

Does Honesty Require Time? Two Preregistered Direct Replications of Experiment 2 of Shalvi, Eldar, and Bereby-Meyer (2012)



Ine Van der Cruyssen¹, Jonathan D'hondt², Ewout Meijer³,
and Bruno Verschuere¹

¹Department of Clinical Psychology, University of Amsterdam; ²Department of Experimental Psychology, Ghent University; and ³Clinical Psychological Science Department, Maastricht University

Psychological Science
2020, Vol. 31(4) 460–467
© The Author(s) 2020



Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0956797620903716
www.psychologicalscience.org/PS



Abstract

Shalvi, Eldar, and Bereby-Meyer (2012) found across two studies ($N = 72$ for each) that time pressure increased cheating. These findings suggest that dishonesty comes naturally, whereas honesty requires overcoming the initial tendency to cheat. Although the study's results were statistically significant, a Bayesian reanalysis indicates that they had low evidential strength. In a direct replication attempt of Shalvi et al.'s Experiment 2, we found that time pressure did not increase cheating, $N = 428$, point biserial correlation (r_{pb}) = .05, Bayes factor $(BF)_{01} = 16.06$. One important deviation from the original procedure, however, was the use of mass testing. In a second direct replication with small groups of participants, we found that time pressure also did not increase cheating, $N = 297$, $r_{pb} = .03$, $BF_{01} = 9.59$. These findings indicate that the original study may have overestimated the true effect of time pressure on cheating and the generality of the effect beyond the original context.

Keywords

intuition, cheating, lying, honesty, replication, moral decision making, time pressure, open data, open materials, preregistered

Received 4/3/19; Revision accepted 12/6/19

Whether it is Lance Armstrong using doping, Diederik Stapel publishing fake data, the Enron executives overstating the company's earnings, or people understating their earnings on tax forms, cheating occurs in many aspects of life. 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 might wonder whether the tendency to serve one's best interest by cheating is automatic and whether honesty requires deliberation (Bereby-Meyer & Shalvi, 2015).

A straightforward way to examine how Kahneman's framework of thinking fast versus thinking slow affects moral decision making is to manipulate the time allotted to decide on whether to cheat or to 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 corresponding to higher pay. Because the payout was based entirely on self-report, participants had both the opportunity and the financial incentive to cheat and report a higher-than-actual outcome. Critically, participants had to roll and report their die-roll outcome either within 8 s or without any time limit. In both experiments, the average reported die-roll outcome was higher—that is, there was more

Corresponding Author:

Bruno Verschuere, University of Amsterdam, Department of Clinical Psychology, Nieuwe Achtergracht 129B, 1018 VZ Amsterdam, The Netherlands
E-mail: b.j.verschuere@uva.nl

Table 1. Self-Reported Die-Roll Outcomes in the Time-Pressure Condition and the Self-Paced Condition of Shalvi, Eldar, and Bereby-Meyer (2012) and in Our Replications

Study	Average reported die roll		Z from Mann-Whitney U test	<i>p</i>	Rank biserial correlation	Bayes factor (BF)
	Self-paced condition	Time-pressure condition				
Shalvi et al., Experiment 2	3.42 (1.84)	4.38 (1.50)	2.17	.030	−0.28 [−0.48, −0.05]	BF ₁₀ = 1.15
Preregistered Direct Replication 1 (mass testing)	3.76 (1.70)	3.61 (1.73)	−0.87	.808	0.05 [−∞, 0.14]	BF ₀₁ = 16.06
Preregistered Direct Replication 2 (small groups)	3.76 (1.63)	3.67 (1.67)	−0.44	.671	0.03 [−∞, 0.14]	BF ₀₁ = 9.59

Note: For average reported die rolls, values in parentheses are standard deviations; for rank biserial correlations, values in brackets are 95% confidence intervals. The null hypothesis (H0) is that there is no difference between conditions. For Shalvi et al.'s original study, the alternative hypothesis (H1) was that time pressure would affect cheating. For Preregistered Direct Replications 1 and 2, the one-sided alternative hypothesis (H1) was that time pressure would increase cheating. BF₁₀ expresses how much more likely the data are to occur under H1 than under H0. BF₀₁ expresses how much more likely the data are to occur under H0 than under H1. The Bayesian analyses were not preregistered.

