Collaborative cheating brings people closer:

How people evaluate and interact with partners in crime

Simon Tobias Karg[1], Minjae Kim[2], Panagiotis Mitkidis[1,3], Liane Young[2]

[1]Aarhus University

[2]Boston College

[3]Duke University

Corresponding author

Simon Tobias Karg, Department of Management, Aarhus University, Denmark.

Email: simonkarg@mgmt.au.dk

Abstract

People can collaborate for immoral reasons. We investigated how people are influenced by others' signaling a willingness to engage in collaborative cheating, and how people perceive such potential partners in crime. Participants (N = 308) performed a task in which they could either report their performance honestly, or cheat for financial gains. Each participant was assigned a Leader who could choose to check the veracity of the report, yet equally benefitted from participants' cheating. Manipulating Leader checking behavior revealed that lowering rates of checking increased participant cheating. Moreover, Leaders who checked less often were perceived as more moral, trustworthy, competent, and psychologically closer than Leaders who checked more often. This trustworthiness bonus translated to actual behavior: participants were more likely to invest in "trustworthy" Leaders in a subsequent trust game. We conclude that collaborative cheating shares the benefits of other forms of collaborations, bringing people closer together, and cultivating trust.

Keywords: *collaborative corruption, ethical decision-making, person perception, trust, moral psychology*

Statement of Relevance

Cases of organizational corruption have shown that collaborative corruption has detrimental consequences for society. For example, the collaborative cheating observed in the 2015 Volkswagen scandal imposed large costs upon consumers, the environment, and potential down-stream health care expenses. Although collaborative corruption occurs in many contexts, many important aspects of it remain understudied. The present work makes two contributions. First, we show that collaborative corruption creates strong bonds between collaborators, with people judging their partners in crime to be trustworthy, competent, and psychologically close. Second, we present a novel paradigm that allows the investigation of real-time cheating at the individual and group levels that can be translated to explore the psychology of leadership, personality, and morality. The present work lays the foundation for a psychological account of why and how people engage in collaborative corruption, which will ultimately inform safeguards against it.

Introduction

Collaboration is undoubtedly a key requirement of a flourishing society. Yet, collaboration can also serve immoral aims. One prominent example is the 2015 Volkswagen scandal, in which the company's top-management turned a blind eye to a fraudulent method devised to seemingly lower emissions (Goodman, 2015). Collaborative corruption, which we use to refer to situations where people cheat collaboratively, is a perilous phenomenon; it is both widespread and can lead to higher degrees of cheating compared to individual settings (Weisel & Shalvi, 2015). Despite recent empirical interest in the phenomenon of collaborative corruption (e.g. Gross, Leib, Offerman, & Shalvi, 2018; Shalvi, 2016; Soraperra et al., 2017), two important questions remain that we address in the current work. First, how are people influenced by others' signaling a willingness to engage in collaborative corruption? Second, how do people receiving such signals perceive their potential partners in crime?

Collaborative corruption presents a remarkable ambiguity for the people engaging in it. On the one hand, knowledge that someone is behaving dishonestly could translate to generally negative character evaluations (Rozin & Royzman, 2001). On the other hand, successful collaborative efforts create common ground and enhance in-group trust (Rousseau & McLean Parks, 1993; Zelmer, 2003). This tension between positive and negative consequences of collaborative cheating provides a rich background for the study of behavior and character evaluations. For instance, while people might evaluate their partners in crime as competent, they may not judge them to be moral (Bhattacharjee, Berman, & Reed, 2013). Similarly, although they may trust them in one context, they might not in another.

The present research examines cheating behavior and character evaluations in a group setting, in which short-term outcomes for the group are dependent solely on *reported* rather than *actual* performance. Since collaborators often also profit from their partners' performing well, they may be motivated to *not* check potential misreports. Our experimental design models this situation using a two-player rely-or-verify game coupled with a trust game. In the rely-or-verify game, a Player engages in a task and conveys information of their performance to their Leader, who then decides whether to verify (check) or rely on (not check) that information. Players are financially motivated to inflate their reports, and Leaders' outcomes are equally tied to this report. Finally, Players have the opportunity to invest in Leaders in the context of a trust game. In this study, all participants were assigned to the role of Player, and believed their Leader was a human partner, whereas in reality, Leader responses were preprogrammed.

Being checked and the impact on subsequent cheating

Rational choice theory predicts that higher chances of being caught will lead to lower rates of cheating, and empirical evidence (Thielmann & Hilbig, 2018) confirms this prediction. However, whereas the design of earlier studies highlights the probabilistic nature of being checked, our study highlights the role of the Leader in checking. Players may attribute intentions to the Leader, and anticipate their Leader's actions based on past behavioral history.

From this perspective, checking (as opposed to non-checking) serves as a signal: The Leader does not trust the report, and cares about honesty. Repeated checking implies that cheating will not be tolerated, and therefore not be profitable for the Player.

Repeated *non*-checking presents a more complex picture, since it may be interpreted as one of two conflicting signals. Non-checking can be interpreted as the Leader believing that the Player's report is correct. Alternatively, repeated non-checking by the Leader even when faced with "too good to be true" reports may be interpreted as a willingness to engage in collaborative cheating. Regardless, a Player should generally infer a Leader's tendency to not check from repeated non-checking, thereby increasing the Player's inclination to cheat.

Hypothesis 1: The more often a Player is checked, the less likely they will cheat in subsequent rounds.

The influence of being checked on character evaluations

Collaborative corruption presents an inherent ambiguity regarding how partners in crime should evaluate each other. Importantly, character evaluations are not a uniform construct but can be decomposed into several dimensions. For the current study, we focus on five dimensions of judgments: evaluations of morality, trustworthiness, competence, risk aversion, as well as the experienced psychological closeness. These measures allow us to explore potential patterns of diverging evaluations, such that people may for instance come to see their partners in crime as competent, but not as moral. In addition, tracking perceptions of risk aversion allows us to gauge whether participants believe Leader checking is due to a Leader being risk averse, or being concerned about honesty.

Based on prior work revealing that people are happy to freeride on their collaborators' dishonesty (Gross et al., 2018), we expect a generally negative impact of Leader checking on evaluations. We formed no specific hypotheses regarding individual character dimensions.

Hypothesis 2: Checking will lead to lower evaluations of the Leader than non-checking.

Updating evaluations over time

Updating character evaluations is a complex process, with the integration of new information with existing beliefs exhibiting a range of biases (Gershman, 2018; Reeder & Brewer, 1979). Most prominently, in what has been termed negativity bias (Rozin & Royzman, 2001), immoral or otherwise negative acts are usually more effective at reversing a positive character impression than vice versa.

So long as Hypothesis 2 is supported (i.e., checking will be perceived negatively), we expect that the direction of change in a Leader's checking behavior will have differential effects on final evaluations: If a Player interacts with a Leader who moves from checking to non-checking, this experience should result in more positive final evaluations. By contrast, if a Leader starts to repeatedly check results (after mostly not checking), this shift should be perceived as a strong break in an established collaboration, and lead to more negative final evaluations.

Hypothesis 3: A Leader shifting from checking to non-checking will be perceived more positively, as indexed by the final evaluation, than a Leader shifting from non-checking to checking.

Being checked and subsequent trust

Finally, we investigate how Leader evaluations translate to actual behavior in a trust game (Berg, Dickhaut, & Mccabe, 1995): Players have the opportunity to send money to their Leader, which will get tripled, and the Leader can choose to send money back or not. We expect a Player to evaluate their Leader as more trustworthy, and to invest more in their Leader, when they stop checking as opposed to start checking.

Hypothesis 4: Participants will send more money to a Leader shifting from checking to non-checking than to a Leader shifting from non-checking to checking.

