .107$, $V = 0.20$.

Exploratory analyses.

Time pressure manipulation check. Data from an extreme outlier (91s; more than 5*SD*s from *M*) was excluded for the time pressure manipulation check. Participants in the time-pressure condition took less time to report the outcome of the die roll ($M = 4.98$ s, $SD = 1.39$ s) than those in the self-paced condition ($M = 9.10$ s, $SD = 5.43$ s), $t(425) = 10.38$, $p < .001$, $d = 1.01$, 95% CI [0.80, 1.21], indicating a successful time-pressure manipulation.

Exclusions. Repeating the analyses without any exclusions, or in the subsample that expressed (strong) belief in the payment scheme (i.e., agreed or strongly agreed on the statement "Several students will receive a monetary reward for the dice under cup game") did not alter the pattern of findings, see https://osf.io/zqpw8/.

Self-report ratings. Self-report scales showed that most (90%) participants answered the control question regarding the payment scheme correctly. Most (84%) participants reported that they (strongly) believed their report was anonymous. Participants estimated their chance of winning the monetary reward at 23% ($SD = 27\%$). Unexpectedly, only a minority (35%) of the participants reported (strong) belief that several students would be paid for the die roll game.

---

[1] Cramer's V varies from 0 to 1, expressing the strength of association between two variables.

Discussion

We found no evidence that time-pressure increases cheating in the die roll paradigm. Self-report ratings revealed that participants may not have fully appreciated the financial benefit of cheating. The difference in session size may have resulted in a different social dynamic potentially influencing cheating behavior (Amir, Mazar, & Ariely, 2018).

Preregistered Direct Replication 2 (PDR2)

To rule out that the difference in results between our first replication study and the original study were due to the use of different session sizes, we ran another replication that used the same session size as the original, allowing up to 6 participants at once.

Method

The method was preregistered: https://osf.io/9bg3z. All materials, data, and analytic scripts are available at https://osf.io/xwzpc/.

To make maximum use of our resources, we preregistered to terminate data collection as soon as decisive evidence was found (Stefan, Gronau, Schönbrodt, & Wagenmakers, 2019). Specifically, after having tested double the sample size of the original study (i.e., 148 participants) we calculated, after each additional session, the Bayes Factor for the Bayesian Mann-Whitney test that assesses the differences between the time pressured and self-paced condition on the self-reported die roll outcome. We used a zero-centered Cauchy prior scaled at $r = 0.707$ (the JASP default setting) in all Bayesian analyses. If decisive evidence was reached for either the alternative hypothesis (i.e., that time pressure leads to higher reported outcome compared to self-paced condition; BF_{10} larger than 10) or the null hypothesis (i.e., that the time pressure manipulation does not affect reported outcome; BF_{01} larger than 10) we

would terminate data collection. At $N = 319$ ($N = 297$ inclusions), we reached decisive evidence for the null hypothesis ($BF_{01} = 10.14$) and we ended data collection.[2]

Participants. Participants were recruited in a university building at both the University of Amsterdam and Maastricht University to participate in a die rolling study. They received €2 for participation in the 10-minute study and were informed during recruitment that they could earn a bonus payment. Other than gathering a minimum of 3 and a maximum of 6 participants per session, there were no in- or exclusion criteria during recruitment. In the time-pressure condition, 22 participants were excluded because they did not report their die roll outcome within the 8s time limit. The final sample contained 297 participants (55% female) with a mean age of 21.60 years ($SD = 3.23$ years). About half of the participants had the Dutch nationality (62%) and also about half of the participants had Dutch as their mother tongue (60%). The time-pressure condition contained 138 participants (54% female, 46% male) with a mean age of 21.92 years ($SD = 3.61$ years). The self-paced condition contained 159 participants (43% female, 57% male) with a mean age of 21.31 years ($SD = 2.84$ years).