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 giving people time to think rather than pushing 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>). This study was theory driven, relied on an established manipulation, and included manipulation checks; in addition, the materials and data for the study are publicly available, and the study has been frequently cited.

There are, however, also reasons to question whether time pressure would increase dishonesty. First, limited cognitive capacity—also assumed to trigger automatic tendencies—has 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 in which 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, Verschuere, Bereby-Meyer, Rand, & Shalvi, 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, time pressure was found to slightly increase cheating in a multiple-die-roll paradigm using a virtual die (D'hondt, Van der Cruyssen, Meijer, & Verschuere, 2019) but not in a single-die-roll paradigm (Van der Cruyssen, D'hondt, Meijer, & Verschuere, 2019).

The extent to which 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 (Bayes factor, or BF = 1.15; see Table 1), we set up an attempt to replicate Experiment 2 of Shalvi et al. (2012).

Reproducing the 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 $r_{rb} = -.22$, 95% confidence interval, or CI = [−.43, .01], corresponding to a Cohen's *d* of 0.44; Experiment 2: $r_{rb} = -.28$, 95% CI = [−.48, −.05], corresponding to a Cohen's *d* of 0.58). We also reproduced ancillary effects (see <https://osf.io/fjca2/>).

Preregistered Direct Replication (PDR) 1

PDR 1 was a preregistered replication of Experiment 2 reported by 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 PDR 1 and the original studies was the session size. Sessions in the original study consisted of up to 6 participants, but in PDR 1, 228 and 233 participants were used. 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 <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 we excluded participants who did not meet the 8-s deadline—the sample sizes of the time-pressure condition 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 randomly assigned to either the time-pressure or the self-paced condition. Such between-session randomization is undesirable, as the experimenter is no longer blind to condition and could influence the results (Rosenthal, Persinger, Vikan-Kline, & Fode, 1963). We therefore chose to randomly assign participants to the time-pressure condition and the self-paced condition 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: E-Prime; current study: Qualtrics), the test language (original study: Hebrew; current study: English), the country (original study: Israel; current study: The Netherlands), and whether participants were tested in their first language (original study) or in English, which for most participants was their second language (our study). We will return to these differences in the General Discussion.

Method

The design and analysis plans were preregistered on the Open Science Framework (<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 the 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. Camerer et al. (2018) showed that the effect size of replications is on average about 50% of the original effect size, so 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 366). For a one-sided independent Mann-Whitney U test with an alpha of .05, the minimum sample size is 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. In each session, students performed a battery of tasks of which our task was the first. Four hundred sixty-one first-year psychology students participated. Thirty-three participants in 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 ($SD = 2.58$): 198 participants in the time-pressure condition and 230 participants in the self-paced condition.

Procedure. Participants first gave informed consent. Each was then randomly assigned to the time-pressure condition (i.e., roll the die and report the outcome within 8 s) or the self-paced condition (i.e., roll the die and report the outcome at their own pace) using Qualtrics (2019) 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 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 Fig. 1). As a financial incentive for cheating, and in accordance with the original instructions, participants



Fig. 1. Illustration of the die-rolling procedure used in Preregistered Direct Replications 1 and 2.

were informed that several of them 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 them to roll the die and report their outcome. The instructions were delivered in English (see <https://osf.io/c2z4f/>).

To evaluate whether the participants believed that 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” and “My dice roll was fully anonymous-only I could know what I rolled.” They were also asked 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, we asked them to answer the multiple-choice question, “The ratio between the dice roll and the possible reward is . . .” by choosing among 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? Following 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, $\chi^2(5, N = 198) = 1.21$, $p = .944$, $V = .08$,¹ nor in the self-paced condition, $\chi^2(5, N = 230) = 9.06$, $p = .107$, $V = .20$.

Exploratory analyses.

Time-pressure manipulation check. Data from an extreme outlier (91 s; more than 5 *SDs* from the mean) were excluded from 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 that the time-pressure manipulation was successful.

Exclusions. Repeating the analyses without any exclusions, or using the subsample that expressed strong belief in the payment scheme (i.e., agreed or strongly agreed with 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 participants (90%) answered the control question regarding the payment scheme correctly. Most participants (84%) 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 of the participants (35%) reported a strong belief that several students would be paid for the die-roll game.

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).