Methods

This study was preregistered (see https://osf.io/afmrw [*Note: this original preregistration is author identified. For an anonymous version, see: https://osf.io/tvuwa/*]), and follows open science practices by sharing all material (including the source code used to run the study), data and code (see supplementary online materials on OSF https://osf.io/p2esr/?view_only=ef604a56ad9d4639ae5e670271b9d76c [*this link is anonymized for peer review*]). Unless otherwise noted, all hypotheses, measures, and statistical analyses reported here follow the preregistered procedure. Because the preregistered protocol was extensive, and given limitations of space in this article, we are not able to provide all analyses in the main text. Additional preregistered analyses are therefore provided in the online materials. The study was fully computerized, implemented in oTree (Chen, Schonger, & Wickens, 2016), and conducted over the internet.

Participants

Three-hundred and fifty participants were recruited from Amazon Mechanical Turk (mTurk). We recruited participants living in the US, and who either had at least a 95% acceptance rate, or completed between 25 and 100 HITs on mTurk. Following preregistered exclusion criteria, 42 participants were removed because they either failed the attention check, or used more than 3 attempts at the comprehension check (see procedure described below). This left 308 participants (mean age = 38.06, SD = 11.38, 138 female, 168 male, and 2 participants specifying other), with 155 in the High Check to Low Check condition, and 153 in the Low Check to High Check condition. This satisfied our preregistered sample size goal of 150 participants per condition. Because we employed a novel paradigm with unclear effect sizes, and we planned a variety of statistical tests (including repeated measures, as well as analysis of aggregated scores), a definitive power analysis was unfeasible. We chose N = 150 per cell, in order to allow testing for relatively small differences between our two conditions (achieving power of 80% for a t-test with $d \approx 0.32$).

Most participants (39.9%) had finished a BA or BS degree, and 88.3% had at least some college education. Participants reported an average of 16.14 ($SD = 11.64$) years of employment history. Moreover, 62.3% of the participants reported previous experience working in a supervisory role. All participants were payed \$0.50 for completing the study, plus 5 cents for each point they earned during the study (the average payment for this study was \$3.21, $SD =$ \$0.26).

Experimental task and procedure

The general structure of the experimental task follows two stages, the first consisting of a rely-or-verify game (adapted from Levine & Schweitzer, 2015), and the second consisting of a trust game (Berg et al., 1995). The rely-or-verify game in this study is a 20 round two player game, with two different roles (Player and Leader). Importantly, all participants in this study were assigned the role of Player, whereas Leaders were (unbeknownst to participants) played by a computer. The Player performs a die roll task similar to the cheating paradigm of Fischbacher and Föllmi-Heusi (2013), whereby the participant privately rolls a die, and reports the number they rolled. Players can report any number they want, with higher numbers leading to higher payoffs, incentivizing cheating. Because our version of this task is computerized, Players do not roll physical dice, but rather click a button which triggers a video of a die being rolled (Kocher, Schudy, & Spantig, 2018). The die rolls are pseudo randomized, such that for each half of the game (10 rounds) participants roll exactly one 1, two 2s, three 3s, two 4s, one 5, and one 6. The order of these die rolls is randomized within each game half. We chose this procedure to navigate two trade-offs: Firstly, the die results are slightly skewed to increase cheating (as lower die rolls are more likely to be inflated), yet without making participants skeptical about die rolls being fixed. Secondly, fixing die roll outcomes across conditions reduces noise, and facilitates better comparison between the conditions, without making results dependent on a specific sequence of die rolls.

After rolling the die, the Player is able to report any number (i.e. either the number they actually rolled, or an inflated number) to the Leader, who is described to them as the Leader of the dyad, who is responsible for checking results. The Leader then decides whether to rely on or verify the Player's report. Finally, depending on the outcome of the die roll, and the Leader's decision, payoffs are calculated. Four different scenarios are possible (see Figure 1). If the Leader checks and the report is correct, both the Player and the Leader receive half of the points the die shows (e.g. if the Player rolled and reported a 4, both the Player and the Leader will receive 2 points this round). In case the Leader checks and the report is incorrect, the Player receives 0 points, whereas the Leader receives half the points reported (e.g. if the Player reports a 5, but actually rolled a 1, the Player would receive 0 points, but the Leader would receive 2.5 points). In case the Leader does not check, and the Player's report was correct, both agents receive the number of points reported. Finally, if the Leader does not check an inflated report, both agents still receive the number of points reported. However, in this case, there is a 0.3% chance that the team will lose the game, meaning that they will lose all earnings they had made so far, and the game ends. This method was implemented to simulate similar low probability - high severity risk situations in the real world. The precise probability was chosen to be low enough such that cheating was clearly the most profitable strategy while presenting a credible threat of losing the game (undetected cheating in each of the 20 rounds accumulates to a 5.83% chance of losing). Importantly, even though participants are informed that the chance of losing the game if cheating goes undetected is 0.3%, we set the actual probability of losing the game to 0. This approach was taken to assure that participants did not actually lose the game, since studying the effects of losing the game was not the main purpose of this study.

[Fig 1 about here]

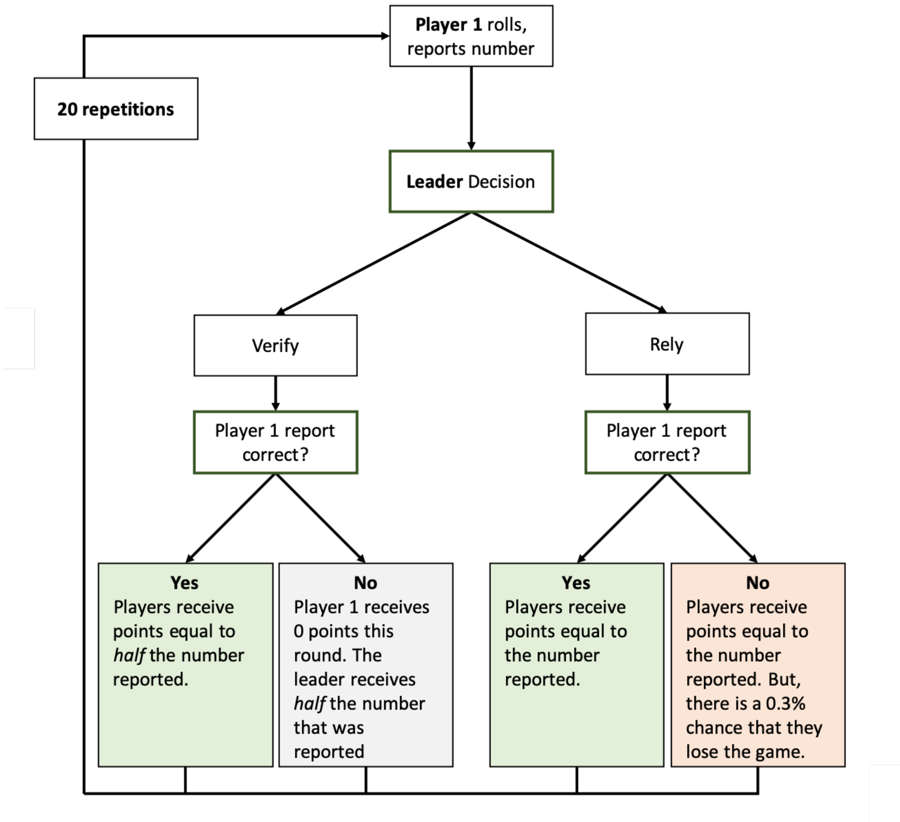

Figure 1: Structure of the rely-or-verify game as presented to the study participants.

In addition to the die roll task, each Player evaluates their Leader every 5 rounds (with one baseline evaluation before the game, and ratings at rounds 5, 10, 15, and 20), see also measures below.

Before participants are allowed to start the game, they answer six comprehension check questions, which ensure they understand the basic payoff scheme, and that they are always playing with the same Leader. All instructions presented to the participants can be found in the supplementary online material.