Procedure. Participants chose a die from a box with dice, then took a seat at one of the six individual tables[3] equipped with a laptop and a cup with a lid on it. General oral instructions were given to the group. After providing informed consent, all further instructions were given individually via the computer screen. Participants were invited to test whether the die was fair by rolling it a few times. Then, they were asked to put the die in the cup and close it with the lid. Participants were randomly assigned to either the time-pressure condition (8s

---

[2] This was the result of the first calculation. Due to the bootstrapping approach, there is variance in JASP 0.9's estimation of the BF. To illustrate, with 10 runs, we found the BF estimation to vary between 9.51 and 10.27. JASP 0.10, that was released after we ended data collection, has enhanced stability (BF varied only at the second decimal), and we therefore rely on its estimate for the BF´s. For PDR2, using JASP 0.10, we found $BF_{01} = 9.59$.

[3] At both test sites, we assured that participants could not see what other participants rolled or reported. The test location at Maastricht University had screens between tables. At University of Amsterdam, participants were seated with sufficient distance from each other, all facing a wall.

deadline) or the self-paced condition using Qualtrics permuted block randomization. As a financial incentive for cheating, and per original instructions, participants were informed that *several participants would be randomly selected to receive a monetary reward* according to their reported die roll outcome. Specifically, they learned that the bonus pay would be twice the reported outcome (1 = €2, 2 = €4, etc.), leading to a bonus of up to €12. The instruction page explaining the reward and the die under the cup task was displayed for a minimum of 30s, to prevent participants from going through the instructions without paying proper attention. After 30s, the next button appeared and pressing it started a timer measuring the time it took participants to roll their die and report their outcome (see Figure 1). The task was either presented in English or in Dutch (materials: https://osf.io/6c9qr/; for an exploratory analysis on the subsample tested in their native language see https://osf.io/c2n6h/).

After reporting their die roll outcome, participants were asked to provide their gender, major, age, nationality, and mother tongue. To gain insight how participants perceived the task, we collected self-report ratings after the die roll. All questions are reported on https://osf.io/6c9qr/. Here, we highlight that participants were asked to rate the statements "Several students will receive an extra monetary reward for the dice under cup task" and "My dice roll was fully anonymous-only I could know what I rolled" on a 5-point Likert scale ranging from strongly disagree to strongly agree. To evaluate whether the participants had read the instructions attentively, they were asked to answer the multiple-choice question: "The ratio between the dice roll and the possible extra reward is..." by choosing between the following options: the reward (in euro) is ...equal to / two times / half of / four times …the outcome of the die roll. They were also asked to indicate "What was your perceived time-pressure during the die roll?" on a 5-point scale ranging from very high to very low.

Deviations from original. Except for session size, the deviations between the current study and the original study are the same as for our first replication study (i.e., precise die rolling procedure, software, test language, and country).

Results

Preregistered analyses.

Time pressure manipulation check. Participants in the time-pressure condition took less time to report the outcome of the die roll ($M = 5.25$ s, $SD = 1.46$ s) than those in the self-paced condition ($M = 7.88$ s, $SD = 4.62$ s), $t(295) = 6.43$, $p < .001$, $d = 0.75$, 95% CI [0.51, 0.98], indicating a successful time-pressure manipulation.

Effect of time pressure on reported die roll outcome. Participants in the time-pressure condition did not report significantly higher die roll outcomes than participants in the self-paced condition, see Table 1.

Was there cheating? We found no evidence for cheating in the time-pressure condition, $X^2(5, N = 138) = 1.91$, $p = .861$, $V = 0.12$, nor in the self-paced condition, $X^2(5, N = 159) = 8.96$, $p = .111$, $V = 0.24$.

Exploratory analyses.

Exclusions. Repeating the analyses without any exclusions did not alter the pattern of findings, see https://osf.io/c2n6h/.

Self-report ratings. Most (84%) participants answered the control question regarding the payment scheme correctly. Most (89%) participants reported that they (strongly) believed their report was anonymous. Participants estimated their chance of winning the monetary reward at 36% ($SD = 27\%$). A majority (77%) of the participants reported that they (strongly) believed that several students would receive an extra reward for the die roll game. The perceived time pressure was higher in the time pressure group ($M = 3.11$, $SD = 1.04$) than in

the self-paced group ($M = 2.34$, $SD = 1.04$) $t(295) = 6.36$, $p < .001$, Cohen´s $d = 0.74$, 95% CI [0.50, 0.98].