PDR 2

To rule out the possibility that the difference in results between our first replication study and the original study was 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. We also explored the possibility that testing participants in their first language versus their second language would modulate the effect.

Method

The design and analysis plans for PDR 2 were preregistered on the Open Science Framework (<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 our intention 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 BF for the Bayesian Mann-Whitney test that assessed the differences between the time-pressure condition and the

self-paced condition on the self-reported die-roll outcome. We used a zero-centered Cauchy prior (r) scaled at 0.707 (the default setting in JASP; JASP Team, 2019) in all Bayesian analyses. If decisive evidence were reached for either the alternative hypothesis (i.e., that time pressure leads to a higher reported outcome compared with the self-paced condition; $BF_{10} > 10$) or the null hypothesis (i.e., that the time-pressure manipulation does not affect reported outcome; $BF_{01} > 10$), we would terminate data collection. After running 319 participants ($N = 297$ 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 for a die-rolling study. They received €2 for participation in the 10-min 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 inclusion 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 8-s 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 Dutch nationality (62%), and about half of the participants spoke Dutch as their native language (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 and then took a seat at one of the six individual tables.³ On each table was a laptop and a cup with a lid on it. General oral instructions were given to the group. After obtaining informed consent, we gave all further instructions 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 the lid. Each participant was randomly assigned to either the time-pressure condition (8-s deadline) or the self-paced condition using Qualtrics permuted block randomization. As a financial incentive for cheating, and in accordance with the original instructions, participants were informed that several of them 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 30 s to prevent participants from going through the instructions without paying proper attention. After 30 s,

the next button appeared and pressing it started a timer measuring how long it took participants to roll their die and report their outcome (see Fig. 1). Participants could choose to take the task in English or in Dutch (for the materials, see <https://osf.io/6c9qr/>).

After reporting their die-roll outcome, participants were asked to provide their gender, major, age, nationality, and native language. To gain insight into 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 1, *strongly disagree*, to 5, *strongly agree*. To evaluate whether the participants had read the instructions attentively, we asked them to answer the multiple-choice question, “The ratio between the dice roll and the possible extra reward is equal to / two times / half of / four times . . . the outcome of the die roll.” They were also asked, “What was your perceived time-pressure during the die roll?” Responses were made on a 5-point scale ranging from *very high* to *very low*.

Deviations from the original study. Except for session size, the deviations between the current study and the original study were 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 that the time-pressure manipulation was successful.

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, $\chi^2(5, N = 138) = 1.91$, $p = .861$, $V = .12$, or in the self-paced condition, $\chi^2(5, N = 159) = 8.96$, $p = .111$, $V = .24$.

Exploratory analyses.

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

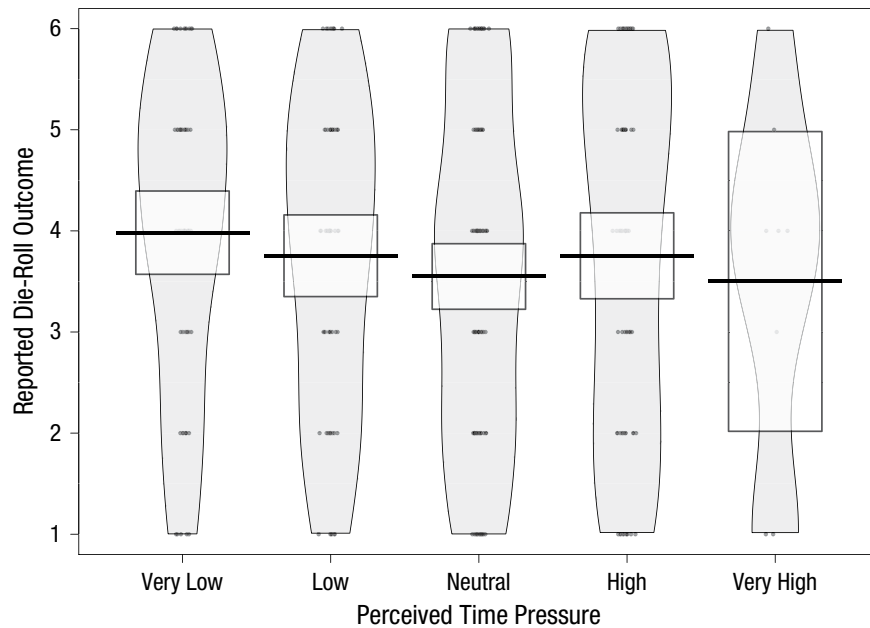


Fig. 2. Pirate plot representing the reported die-roll outcome at different levels of perceived time pressure within the time-pressure condition of Preregistered Direct Replication 2. The bold lines represent the sample mean (with the upper and lower bounds of the rectangle representing the 95% confidence interval around the mean). The dots represent individual raw data points. The width of the gray areas indicates the density of the data.