Leader behavior is manipulated in two conditions. In the High Check to Low Check condition, the Leader checks Player reports with a 90% chance in the first 10 rounds. Then, in the next 10 rounds, the Leader checks only with a 10% chance. This pattern of checking is reversed in the Low Check to High Check condition, such that the Leader checks with only 10% chance in the first 10 rounds, and with 90% chance in the last 10 rounds. In both conditions, the Leader never checks reports of a 2 or a 1, so as to not raise suspicion in participants that they are playing a computer.

After completing 20 rounds, participants are asked to assess the amount of power they experienced over the outcome, and how much power they experienced the Leader had, both on a Likert scale ranging from 1-7. This measure serves as a manipulation check, as we expect participants to perceive themselves as less powerful than their Leader.

Subsequently, Players play a one-shot trust game (Berg et al., 1995) with their Leader. Each participant acts as the sender, thus allowing them to send any amount of points they have earned in the previous game to their former Leader, which will be tripled. Players are instructed that their former Leader may choose to send some points back but is under no obligation to do so.

After deciding how much money to send, participants fill out personality measures (Moral Foundations subscales for Fairness, Authority as well as Ingroup/Loyalty (Graham et al., 2011), the DOSPERT financial risk-taking subscale (Blais & Weber, 2006), and demographics). In this section, we also included one attention check item ("Please select slightly disagree").

Subsequently, participants learn about the outcome of the trust game. In this study, the Leader never chooses to send any money back. Participants were then presented with a final Leader evaluation questionnaire using the same questions as in the rely-or-verify game, allowing us to measure updating Leader of evaluations in a new context, and after a severe violation of trust. However, in order to not punish participants for sending money to the Leader, we re-instantiate all points sent to the Leader.

Finally, participants go through a funnel debrief, including questions probing participants for their perception of the purpose of the study, whether they noticed any change in the Leader's behavior, and whether they believed they had played with a real person. The experiment lasted around 22 minutes.

Measures

This study comprises several dependent and independent variables. To analyze cheating, we created a binary measure of whether a participant inflated their rolled outcome in a given round. Unlike other studies employing die roll tasks, where researchers need to infer cheating from overall dice roll performance (Fischbacher & Föllmi-Heusi, 2013), the computerization of the experiment allowed us to accurately monitor whether, when, and how much participants actually cheated. Additionally, we also calculated mean cheating scores for each evaluation period, i.e. the 5 rounds of the game prior to an evaluation. Thus, for each Player, we calculated mean cheating in rounds 1-5, 6-10, 11-15, and 16-20. This way, we could investigate the influence of cheating in an evaluation period on subsequent evaluations.

Leader evaluations are made on a set of five character traits. First, morality evaluations ("how moral is your leader as a person?") are used to track perceptions of moral character (Goodwin, Piazza, & Rozin, 2014). Second, even though trustworthiness is usually subsumed under moral character evaluations (Lapsley & Lasky, 2001), we give it special focus here, given the above mentioned possibility of diverging evaluations ("how trustworthy is your leader as a person?"). Third, competence assessments ("how competent is your leader as a person?") measure perceived skill, intelligence, or talent of the Leader (Fiske, Cuddy, & Glick, 2007). Fourth, measuring perceived risk aversion ("how risk averse is your leader as a person?") allows the assessment of whether a Player believes that checking is due to the Leader being risk-averse rather than concerned about honesty. Finally, closeness ("how close do you feel towards your leader?") is not a character evaluation per se, but is used as an indicator of the collaborative spirit of the team. All ratings are made on a seven-point Likert scale, ranging from -3 to 3, with the endpoints labelled "Not at all", and "Extremely".

Regarding the trust game, the amount of money participants sent is measured as the percentage of the maximum amount a participant could have sent. This is done because the amount participants could send differs depending on their performance in the rely-or-verify game.

For the personality measures, we created z-scores of the sum score for each moral foundation (Fairness (std. alpha = .50), Ingroup/Loyalty (std. alpha = .57), as well as Authority (std. alpha = .72)). Since these reliability indices are below what is commonly considered adequate, we also adopted an alternative analysis approach by conducting an exploratory factor analysis (EFA) on the questionnaire data, which revealed a three factor structure that was slightly different than the one proposed by Graham et al. (2011). Because we preregistered an analysis using sum scores using the original structure, we report this originally planned analysis in the main text. In addition, we report extra analysis using factor scores from the EFA in the supplementary online materials.

Regarding the DOSPERT financial risk taking subscale, we had preregistered a parallel analysis, due to recent discussions about the DOSPERT's factor structure (Highhouse, Nye, Zhang, & Rada, 2017). This analysis indicated a two-factor solution. A subsequent EFA revealed two factors, one related to betting (items 1,3, and 5 on the financial risk-taking scale, std. alpha = 0.90), and the other related to investment (items 2,4, and 6, std. alpha = .81), for which we calculated sum scores.

Results

All data analyses were performed using R (R Core Team, 2014). Mixed effects models were built using the lme4 package (Bates, Mächler, Bolker, & Walker, 2015), as well as lmerTest (Kuznetsova, Brockhoff, & Christensen, 2017). Unless otherwise noted, all mixed effects models specify random intercepts on the participant level, and random slopes for round number. If models failed to converge, we simplified the random effect structure by removing random slopes. This follows common procedure (Bates et al., 2015), and is line with the preregistered protocol. Bayesian models were fit using brms (Bürkner, 2017).

Manipulation checks and descriptive measures

First, we tested and confirmed that Players perceived themselves to have less power than the Leader in determining the outcome of the rely-or-verify game (paired t-test $t(307) = 12.89$, $p < .001$, $d = 0.73$). Moreover, participants also perceived themselves to be on a relatively lower level on the social ladder (paired t-test $t(307) = -6.39$, $p < .001$, $d = -0.36$). Thus, Players did perceive their Leader as generally more powerful than themselves.

Turning to a descriptive overview of Leader checking and Player cheating, we find that, overall, participants cheated in 11.67 % (SD = 32.10 %) of the rounds in which they could cheat (i.e. rounds in which they did not roll a 6). Notably, only 151 participants cheated at least once during the study. We therefore performed robustness analyses for all reported analyses using only the subset of cheating participants. In almost all cases, model parameter estimates of models using the full dataset (as preregistered) or only cheaters were comparable (see Table S2, Figure S1, and online supplementary materials). Cheating was considerably more likely when participants rolled lower numbers (e.g. when rolling a 1, 25.32% (SD = 43.52%) of participants cheated, yet when rolling a 4, only 5.60 % (SD = 23.00%) cheated).

Leader checking, in turn, occurred on 37.22% (SD = 48.34%) of rounds. Overall, Leaders discovered 40.83% (SD = 49.19%) of all cheats.

How does checking influence cheating?

Next, we analyzed the influence of Leader behavior on Player cheating. To recall, we hypothesized that checking would be negatively related to subsequent Player cheating. To gauge this effect, we built a generalized linear mixed effects model specifying an interaction between condition and the cubic polynomial of round number, controlling for the number participants had actually rolled, as well as participants' scores on moral foundations and financial risk aversion. We chose a cubic polynomial for round number since we expected Players' cheating to develop nonlinearly as they develop cheating strategies, and adapt to Leader checking behavior, which switches after round 10. Yet, because polynomials are criticized for being too strict (Magee, 1998), we also modelled all nonlinear models with b-splines as preregistered robustness checks, which are less biased to non-local patterns but are more difficult to interpret. In general, the b-spline model results follow the polynomial models (model results can be found in the supplemental online material).

In line with our expectations, this linear mixed effects model showed significant interactions for round number and condition, such that participants in the Low Check to High Check condition cheated more in the first half of the game than participants in the High Check to Low Check condition. This pattern reversed following the switch in Leader behavior (see Figure 2). Thus, Hypothesis 1 finds support: the more a Leader checked, the less participants cheated. Interestingly, there was no long-term impact of being checked: when a Leader stopped checking, cheating increased quickly to degrees similar to when Leaders started out not checking. Thus, there was no overall difference in the amount of cheating between conditions (Wilcox W = 11568, p = .69, d = 0.064).