Self-reported time pressure provides for an additional test of the time pressure effect. Within the time pressure condition, we examined whether greater perceived time pressure was related to higher reported die roll outcomes. The Kruskal-Wallis test on average reported die roll outcome for each of the 5 levels of perceived time pressure (very low, low, neutral, high, very high) was not significant, $\chi2$ (4, $N = 138$) = 3.16, $p = .532$, see Figure 2.
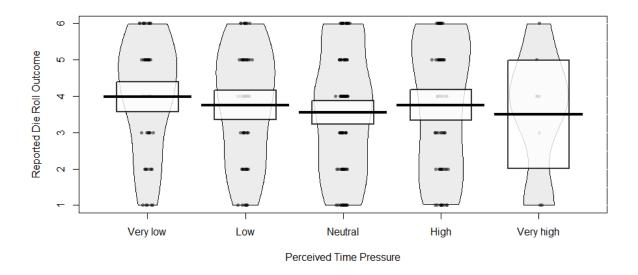

Figure 2. A pirate plot representing the reported die roll outcome at different levels of perceived time pressure. The bold lines represent the sample mean (with the upper and lower bounds of the rectangle representing the 95% CI around the mean). The dots represent individual raw data points.

General Discussion

What is people's automatic tendency in a tempting situation? Shalvi et al. (2012) found time-pressure, a straightforward manipulation to spark 'thinking fast' over 'thinking slow', provoked more cheating, and concluded that people's initial response is to serve their self-interests and cheat. We found no evidence that time-pressure increases cheating in the die

roll paradigm. There are three possible reasons why replications studies do not produce the same results as the original study: (1) Methodological problems of the replication study, (2) The original study overestimated the true effect size, or (3) Differences between the studies that moderate the effect (Wicherts, 2018).

The first possibility is methodological limitations of the replication study. In our first replication study, participants may not have fully appreciated the financial benefits of cheating. In our second replication study, relying on two test sites and offering the task in two languages may have increased error variance.

The second possible explanation is that the original effect overestimated the true effect size. The use of between-session rather than within-session randomization in the original study makes the experimenter aware of condition assignment and raises the possibility that the experimenter influenced the results (Rosenthal et al., 1963). Also, a single observation (i.e., single reported die roll outcome) per participant is likely to provide for a noisy measure. With low reliability, the results are more likely to vary per sample.

The third possible explanation is that the time pressure effect on cheating is influenced by the context and that differences between the studies explain the different results. Our replications differed in several ways from the original, most prominently the country where the study was run, namely Israel in the original versus The Netherlands in the replications. The difference in test site raises the possibility of cross-cultural differences in intuitive dishonesty. Perceived country corruption, for instance, is related to amount of cheating in the die under the cup game (Gächter & Schulz, 2016). Then again, the large meta-analysis by Abelerer et al (2018) found that cheating behavior varies little by country. Still, it seems worthwhile to explore whether the automatic tendency to cheat may vary with culture.

In both our studies, people were predominantly honest, and we found in fact no evidence of cheating[4]. Whereas it was originally reasoned that ″time pressure evokes lying even in settings in which people typically refrain from lying″ (Shalvi et al., 2012; pp. 1268), our findings point to the possibility that the time pressure effect is bound to settings that produce more pronounced cheating (e.g., when providing justifications for cheating).

In sum, our findings indicate that the original study may have overestimated the true effect of time pressure on cheating and/or the generality of the effect beyond the original context. The vast majority of our participants were honest – even under time-pressure. This finding casts doubt on whether peoples' intuitive tendency is to cheat, and fits better with a preference for honest behavior.

---

[4] This is far from exceptional and many studies found even lower average report and complete honesty (see Figure 1 of Abeler et al. 2019 or http://www.preferencesfortruthtelling.com/). That people could cheat to maximize self-profit without any punishment, but did not, fits with the meta-analytic conclusion of Abelerer et al (2019) that people cheat surprisingly little.

Author Contributions.