Test language. Using a similar die-rolling paradigm, Bereby-Meyer et al. (2018; but see Köbis et al., 2019) found that participants cheated more when the experiment was conducted in their native language. Given that the original study also tested participants in their native language, we separately analyzed the subsample tested in their native language ($n = 177$). Participants in the time-pressure condition ($n = 80$; $M = 3.86$, $SD = 1.65$) did not report significantly higher die-roll outcomes than participants in the self-paced condition ($n = 97$; $M = 3.68$, $SD = 1.70$), $Z = 0.70$, $p = .241$, rank biserial correlation (r_{bc}) = $-.05$, 95% CI = $[-\infty, .09]$. (See <https://osf.io/c2n6h/> for the full results.)

Self-report ratings. Most participants (84%) answered the control question regarding the payment scheme correctly. Most participants (89%) also 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 indicated that the die-roll outcome for each of the five levels of perceived time pressure (*very low*, *low*, *neutral*, *high*, *very high*) was not significant, $\chi^2(4, N = 138) = 3.16$, $p = .532$ (see Fig. 2).

General Discussion

What is people's automatic tendency in a tempting situation? Shalvi et al. (2012) found that time pressure, a straightforward manipulation to spark "thinking fast" over "thinking slow," provoked more cheating, and they concluded that people's initial response is to serve their self-interest and cheat. We found no evidence that time pressure increased cheating in the die-roll paradigm. There are three possible reasons why replication studies do not produce the same results as the original study: (a) methodological problems in the replication study, (b) overestimation of the true effect size in the original study, or (c) differences between the studies that moderate the effect (Wicherts, 2018).

The first possibility is that methodological limitations in the replication study produced different results. 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. But even for participants who performed the task in their native language, there was anecdotal support for the absence of a time-pressure effect ($BF_{01} = 2.90$).

The second possible explanation is that the original study 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 (in this case, a 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, the most prominent being 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 the amount of cheating in the die-under-the-cup game (Gächter & Schulz, 2016). Then again, the large meta-analysis by Abeler, Nosenzo, & Raymond (2019) 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 PDRs, people were predominantly honest, and we in fact found no evidence of cheating.⁴ Whereas Shalvi et al. (2012) originally reasoned that “time pressure evokes lying even in settings in which people typically refrain from lying” (p. 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 by Shalvi et al. (2012) may have overestimated the true effect of time pressure on cheating 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 people’s intuitive tendency is to cheat and fits better with a preference for honest behavior.

Transparency

Action Editor: D. Stephen Lindsay

Editor: D. Stephen Lindsay

Author Contributions

I. Van der Cruyssen and J. D’hondt share first authorship of this article. B. Verschuere developed the study concept. I. Van der Cruyssen and J. D’hondt designed Preregistered Direct Replication 1, and B. Verschuere and E. Meijer provided

critical input. B. Verschuere and E. Meijer designed Preregistered Direct Replication 2. Data for Preregistered Direct Replication 1 were collected by I. Van der Cruyssen and J. D’hondt. I. Van der Cruyssen and J. D’hondt analyzed the data with critical input from B. Verschuere and E. Meijer. All authors contributed to and approved the final version of the manuscript for submission.

Declaration of Conflicting Interests


B. Verschuere has past and ongoing collaborations with S. Shalvi and Y. Bereby-Meyer, the first and last authors of the article addressed in the present work. The authors declared that there were no other conflicts of interest with respect to the authorship or the publication of this article.