[Figure 2 about here]

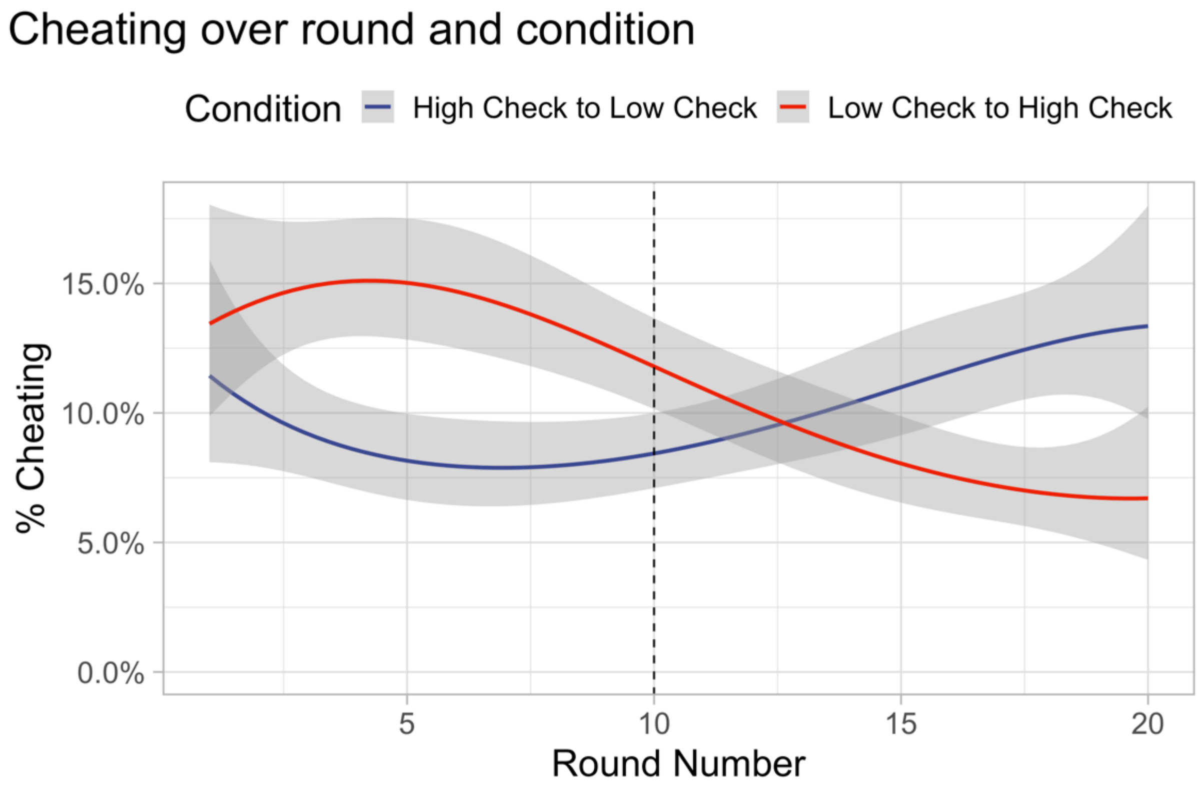

Figure 2: Likelihood of cheating by round and condition. Lines represent b-spline curves modelling Player cheating occurring during the game, with 3 degrees of freedom. The dotted line at round 10 indicates the point when Leader checking behavior switched. Grey ribbons indicate standard error of the mean.

In addition to the hypothesized interaction of round number and condition, the actual number a Player rolled was significantly related to Player cheating, such that higher numbers were less likely to be inflated (ß = -1.03, 95% CI = [-1.14, -0.92], p < .001). Regarding personality, Moral Foundation scores did not significantly predict cheating. Yet, participants reporting greater likelihood to engage in betting behavior were also more likely to cheat (ß = 0.54, 95% CI = [0.18, 0.91], p < .001). Full model results can be found in Table 1.

How does checking influence Leader evaluations?

To analyze Hypothesis 2, that checking has adverse effects on Leader evaluations, we again built linear mixed effects models, this time modelling Leader evaluations. We built one model for each evaluation dimension: morality, trustworthiness, closeness, competence, and risk aversion.[1] All models predicted the given evaluation by a three-way interaction between condition, the cubic polynomial of round number, and the amount a participant cheated in a given evaluation period (i.e. the 5 rounds leading up to an evaluation), further controlling for Moral Foundation scores.

Regarding the effect of condition on evaluations, we find strong interactions of round number and condition in the predicted direction for all character evaluations, providing evidence for Hypothesis 2 (see Figure 3). Thus, as hypothesized, Leader checking generally had a negative effect for all Leader evaluation dimensions except risk aversion, for which no clear pattern emerged.

[Figure 3 about here]

[1] This follows the preregistered protocol. One may argue that, since all dimensions (except risk aversion) are correlated with each other, a single latent factor may drive the evaluations. Indeed, parallel analysis suggested that a single factor structure is sufficient to explain variation in these dimensions (explaining around 60% of their variance at each evaluation timepoint). However, it is important to note that the factor structure of this potential latent factor varies over time. More precisely, we could not reject weak (factorial) invariance for this single factor model over time. This analysis indicates that participants understood the various evaluation dimensions differently, and suggests that the effect of condition differentially affected evaluations (see online supplemental material).

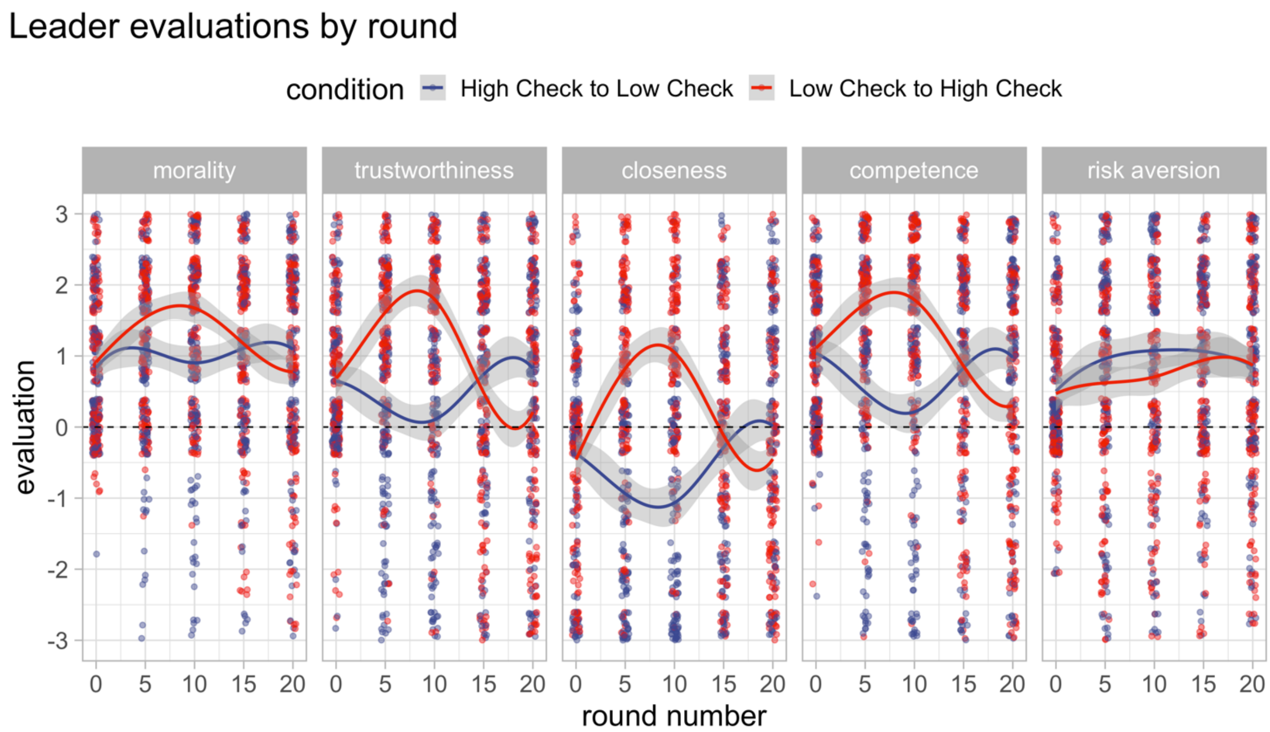

Figure 3: Leader evaluations by round number and condition. Leader evaluations occurred every 5 rounds, with a baseline evaluation at round 0. Lines represent b-splines with knots at round 10, and 3 degrees of freedom. Grey ribbons represent standard error of the mean.