BV developed the study concept. IvdC and JD developed the study design of PDR1 and BV and EM provided critical input. BV and EM developed the study design of PDR2. Data collection for PDR1 was performed by IvdC and JD. Data collection for PDR2 was performed by SF, SV, and GD (see Acknowledgements). IvdC and JD performed the data analysis with critical input from BV and EM. All authors contributed to and approved the final version of the manuscript for submission

Declaration of Conflicting Interests. The author(s) declared no conflicts of interest with respect to the authorship or the publication of this article. BV has past and ongoing collaboration with the first and last authors of the original paper.

References

Abeler, J., Nosenzo, D., & Raymond, C. (2019). Preferences for Truth-Telling. *Econometrica*, *87*, 1115–1153. https://doi.org/10.1017/CBO9781107415324.004

Amir, O., Mazar, N., & Ariely, D. (2018). Replicating the Effect of the Accessibility of Moral Standards on Dishonesty: Authors' Response to the Replication Attempt. *Advances in Methods and Practices in Psychological Science*. https://doi.org/10.1177/2515245918769062

Bereby-Meyer, Y., & Shalvi, S. (2015). Deliberate honesty. *Current Opinion in Psychology*, *6*, 195–198. https://doi.org/10.1016/j.copsyc.2015.09.004

Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T. H., Huber, J., Johannesson, M., … Wu, H. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nature Human Behaviour*, *2*, 637–644. https://doi.org/10.1038/s41562-018-0399-z

Capraro, V. (2017). Does the truth come naturally? Time pressure increases honesty in one-shot deception games. *Economics Letters*, *158*, 54–57. https://doi.org/10.1016/j.econlet.2017.06.015

Capraro, V., Schulz, J., & Rand, D. G. (2019). Time pressure and honesty in a deception game. *Journal of Behavioral and Experimental Economics*, *79*(July), 93–99. https://doi.org/10.1016/j.archger.2007.01.047

Foerster, A., Pfister, R., Schmidts, C., Dignath, D., & Kunde, W. (2013). Honesty saves time (and justifications). *Frontiers in Psychology*, *4*, 473. https://doi.org/10.3389/fpsyg.2013.00473

Gächter, S., & Schulz, J. F. (2016). Intrinsic honesty and the prevalence of rule violations

across societies. *Nature*, *531*(7595), 496–499. https://doi.org/10.1038/nature17160

Hallsson, B. G., Siebner, H. R., & Hulme, O. J. (2018). Fairness, fast and slow: A review of dual process models of fairness. *Neuroscience and Biobehavioral Reviews*, *89*, 49–60. https://doi.org/10.1016/j.neubiorev.2018.02.016

Kahneman, D. (2011). *Thinking fast, thinking slow*. *London, UK: Farrar, Straus and Giroux*.

Köbis, N. C., Verschuere, B., Bereby-Meyer, Y., Rand, D. G., & Shalvi, S. (2019). Intuitive Honesty Versus Dishonesty : Meta-Analytic Evidence. *Perspectives on Psychological Science*, *14*, 778–796. https://doi.org/10.1177/1745691619851778

Rosenthal, R., Persinger, G. W., Vikan-Kline, L. L., & Fode, K. L. (1963). The effect of experimenter outcome-bias and subject set on awareness in verbal conditioning experiments. *Journal of Verbal Learning and Verbal Behavior*, *2*, 275–283. https://doi.org/10.1016/S0022-5371(63)80094-8

Shalvi, S., Eldar, O., & Bereby-Meyer, Y. (2012). Honesty Requires Time (and Lack of Justifications). *Psychological Science*, *23*(10), 1264–1270. https://doi.org/10.1177/0956797612443835

Shalvi, S., Eldar, O., & Bereby-Meyer, Y. (2013). Honesty requires time—a reply to Foerster et al. (2013). *Frontiers in Psychology*, *4*(September), 1–2. https://doi.org/10.3389/fpsyg.2013.00634

Stefan, A. M., Gronau, Q. F., Schönbrodt, F. D., & Wagenmakers, E. J. (2019). A tutorial on Bayes Factor Design Analysis using an informed prior. *Behavior Research Methods*. https://doi.org/10.3758/s13428-018-01189-8

Suchotzki, K., Verschuere, B., Van Bockstaele, B., Ben-Shakhar, G., & Crombez, G. (2017). Lying takes time: A meta-analysis on reaction time measures of deception.

