
Jeroen Wouda, Gijsbert Bijlstra, Willem E. Frankenhujs, and Daniel H. J. Wigboldus
Radboud University, Behavioural Science Institute

Words: 937

Send correspondence to:
Jeroen Wouda MSc.
Radboud University
Behavioural Science Institute
PO Box 9104
6500 HE Nijmegen, The Netherlands
Phone:
E-mail: jeroenwouda@gmail.com

People are social animals; many of our good and bad behaviors take place in groups. A recent study by Weisel and Shalvi shows that “collaborative settings led people to engage in excessive dishonest behavior” (2015, p. 10655). The effects are large, spurring concern about harmful real-life consequences. Here, we report two preregistered studies that replicate the original findings, but with a smaller effect size. Moreover, our findings suggest that context moderates corruption in collaboration.

Weisel and Shalvi (2015) examined corrupt collaboration using a novel sequential dyadic die-rolling paradigm. In their Aligned outcomes condition, player A privately rolls a die and reports the outcome to Player B (anonymously). Next, Player B privately rolls a die and reports the outcome to player A (anonymously). If both players reported the same number, they earned money; otherwise, they earned nothing. This interaction was repeated for 20 trials. The number of reported doubles was the dependent variable. Participants reported a double on 81.5% of trials. This is a staggering 489% more than the chance expectation of 16.7%, and vastly more than the 54.9% doubles that lone players throwing twice report.

Weisel and Shalvi (2015) tested students used to participating in economic studies. In Study 1, we conducted a preregistered replication study\(^1\) of their Aligned outcomes condition to test whether their effect generalizes to students used to participating in psychological but not economic studies (see Simons, Shoda, & Lindsay, in press). Our results\(^2\) are consistent with those of Weisel and Shalvi: participants reported a higher percentage of doubles (29.6%) than expected by chance (16.7%; generalized linear mixed model (GLMM): \(\chi^2(1) = 10.63, p < .002\); see Appendix A for details). However, our results indicate a lower rate of corruption,

\(^1\)www.osf.io/gh5pd
\(^2\)All data was processed and analyzed in RStudio (RStudio, 2012), which is an integrated development environment for R (R Core Team, 2015). Analyses were run with either MLwiN (Rasbash, Charlton, Browne, Healy, & Cameron, 2009) and / or the lme4 package (Bates, Mächler, Bolker, & Walker, 2014).
with participants reporting fewer doubles than found by Weisel and Shalvi (GLMM: \( \chi^2(1) = 31.01, p < .001; \) Table 1).  

There are multiple, mutually compatible explanations for the observed difference in effect sizes. Research shows that published effect sizes tend to overestimate true effect sizes, and such overestimation tends to be greater in pioneering studies that are the first to report an effect, a ‘decline effect’ (Anderson, Kelley, & Maxwell, 2017; Ioannidis, 2008; Simonsohn, 2015). It is possible that the effect sizes observed by Weisel and Shalvi (2015) overestimated the true effect. Further, contextual factors may have affected the difference in effect sizes, and behavioral norms in particular (e.g., Grube, Morgan, & McGree, 1986; Nucifora, Gallois, & Kashima, 1993). Some research suggests that the norm for students used to participating in economic studies is to maximize payoffs, more so than for students used to participating in psychological studies (Cappelen, Nygaard, Sørensen, & Tungodden, 2015; Carter & Irons, 1991; Gerlach, 2017). We do not compare the behavioral norms of these groups. Rather, we directly examine whether causally manipulating behavioral norms affects corruption in collaboration.  

To this end, we included norms as a moderator in Study 2. We manipulated the norm by showing participants a visual representation of the findings of the two previous studies (for a similar manipulation, see Kroher & Wolbring, 2015; Rauhut, 2013). Participants were either shown a representation of a distribution of results in which participants lied very often (High behavioral norm), i.e., Weisel and Shalvi’s (2015) data, or less often (Low Behavioral norm), i.e., our Study 1 data (see Appendix B for details). We tested a sample similar to Weisel and Shalvi’s: students who are used to participating in economic studies (recruited with ORSEE; Greiser, 2015). The results showed that participants in the High behavioral norm condition

---

3 For Studies 1 and 2, we also investigated whether the dishonesty of player A would influence the dishonesty of player B. These results are discussed in Appendix C.
reported more doubles ($M = 67\%$, $SD = 31\%$) than participants in the Low behavioral norm ($M = 47\%$, $SD = 30\%$; see Table 1).