Apart from the effect of condition, the effect of checking on evaluations is moderated by the degree of Player cheating for morality and closeness evaluations, but not for trustworthiness or competence. For closeness ratings, unchecked cheating leads to higher closeness ratings than when cheating is checked (ß = -29.37, 95% CI = [-55.13, -3.62], p = .026). For morality evaluations, the picture is more complex. Overall, participants who cheated more perceived their Leaders as less moral than participants that did not cheat as much (ß = -0.54, 95% CI = [-1.07, -0.01], p = .045). Yet, those participants that cheated often evaluated Leaders who switched to non-checking in the High Check to Low Check condition as more moral than Leaders who switched to checking in the Low Check to High check condition (ß =

14.55, 95% CI = [-0.42, 29.53], p = 0.057). However, these effects are only marginally significant, and await further investigation.

[Table 1 about here]

	Dependent variable:				
	Cheating	**Morality**	**Trustworthiness**	**Competence**	**Closeness**
	(1)	(2)	(3)	(4)	(5)
Intercept	**-1.19*****	**1.11*****	**0.53*****	**0.63*****	**-0.53*****
	(-1.77, -0.61)	(0.90, 1.31)	(0.30, 0.76)	(0.39, 0.87)	(-0.78, -0.27)
Condition†	0.20	0.23	**0.55****	**0.63*****	**0.85*****
	(-0.49, 0.89)	(-0.07, 0.52)	(0.21, 0.88)	(0.29, 0.98)	(0.49, 1.22)
Cheating		**-0.54***	-0.41	0.01	-0.44
		(-1.07, -0.01)	(-1.06, 0.24)	(-0.65, 0.68)	(-1.07, 0.19)
Round	**19.95*****	1.51	**9.47*****	**6.92****	**13.07*****
	(8.28, 31.61)	(-2.02, 5.05)	(5.22, 13.73)	(2.44, 11.40)	(8.95, 17.19)
Round^2	**20.16*****	**2.74***	2.44	**4.34****	**4.79*****
	(8.47, 31.85)	(0.65, 4.83)	(-0.21, 5.08)	(1.67, 7.02)	(2.34, 7.23)
Round^3	-9.36	-1.51	**-5.39*****	**-5.21*****	**-4.95*****
	(-21.12, 2.41)	(-3.55, 0.54)	(-7.98, -2.81)	(-7.83, -2.59)	(-7.34, -2.56)
Actual number rolled	**-1.03*****				
	(-1.14, -0.92)				
MF: Authority	-0.29	0.03	0.02	0.17	-0.03
	(-0.75, 0.17)	(-0.16, 0.21)	(-0.19, 0.23)	(-0.05, 0.40)	(-0.27, 0.21)
MF: Ingroup/Loyalty	-0.26	0.16	**0.25***	0.18	**0.39*****
	(-0.71, 0.19)	(-0.02, 0.35)	(0.04, 0.46)	(-0.03, 0.40)	(0.16, 0.63)
MF: Fairness	-0.31	0.04	0.07	0.09	0.12
	(-0.66, 0.03)	(-0.10, 0.19)	(-0.09, 0.24)	(-0.08, 0.26)	(-0.07, 0.30)
DOSPERT: investment	0.04				
	(-0.33, 0.41)				
DOSPERT: betting	**0.54****				
	(0.18, 0.91)				
Condition * Round	**-65.11*****	**-13.03*****	**-30.98*****	**-30.04*****	**-31.15*****
	(-82.56, -47.65)	(-18.04, -8.02)	(-37.02, -24.95)	(-36.40, -23.69)	(-36.99, -25.30)
Condition * Round^2	**-37.94*****	**-7.85*****	**-6.92*****	**-10.42*****	**-9.90*****
	(-54.81, -21.08)	(-10.83, -4.87)	(-10.70, -3.15)	(-14.24, -6.60)	(-13.39, -6.42)
Condition * Round^3	**21.60***	**4.02****	**13.96*****	**10.80*****	**12.96*****
	(4.62, 38.59)	(1.08, 6.96)	(10.23, 17.68)	(7.04, 14.57)	(9.53, 16.40)
Cheating * Round		-2.66	-2.24	12.46	13.70
		(-17.86, 12.55)	(-20.78, 16.30)	(-6.88, 31.81)	(-4.16, 31.56)
Cheating * Round^2		-9.18	-1.09	-6.63	-9.34
		(-19.92, 1.56)	(-14.65, 12.47)	(-20.37, 7.12)	(-21.93, 3.24)
Cheating * Round^3		-9.65	-0.40	-1.18	-6.01
		(-20.12, 0.82)	(-13.65, 12.86)	(-14.59, 12.22)	(-18.27, 6.25)
Condition * Cheating		0.14	-0.13	-0.75	0.30
		(-0.63, 0.92)	(-1.09, 0.82)	(-1.73, 0.23)	(-0.62, 1.22)
Condition * Cheating * Round		5.21	-3.54	6.24	**-29.37***
		(-16.79, 27.20)	(-30.39, 23.31)	(-21.72, 34.20)	(-55.13, -3.62)
Condition * Cheating * Round^2		14.55	4.97	16.54	8.01
		(-0.42, 29.53)	(-13.93, 23.87)	(-2.62, 35.69)	(-9.52, 25.54)
Condition * Cheating * Round^3		13.60	12.45	0.21	11.99
		(-0.61, 27.81)	(-5.54, 30.43)	(-17.98, 18.40)	(-4.64, 28.62)
Observations	6,160	1,232	1,232	1,232	1,232
Akaike Inf. Crit.	2,732.35	3,617.19	4,093.85	4,159.55	4,044.75
Note:					$^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$
					† Baseline Condition: High Check to Low Check

Table 1: Linear Mixed Effects model results for Hypotheses 1), and 2). Model 1 is a generalized mixed effects model that predicts cheating (as a binary value), testing Hypothesis 1. Parameter

estimates are log odds. Models 2-5 test for Hypothesis 2. Values in parentheses are 95% confidence intervals.

Because of the complex relationship emerging between checking and cheating, we analyzed their relationship in an exploratory way. More particularly, we analyzed whether the specific type of checking matters for evaluations, i.e. whether a check or a non-check was correct or not. The most interesting types of checks in our case are correct checks and incorrect *non*-checks, as both carry the strongest cues of a willingness (or lack thereof) to engage in collaborative cheating: on the one hand, correct checks indicate a Leader that wants to behave ethically, and who is paired with a Player that wants to cheat. Incorrect non-checks, on the other hand, signal collaborative corruption, as they can be interpreted as turning a blind eye.

To model the effects of correct checking and incorrect non-checking on Leader evaluations, we again created linear mixed effects models for each dimension. In these models, correct checking and incorrect non-checking were represented as counts ranging from 0 to 5 (e.g. in the five rounds leading up to an evaluation, a Leader could for example carry out three correct checks). However, because these check types are mutually exclusive (e.g. five correct checks imply that there can be no other type of check), we built individual models each using a different check type as predictor. In addition, we also specified interactions between check type (i.e. correct checks vs. incorrect non-checks) and personality traits of Fairness, Authority, as well as Ingroup/Loyalty, since we expected that check types might be interpreted differently by participants depending on how much they valued these norms. Because the number of checks systematically differed by condition, we did not include condition as a predictor. For the random effects structure, we specified random intercepts for participants, and random slopes for the specific check type.