Open Practices

The design and analysis plans for Preregistered Direct Replications (PDRs) 1 and 2 were preregistered on the Open Science Framework at <https://osf.io/jez3g> and <https://osf.io/9bg3z>, respectively. Changes to the preregistration are noted in the text. All materials, data, and analysis scripts have been made publicly available via the Open Science Framework (PDR 1: <https://osf.io/fnh9u/>; PDR 2: <https://osf.io/xwzpc/>). The complete Open Practices Disclosure for this article can be found at <http://journals.sagepub.com/doi/suppl/10.1177/0956797620903716>. This article has received the badges for Open Data, Open Materials, and Preregistration. More information about the Open Practices badges can be found at <http://www.psychologicalscience.org/publications/badges>.



ORCID iD

Bruno Verschuere  <https://orcid.org/0000-0002-6161-4415>

Acknowledgments

We thank Sam Fischer, Sanne Visser, and Gijs Dollekens for collecting the data of Preregistered Direct Replication 2.

Notes

1. Cramer’s V varies from 0 to 1, expressing the strength of the association between two variables.
2. This was the result of the first calculation. Because of the bootstrapping approach, there is variance in the estimation of the BF in JASP Version 0.9 (JASP Team, 2018). To illustrate, with 10 runs, we found the BF estimation to vary between 9.51 and 10.27. JASP 0.10 (JASP Team, 2019), which was released after we ended data collection, has enhanced stability (the BF varied only at the second decimal), and we therefore relied on its estimate for the BF’s. For PDR 2, using JASP 0.10, we found a BF_{01} of 9.59.
3. At both test sites, we made sure that participants could not see what other participants rolled or reported. The test location at Maastricht University had screens between tables. At the University of Amsterdam, participants were seated a sufficient distance from each other, all facing a wall.

4. This is far from exceptional, and many studies have found even lower average reports and complete honesty (see Fig. 1 of Abeler et al., 2019 or <http://www.preferencesfortruth-telling.com/>). The finding that people could cheat to maximize personal gain without any punishment but did not fits with the meta-analytic conclusion of Abeler et al. that people cheat surprisingly little.

References

- Abeler, J., Nosenzo, D., & Raymond, C. (2019). Preferences for truth-telling. *Econometrica*, 87, 1115–1153. doi: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*, 1, 318–320. doi:10.1177/2515245918769062
- Bereby-Meyer, Y., Hayakawa, S., Shalvi, S., Corey, J. D., Costa, A., & Keysar, B. (2018). Honesty speaks a second language. *Topics in Cognitive Science*. Advance online publication. doi:10.1111/tops.12360
- Bereby-Meyer, Y., & Shalvi, S. (2015). Deliberate honesty. *Current Opinion in Psychology*, 6, 195–198. doi: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. doi: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. doi: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, 93–99. doi:10.1016/j.archger.2007.01.047
- D'hondt, J., Van der Cruyssen, I., Meijer, E., & Verschuere, B. (2019). Does honesty require time? A preregistered online replication of SEB2012 Exp1. Retrieved from <https://osf.io/j7pbc/>
- Foerster, A., Pfister, R., Schmidts, C., Dignath, D., & Kunde, W. (2013). Honesty saves time (and justifications). *Frontiers in Psychology*, 4, Article 473. doi:10.3389/fpsyg.2013.00473
- Gächter, S., & Schulz, J. F. (2016). Intrinsic honesty and the prevalence of rule violations across societies. *Nature*, 531, 496–499. doi:10.1038/nature17160
- JASP Team. (2018). JASP (Version 0.9) [Computer software]. Retrieved from <https://jasp-stats.org/>
- JASP Team. (2019). JASP (Version 0.10.2) [Computer software]. Retrieved from <https://jasp-stats.org/>
- Kahneman, D. (2011). *Thinking, fast and slow*. New York, NY: 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. doi:10.1177/1745691619851778
- Qualtrics. (2019). Qualtrics XM Platform [Computer software]. Retrieved from <https://www.qualtrics.com/platform/>
- 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. doi: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, 1264–1270. doi: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, Article 473. doi: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*, 51, 1042–1058. doi: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, 428–453. doi:10.1037/bul0000087
- Van der Cruyssen, I., D'hondt, J., Meijer, E., & Verschuere, B. (2019). Does honesty require time? A preregistered online replication of SEB2012 Exp2. Retrieved from <https://osf.io/6fuhm/>
- Wicherts, J. (2018). *Why replications fail*. Paper presented at the workshop “Reflections on Replication: Psychology's Current Crisis.” Utrecht University, The Netherlands.