### Table 1.

*Median, mean, standard deviation and percentages across studies over 20 trials.*

<table>
<thead>
<tr>
<th>Study</th>
<th>Median</th>
<th>Mean (SD)</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study 1&lt;sup&gt;a&lt;/sup&gt; ($N = 46$)</td>
<td>6.0</td>
<td>5.9 (3.8)</td>
<td>30%</td>
</tr>
<tr>
<td>Study 2 Low behavioral norm&lt;sup&gt;b&lt;/sup&gt; ($N = 42$)</td>
<td>8.0</td>
<td>9.3 (5.9)</td>
<td>47%</td>
</tr>
<tr>
<td>Study 2 High behavioral norm&lt;sup&gt;b&lt;/sup&gt; ($N = 40$)</td>
<td>13.0</td>
<td>12.6 (6.2)</td>
<td>67%</td>
</tr>
<tr>
<td>Weisel and Shalvi (2015): Aligned-outcomes&lt;sup&gt;b&lt;/sup&gt; ($N = 40$)</td>
<td>19.5</td>
<td>16.3 (5.1)</td>
<td>82%</td>
</tr>
</tbody>
</table>

<sup>a</sup>Sample used to participating in psychological studies.

<sup>b</sup>Sample used to participating in economic studies.

Our studies have several strengths, such as being preregistered and replicating a reported large effect, which may have real-life consequences. However, our studies also have limitations. First, we did not include Wiesel and Shalvi’s (2015) Individuals condition in our studies. Hence, we did not replicate their core finding that collaboration increases cheating relative to solitary play. Second, we used a lower monetary compensation than Weisel and Shalvi, possibly reducing our participants’ motivation to lie. Although higher incentives do not necessarily increase the magnitude of lies (Fischbacher & Föllmi-Heusi, 2013; Mazar, Amir, & Ariely, 2008), future research may systematically examine the extent to which size of incentives influence the magnitude of dishonesty in collaborative settings.

Finally, our results converge with the idea that collaborative settings can lead to dishonest behavior. Corrupt collaboration can have significant real-life consequences, but the

---

<sup>4</sup> We used to two different models to analyze the data (see Appendix D for details of these models). The results were similar, both indicating a difference between the two conditions (model 1: $\chi^2(1) = 4.18, p = .04$, model 2: $\chi^2(1) = 3.04, p = .09$).
severity of these consequences is likely to depend on context. Previous research highlights the role of social norms and beliefs about such norms in the spreading of dishonest behavior (Keizer, Lindenberg, & Steg, 2008; Rauhut, 2013). Here, we provided evidence suggesting that norms can shape dishonest behavior in a collaborative setting. Investigating what norms increase or decrease dishonesty in real-life settings is a promising avenue for future research.

**Acknowledgements**

We thank Ori Weisel and Shaul Shalvi for their correspondence and sharing of study materials.

**Contributions**

Contributed to conception and design: JW, GB, WF, DW

Contributed to acquisition of data: JW

Contributed to analysis and interpretation of data: JW

Drafted and/or revised the article: JW, GB, WF, DW

Approved the submitted version for publication: JW, GB, WF, DW

**Data accessibility statement**

The data has been uploaded to the Open Science Framework (www.osf.io/gh5pd).