This analysis revealed significant negative effects of correct checking on all evaluations, such that more correct checks by a Leader led to more negative evaluations of them. This effect was strongest for closeness evaluations and smallest for morality evaluations. Subsequent analysis revealed that these differences in effect sizes were significantly different as well: the influence of correct checks on updating evaluations was significantly smaller for morality ($p < .001$) and competence ($p = .0012$) ratings, relative to closeness ratings (see Figure 4; detailed results of these interaction models are presented in the supplementary material, see Table S1).

Turning to the effect of incorrect non-checks (i.e., turning a blind eye to cheating) on evaluations, we again found significant but positive relationships with trustworthiness and closeness evaluations, such that a Leader that let cheating go undetected was evaluated as closer and more trustworthy than a Leader that did not (see full results in Table 2). In addition, the positive effect of incorrect non-checks on perceived trustworthiness was negatively moderated by Fairness concerns, such that the relationship was stronger for participants who were less concerned with Fairness norms. Again, follow-up analysis revealed that the difference in the effects of incorrect non-checks on the different evaluation dimensions was significant. Thus, the effect of incorrect checks on Leader evaluations was smaller for morality ($p < .001$), and competence evaluations ($p = .045$), compared to closeness evaluations (see Table S1).

In sum, this analysis supports the notion that collaborative cheating can enhance evaluations of partners in crime compared to people refusing to engage in such practices, and that this effect is especially strong for closeness and trustworthiness ratings. These results are illustrated in Figure 4.

	Check type: correct check				*Check type: incorrect non-check*			
	Dependent variable							
	Morality (1)	Trustworthiness (2)	Competence (3)	Closeness (4)	Morality (5)	Trustworthiness (6)	Competence (7)	Closeness (8)
Intercept	**0.845*****	**0.786*****	**0.577*****	**0.392*****	**0.808*****	**0.650*****	**0.635*****	0.038
	(0.698, 0.992)	(0.596, 0.977)	(0.378, 0.777)	(0.204, 0.581)	(0.656, 0.961)	(0.452, 0.848)	(0.427, 0.843)	(-0.158, 0.234)
Previous rating	**0.739*****	**0.581*****	**0.663*****	**0.618*****	**0.715*****	**0.478*****	**0.522*****	**0.453*****
	(0.697, 0.781)	(0.533, 0.630)	(0.616, 0.710)	(0.574, 0.662)	(0.670, 0.760)	(0.427, 0.530)	(0.470, 0.574)	(0.405, 0.502)
Check type	**-0.259****	**-0.534*****	**-0.379*****	**-0.584*****	-0.038	**0.159***	0.065	**0.357*****
	(-0.423, -0.095)	(-0.709, -0.359)	(-0.575, -0.183)	(-0.730, -0.438)	(-0.129, 0.053)	(0.035, 0.282)	(-0.068, 0.198)	(0.230, 0.484)
MF: Fairness	0.013	-0.036	0.020	0.001	0.018	0.110	0.030	0.128
	(-0.055, 0.081)	(-0.134, 0.062)	(-0.069, 0.110)	(-0.097, 0.099)	(-0.060, 0.095)	(-0.012, 0.232)	(-0.090, 0.150)	(-0.003, 0.258)
MF: Authority	0.073	0.091	0.109	0.016	0.034	0.078	0.068	0.037
	(-0.013, 0.160)	(-0.034, 0.215)	(-0.004, 0.223)	(-0.109, 0.142)	(-0.064, 0.133)	(-0.076, 0.233)	(-0.084, 0.221)	(-0.129, 0.203)
MF: Ingroup	0.035	0.084	0.058	**0.141***	0.073	0.090	**0.161***	**0.170***
	(-0.050, 0.120)	(-0.038, 0.207)	(-0.052, 0.169)	(0.018, 0.264)	(-0.023, 0.169)	(-0.062, 0.241)	(0.012, 0.311)	(0.006, 0.333)
DOSPERT: betting	-0.042	-0.011	-0.013	0.052	-0.069	-0.099	-0.074	-0.026
	(-0.106, 0.021)	(-0.105, 0.083)	(-0.100, 0.075)	(-0.042, 0.147)	(-0.137, 0.0002)	(-0.211, 0.013)	(-0.183, 0.035)	(-0.153, 0.102)
DOSPERT: investment	0.021	0.042	-0.009	**0.101***	0.027	0.080	0.011	**0.159***
	(-0.042, 0.084)	(-0.051, 0.136)	(-0.095, 0.076)	(0.007, 0.195)	(-0.043, 0.096)	(-0.033, 0.194)	(-0.098, 0.121)	(0.032, 0.286)
Game period	**-0.192*****	**-0.154*****	**-0.109****	**-0.115*****	**-0.184*****	**-0.139*****	**-0.124*****	**-0.081****
	(-0.242, -0.142)	(-0.218, -0.090)	(-0.177, -0.041)	(-0.180, -0.049)	(-0.234, -0.134)	(-0.200, -0.078)	(-0.188, -0.059)	(-0.141, -0.020)
Check type * MF: Fairness	-0.016	0.127	-0.042	0.018	0.011	**-0.111***	-0.018	-0.100
	(-0.165, 0.133)	(-0.027, 0.282)	(-0.215, 0.131)	(-0.100, 0.136)	(-0.065, 0.088)	(-0.216, -0.006)	(-0.132, 0.097)	(-0.206, 0.006)
Check type * MF: Authority	-0.181	**-0.281***	-0.125	-0.108	0.009	0.023	0.105	0.001
	(-0.391, 0.030)	(-0.509, -0.053)	(-0.378, 0.129)	(-0.305, 0.088)	(-0.111, 0.128)	(-0.141, 0.187)	(-0.072, 0.281)	(-0.166, 0.169)
Check type * MF: Ingroup	0.152	0.085	0.069	-0.053	0.00003	0.064	-0.079	0.079
	(-0.056, 0.361)	(-0.141, 0.311)	(-0.183, 0.320)	(-0.245, 0.139)	(-0.106, 0.106)	(-0.083, 0.211)	(-0.239, 0.081)	(-0.072, 0.229)
Observations	1,232	1,232	1,232	1,232	1,232	1,232	1,232	1,232
Akaike Inf. Crit.	3,602.253	4,286.970	4,359.306	4,326.553	3,640.791	4,316.830	4,378.787	4,348.680
Note:								*p<0.05; **p<0.01; ***p<0.001

Table 2: Results for linear mixed effects models predicting Leader evaluations (morality, trustworthiness, competence & closeness) based on correct checks (models 1-4), and incorrect non-checks (models 5-8), controlling for participant personality, and game period (the 5 rounds leading up to an evaluation). Models include random intercepts for participants, and random slopes for check type.