Psychological Bulletin, *143*(4), 428–453. https://doi.org/10.1037/bul0000087

Welsh, D. T., & Ordóñez, L. D. (2014). The dark side of consecutive high performance goals: Linking goal setting, depletion, and unethical behavior. *Organizational Behavior and Human Decision Processes*, *123*, 79–89. https://doi.org/10.1016/j.obhdp.2013.07.006

Wicherts, J. (2018). Why replications fail. In *Reflections on replication: Psychology's current crisis*. Utrecht, The Netherlands.

<div align="center">**Exploratory analyses of PDR1**</div>

Exploratory analysis of the data without exclusions ($N = 461$)

When no participants were excluded, the time pressure manipulation was successful; participants in the time-pressure condition rolled and reported the outcome of their dice roll faster ($M = 5.92$ s, $SD = 3.18$ s) than those in the self-paced condition ($M = 9.46$ s, $SD = 7.68$ s), $t(459) = 6.47$, $p < .001$, $d = 0.60$, 95% CI [0.42, 0.79]. Participants in the time-pressure condition ($M = 3.67$, $SD = 1.73$) did not report higher dice roll outcomes than participants in the self-paced condition ($M = 3.76$, $SD = 1.70$), $Z = -0.53$, $p = .701$, $rrb = 0.03$. There was no evidence for lying in the time-pressure condition, $X^2(5, N = 231) = 2.84$, $p = .724$, $V = 0.11$, nor in the self-paced condition, $X^2(5, N = 230) = 9.06$, $p = .107$, $V = 0.20$.

Exploratory analysis of the subsample of participants who (strongly) believed that several participants would be randomly selected to receive a bonus ($N = 149$)

In this analysis we excluded the 33 participants in time pressure condition who responded too slow, and excluded an additional 279 participants who indicated either ''Strongly do not believe'', ''Do not believe'' or ''Neutral' on the 5-point Likert scale statement "Several students will receive a monetary reward for the dice under cup game".

Data from extreme outlier (91s; more than 5SDs from M) was excluded for the time pressure manipulation check. The time pressure manipulation in this subsample ($N=148$) was successful; participants in the time-pressure condition ($M = 4.90$ s, $SD = 1.35$ s) rolled and reported the outcome of their dice roll faster than those in the self-paced condition ($M = 8.21$ s, $SD = 4.60$ s), $t(146) = 5.97$, $p < .001$, $d = 0.98$, 95% CI [0.64, 1.33].

Further analyses showed that the participants in the time-pressure condition ($M = 3.55$,

SD = 1.67) did not report higher dice roll outcomes than those in the self-paced condition (*M*

= 3.73, *SD* = 1.69), *Z* = -0.65, *p* = .743, *rrb* = 0.05. There was no evidence for cheating in the

time-pressure condition, X²(5, *N* = 75) = 2.68, *p* = .749, *V* = 0.19, nor in the self-paced condition,

X²(5, *N* = 74) = 3.68, *p* = .597, *V* = 0.22.

Exploratory analyses of PDR2

No exclusions (N = 319)

Exploratory analysis of the data without exclusions (N = 319). When no participants were excluded, the manipulation was still successful; participants in the time-pressure condition rolled and reported the outcome of their dice roll faster (M = 5.93 s, SD = 2.41 s) than those in the self-paced condition (M = 7.88 s, SD = 4.62 s), t(317) = 4.74, p <.001, d = 0.53, 95% CI [0.31, 0.75]. Participants in the time-pressure condition (M = 3.68, SD = 1.66) did not report higher dice roll outcomes than participants in the self-paced condition (M = 3.76, SD = 1.63), Z = -0.43, p = .667, rrb = 0.02, 95% CI [-∞, 0.13]. There was no evidence for lying in the time-pressure condition, X^2(5, N = 160) = 2.68, p = .750, V = 0.13, nor in the self-paced condition, X^2(5, N = 159) = 8.96, p = .111, V = 0.24.