**Competing Interests**

The authors have no competing interests to declare.
References


Appendix A

Here we report a detailed description of the model in MLwiN we used to analyze the data of Study 1. First we checked the extra binomial variance, which was acceptable. Second, we defined the logit for the reference probability of 1/6 (i.e., the chance of reporting a double if subjects are honest), which was -1.609. Lastly, we examined whether the mean deviated from the reference (i.e., do people report more doubles than based on chance?). We used a two-level model, with trial as the lowest level and dyad as the highest level (i.e., trial is nested within dyad).
Appendix B

**Behavioral norm manipulation**

The materials and procedure in Study 2 were identical to Weisel and Shalvi (2015) and Study 1, with two exceptions. First, after participants indicated they understood the rules and believed the die was fair, they were told that they would receive information about the results from subjects who previously participated in a similar study. More specifically, we showed and explained participants in the Low behavioral norm condition the data-figure from Study 1 (see Figure B1). Participants in the High behavioral norm condition were presented with the data from the Aligned-outcomes condition from Weisel and Shalvi (2015; see figure B2). Thus, the participants received either a data pattern from an earlier study with a similar setting in which previous participants lied relatively often versus lied relatively less. Any questions from participants regarding the figures were answered individually. Second, we asked participants to estimate the percentage of the reported doubles in previous research in order to measure whether our manipulation was successful.

To test whether our manipulation was successful, we conducted an independent-samples *t*-test and compared participants’ estimates of the percentage of reported doubles in previous research for the low and the high behavioral norm condition. There was a significant difference in the percentages for the Low (*M* = 31%, *SD* = 24%) and the High (*M* = 67%, *SD* = 21%) behavioral norm condition; *t*(79.705) = 7.24, *p* < .001. These results suggest that participants understood the figures, indicating our manipulation was successful.
Figure B1. Observed distribution of reported outcomes in Study 1, which served as the Low behavioral norm manipulation. We explained in text to our participants how often people either reported a “double 5” (red square) or how often Player A reported a 4 and player B reported a 3 (yellow square).
Figure B2. Observed distribution of reported outcomes in Study 1, which served as the High behavioral norm manipulation. We explained in text to our participants how often people either reported a “double 5” (red square) or how often Player A reported a 4 and player B reported a 3 (yellow square).
Appendix C

We also tested whether the dishonesty of player A would influence that of Player B. In line with Weisel and Shalvi (2015), reporting a “6” 20 times made player A brazen, whereas reporting a double 20 times made player B brazen. Since none of the player A’s in Study 1 were brazen, we did not test this hypothesis for Study 1. We did, however, test this hypothesis for Study 2. When player A was brazen, 80% (4 of 5 cases) of player B’s were brazen. When A was not brazen, only 6% (2 of 36 cases) player B’s were brazen ($\chi^2(1) = 13.97, p < .001$).
Appendix D

Below is the R-code that we used to specify our models and obtain our $p$-values of Study 2.

```r
#set the working directory
setwd('path to working directory')

#read data
data1 = read.csv2("Data_Study_2_collabra.csv")

#load the library
library(afex)

# the variables contain the following values:
# DyadID: the unique id for each Dyad
# Treatment_name: low or high behavioral norm
# total_trials: always 20
# total_doubles: the total reported number of doubles for each dyad
# proportion_doubles: the variable “total_doubles” divided by “total_trials” (i.e., 20)
# First get model 1 with the mixed function. As the dependent variable we specified the
# proportion of doubles for each dyad, and Treatment_name (i.e., the behavioral condition)
# was the predictor. Furthermore, we wanted to obtain our $p$-values, hence we used the PB
# function, with 1000 simulations. Lastly, we also specified that the data was binomial.

model1 <- mixed(proportion_doubles ~ Treatment_name + (1 | DyadID), method = "PB",
                 family = binomial, data = data1, args.test = list(nsim = 1000))

#Now get the other model with the mixed function. The only difference with the first model is
#that instead of using the proportions as dependent variable, we used the cbind function. This
#function requires the hits (i.e., the number of doubles) and misses (i.e., the total number of
#trials minus the total number of doubles) as the dependent variable for each dyad. The rest is
#still the same.

model2 <- mixed(cbind(total_doubles, total_trials - total_doubles) ~ Treatment_name + (1 | DyadID), method = "PB", family = binomial, data = data1, args.test = list(nsim = 1000))

#get the p-values with the anova function
anova(model1)
anova(model2)

Lastly, below is Figure D1, which shows the output of the R-Code.
THE COLLABORATIVE ROOTS OF CORRUPTION?

Figure D1. Screenshot of the output of model 1 and model 2.