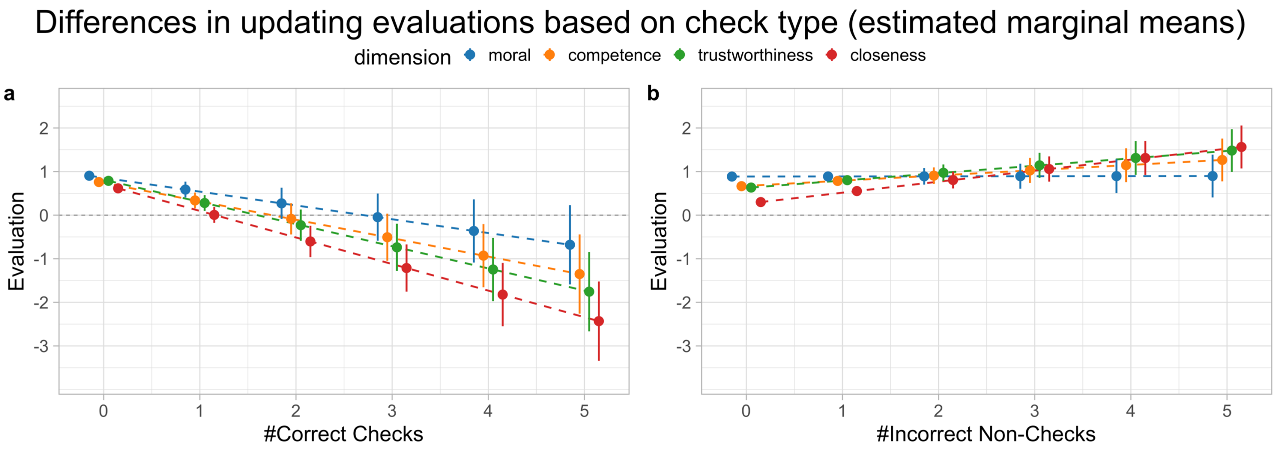

Figure 4: Estimated marginal means from linear mixed effects models predicting Leader evaluations based on the number of a specific check type in the five rounds leading up to an evaluation (i.e. rounds 1-5,6-10,11-15,16-20). Each plot represents a separate model. a: number of correct checks, b: number of correct non-checks. Error bars represent standard error. The models behind these marginal effects plot can be found in the supplemental material (see Table S1).

How does a change in checking behavior influence evaluations?

Next, we tested Hypothesis 3, predicting that a Leader switching from checking to non-checking will be perceived more positively than a Leader switching from non-checking to checking. To do so, we compared final evaluations of Leaders at the end of the rely-or-verify game between both conditions. In line with our hypothesis, we find that Players changed their evaluations to a greater degree when the Leader moved from non-checking to checking behavior than vice versa. Thus, participants evaluated their Leader to be less trustworthy and competent in the Low Check to High Check condition than in the High Check to Low Check condition. The same directional effects of condition existed for morality and closeness evaluations, even though they did not reach statistical significance (closeness ß

= -0.41, 95% CI = [-0.83, 0.01], p = 0.056; morality ß = -0.34, 95% CI = [-0.69, 0.003], p = .053; see Table 3). In addition to condition, participants' overall cheating emerged as a strong predictor for all dimensions, with participants who cheated more rating their Leaders less favorably.

	Dependent variable:				
	Morality (1)	Trustworthiness (2)	Competence (3)	Closeness (4)	Risk Aversion (5)
Intercept	**1.36*****	**1.10*****	**1.18*****	0.28	**0.98*****
	(1.10, 1.62)	(0.80, 1.40)	(0.87, 1.49)	(-0.04, 0.60)	(0.70, 1.27)
Low Check to High Check	-0.34	**-0.62****	**-0.69*****	-0.41	0.06
	(-0.69, 0.003)	(-1.02, -0.23)	(-1.10, -0.28)	(-0.83, 0.01)	(-0.32, 0.44)
MF: Authority	0.19	0.22	0.26	0.001	-0.14
	(-0.04, 0.42)	(-0.04, 0.49)	(-0.02, 0.53)	(-0.28, 0.28)	(-0.39, 0.12)
MF: Ingroup/Loyalty	0.22	0.23	0.23	**0.42****	0.02
	(-0.005, 0.44)	(-0.03, 0.49)	(-0.03, 0.50)	(0.14, 0.69)	(-0.23, 0.27)
MF: Fairness	0.12	0.07	-0.02	0.09	0.07
	(-0.06, 0.30)	(-0.14, 0.28)	(-0.23, 0.20)	(-0.13, 0.31)	(-0.12, 0.27)
DOSPERT: Investment	0.08	0.03	-0.12	0.09	0.01
	(-0.10, 0.27)	(-0.18, 0.24)	(-0.33, 0.10)	(-0.13, 0.31)	(-0.19, 0.21)
DOSPERT: Betting	-0.14	-0.03	0.03	0.20	0.03
	(-0.33, 0.04)	(-0.24, 0.19)	(-0.19, 0.25)	(-0.03, 0.43)	(-0.17, 0.24)
Cheating	**-2.42*****	**-2.61*****	**-1.86****	**-2.79*****	**-1.41***
	(-3.47, -1.38)	(-3.82, -1.41)	(-3.10, -0.62)	(-4.07, -1.52)	(-2.56, -0.26)
Observations	308	308	308	308	308
R^2	0.15	0.14	0.12	0.14	0.03
Adjusted R^2	0.13	0.12	0.10	0.12	0.01
F Statistic (df = 7; 300)	7.60***	7.06***	5.90***	6.97***	1.23
Note:					*p<0.05; **p<0.01; ***p<0.001

Table. 3. Linear models predicting final evaluations of Leaders based on condition, personality factors, and overall cheating during the game. Values in parentheses indicate 95% confidence intervals.

Finally, we turn to Hypothesis 4, which predicted that participants in the High Check to Low Check condition would send more money in the trust game to the Leader than participants in the Low Check to High Check condition. To model behavior in the trust game, we analyzed points sent to the Leader using a zero-one inflated binomial model (ZOIB). This was done because participants' contributions were not normally distributed, but comprised many participants sending either none (28.6% of participants) or all (16.2 % of participants) of the points they had earned during the previous game to their partner (see Figure 5). ZOIB models are ideally suited for modelling such data (Ospina & Ferrari, 2012), as they independently model the mean of a binomial distribution, its precision (the precision of the binomial distribution, determining how wide or narrow it is, called phi), the zero-one inflation parameter (the degree to which the distribution is zero-one inflated, called zoi), and the conditional one-inflation parameter (the likelihood that a value is 1, given it is either zero or one, called coi).

We predicted these model parameters by condition and the final Leader trustworthiness evaluation of the preceding game, controlling for evaluations of perceived Leader risk aversion, the mean amount a player had cheated during the rely-or-verify game, and their personality scores. In line with the preregistered protocol, we chose to include only trustworthiness and risk aversion evaluations, as evaluations for morality, trustworthiness, competence, and closeness were highly correlated with each other (all correlations > .63). We specified weakly informative priors for each of model intercepts (mean, phi, zoi, and coi), and the other parameters (all prior specifications were preregistered, see supplement).

Controlling for trustworthiness ratings, overall cheating, and personality factors, the model revealed an effect of condition both for mean and conditional one inflation, such that participants in the Check to No Check condition entrusted overall more money to the Leader ($ß_{mean}$ = -0.08, 95% Credible Interval = [-0.16; -0.01]), and were also more likely to send everything rather than nothing ($ß_{coi}$ = -0.15, 95% Credible Interval = [-0.29; -0.02], see also Figure 5). Further, trustworthiness evaluations predicted conditional one inflation, such that participants rating the Leader as more trustworthy were more likely to send everything rather than nothing ($ß_{coi}$ = 0.09, 95% Credible Intervals = [0.04; 0.15]). These results support Hypothesis 4.

In addition, participants who scored higher on Authority gave less money to the Leader ($ß_{mean}$ = -0.08, 95% Credible Interval = [-0.13; -0.03]). Further predictors of conditional one inflation were the amount of cheating in the previous game ($ß_{coi}$ = 0.46, 95% Credible Interval = [0.01; 0.75]), Leader risk aversion judgments ($ß_{coi}$ = 0.09, 95% Credible Interval = [0.03; 0.15]), and participant investment risk taking propensity ($ß_{coi}$ = 0.16, 95% Credible Intervals = [0.05; 0.27]). Full model results are presented in Table 4.

In sum, participants sent more money to a Leader that stopped rather than started checking, falling in line with Hypothesis 4. Above that, higher Leader trustworthiness ratings, higher levels of Player cheating, higher levels of trait financial risk taking, and higher ratings of Leader risk aversion all positively predicted money given to the Leader.