Subsample tested in native language (N = 177)

In the original study all participants were tested in their native language, therefore we did an exploratory analysis of a subsample with only participants who performed the task in their native language (N = 177) included. The time manipulation was successful; participants in the time-pressure condition rolled and reported the outcome of their dice roll faster (M = 5.19 s, SD = 1.28 s) than those in the self-paced condition (M = 7.86 s, SD = 4.65 s), t(175) = 4.97, p <.001, d = 0.75, 95% CI [0.44, 1.06]. Participants in the time-pressure condition (M = 3.86, SD = 1.65) did not report higher dice roll outcomes than participants in the self- paced condition (M = 3.68, SD = 1.70), Z = 0.70, p = .241, r_{bc} = -0.05, 95% CI [-∞, 0.09]. There was no evidence for lying in the time-pressure condition, X^2(5, N = 80) = 5.2, p = .392, V = 0.25, nor in the self-paced condition, X^2(5, N = 97) = 8.34, p = .139, V = 0.29.

Subsample tested in non-native language (N = 120)

In addition, we also did an exploratory analysis of a subsample with only participants who performed the task in their non-native language (N = 120) included. The time manipulation was successful; participants in the time-pressure condition rolled and reported the outcome of their dice roll faster (M = 5.33 s, SD = 1.69 s) than those in the self-paced condition (M = 7.93 s, SD = 4.61 s), t(118) = 4.05, p <.001, d = 0.74, 95% CI [0.37, 1.11]. Participants in the time-pressure condition (M = 3.41, SD = 1.68) did not report higher dice roll outcomes than

participants in the self-paced condition ($M = 3.89$, $SD = 1.52$), $Z = -1.56$, $p = .940$, $r_{bc} = 0.15$, 95% CI [$-\infty$, 0.32]. There was no evidence for lying in the time-pressure condition, $X^2(5, N = 58) = 0.97$, $p = .965$, $V = 0.13$, nor in the self-paced condition, $X^2(5, N = 62) = 8.65$, $p = .124$, $V = 0.37$.

Does session size influence the perceived chance of bonus payment?

As the group size of each experimental session of PDR1 differed from the group size of each experimental session of the original study (original study: 3-6 participants per session, first replication: 225-275 participants per session), we wanted to examine whether this could influence how the vague instruction "*several participants will be randomly selected to receive a monetary reward*" would be interpreted. This was explored in a vignette study where participants were asked to imagine taking part in a dice rolling study with either 3-6 others, 225-275 others or no others and to indicate what they thought was the chance that they would receive the monetary reward.

Method

Participants. One hundred and sixty-seven participants were recruited through the online platform Prolific.ac. They earned £1 for their participation. Participants who did not answer correctly on either one of the two control questions ("How many other participants were there in the dice roll session in which you took part? " and ''Answer 'three' on this question.'') were excluded from all analysis ($N = 23$). So, we analyzed the data of 144 participant of which 52 in the no-others condition, 47 in the 3-6 others condition and 45 in the 225-275 others condition.

Procedure. Participants were asked to imagine participating in a dice rolling study with either 3-6 others, 225-275 others or no others. All participants received the same instructions (i.e. "At the end of the experiment, the experimenters will randomly select several participants and pay them according to the number that appeared on their dice") and had to indicate what they thought was the chance that they would receive the monetary reward.

Results

The was no significant difference between the no others condition ($M = 65.72\%$, $SD = 34.47\%$), the 3-6 others condition ($M = 58.29\%$, $SD = 32.90\%$) and the 225-275 others condition ($M = 53.88\%$, $SD = 41.47\%$), $F(2,141) = 1.30$, $p = .28$, $d = 0.22$ (95% CI: [-0.190;0.588]) in how they answered the question "What is the chance that you will receive the reward?"

Discussion

This pilot study explored whether the vague reward instructions from the original paper could be differently interpret depending on group session size. We found no evidence that group session size of each experimental influences how participants perceived the chance that they would get the reward.