[Table 4 about here]

Distribution parameter	Term	Estimate	95% CI low	95% CI high
mean	Intercept	0.38	0.31	0.45
phi	Intercept	3.44	2.23	5.02
zoi	Intercept	0.47	0.37	0.57
coi	Intercept	0.29	0.16	0.44
mean	Condition*	**-0.08**	**-0.16**	**-0.01**
	Cheating	0.23	-0.08	0.49
	Leader Trustworthiness	0.01	-0.01	0.03
	Leader Risk aversion	0.02	-0.00	0.04
	MF: Fairness	0.00	-0.04	0.04
	MF: Authority	**-0.08**	**-0.13**	**-0.03**
	MF: Ingroup / Loyalty	0.04	-0.01	0.09
	DOSPERT: investment	-0.00	-0.04	0.04
	DOSPERT: betting	0.02	-0.02	0.05
phi	Condition*	0.05	-0.02	0.13
	Cheating	-0.26	-0.62	0.09
	Leader Trustworthiness	-0.01	-0.03	0.02
	Leader Risk aversion	-0.00	-0.03	0.02
	MF: Fairness	0.04	-0.00	0.08
	MF: Authority	**0.06**	**0.02**	**0.10**
	MF: Ingroup / Loyalty	-0.03	-0.09	0.02
	DOSPERT: investment	0.03	-0.01	0.07
	DOSPERT: betting	0.01	-0.03	0.05
zoi	Condition*	-0.01	-0.14	0.11
	Cheating	0.12	-0.23	0.44
	Leader Trustworthiness	-0.07	-0.10	-0.03
	Leader Risk aversion	0.00	-0.04	0.04
	MF: Fairness	0.04	-0.02	0.11
	MF: Authority	0.07	-0.01	0.15
	MF: Ingroup / Loyalty	-0.07	-0.14	0.01
	DOSPERT: investment	-0.06	-0.12	0.00
	DOSPERT: betting	0.06	-0.00	0.13
coi	Condition*	**-0.15**	**-0.29**	**-0.02**
	Cheating	**0.46**	**0.01**	**0.75**
	Leader Trustworthiness	**0.09**	**0.04**	**0.15**
	Leader Risk aversion	**0.09**	**0.03**	**0.15**
	MF: Fairness	-0.05	-0.12	0.03
	MF: Authority	0.04	-0.07	0.16
	MF: Ingroup / Loyalty	-0.03	-0.12	0.08
	DOSPERT: investment	**0.16**	**0.05**	**0.27**
	DOSPERT: betting	-0.01	-0.10	0.09

*Condition Baseline: High Check to Low Check

Table 4: Bayesian model estimates for contributions in the Trust Game, with 95% Credible Intervals. Estimates are derived assuming a zero-one inflated binomial distribution, which has four parameters: mean, phi (precision of the distribution), zoi (degree of zero-one inflation), and coi (conditional one-inflation). Estimates whose 95% credible intervals do not span 0 are printed in bold.

[Figure 5 about here]

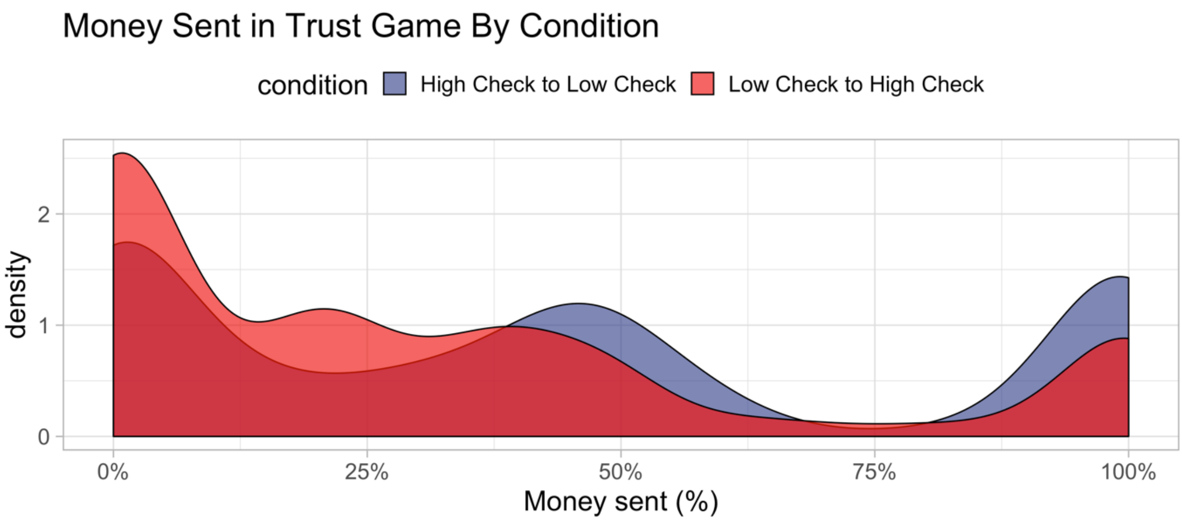

Figure 5: Distribution of money sent by Players to their Leaders in the trust game, by condition. Contributions are measured as percentage of the maximum possible amount.

Discussion

This study investigated how people are influenced by others' signaling a willingness to engage in collaborative corruption, and how people perceive their potential partners in crime. We found that lower rates of checking lead to more cheating, and that Leaders who check often are evaluated more negatively than Leaders who do not. Importantly, consistent non-checking of cheating enhanced trustworthiness ratings and perceived psychological closeness, and had no negative impact on morality and competence ratings. The trustworthiness bonus translated to actual behavior in a subsequent trust game: Players who interacted with a Leader who stopped checking entrusted the Leader with more money. Below we interpret our findings in the context of existing literature, and discuss their implications (see Figure 6 for a summary).

Our results align with work showing that higher likelihoods of being checked result in lower rates of cheating, e.g., Thielmann and Hilbig (2018). Our study extends prior findings by investigating these effects when Players believe that a person rather than a probabilistic process determines the checking. Thus, participants' inferences about likely Leader decisions govern their behavior, a trend that is visible in the nonlinear development of cheating throughout the game. Participants cheated less as they learned that their Leader would be prone to checking. Yet, participants were also highly sensitive to the change in Leader behavior: we observed no overall difference in cheating between conditions. This pattern implies that an initial phase of strong checking is not sufficient to keep dishonesty in check, but that repeated checking is necessary to curb cheating.

Additionally, the results suggest that repeated checking may negatively impact trust, as well as perceived closeness, presenting a clear dilemma for people in supervisory roles. As introduced, the phenomenon of collaborative corruption carries an inherent tension for how people engaging in it should perceive each other. Our study suggests that, at least in the context of rely-or-verify situations, collaborative corruption has largely positive effects on person perception. In fact, letting cheating go unchecked was perceived as particularly positive. This effect also holds for competence evaluations, indicating that participants did not believe that Leaders were placing undeserved trust in them, but instead rated non-checking even of dishonest reports as signs of their Leaders' competence. This effect is underlined by participants' behavior in the trust game: positive evaluations of non-checking Leaders transferred to a new context testing actual costly behavior. Therefore, collaborative corruption appears to comprise many of the benefits usually tied to positive collaborations, such as increasing intra-group trust and enhancing experienced psychological closeness (Zelmer, 2003).

Psychological closeness deserves special discussion, as it showed the strongest sensitivity towards Leader checking behavior. Perceived closeness is an important factor for moral judgment and behavior: people will often engage in bad behavior if they observe close others' engaging in bad behavior (Gino & Galinsky, 2012). However, while earlier findings have been concerned with the influence of close others, here we show that cheating together can itself increase the experience of closeness. This has the potential to spark a vicious cycle: people who cheat together feel closer to each other, and then are even more influenced by their partner's bad behavior. Future work could probe this cycle by for example manipulating perceived closeness via in-group membership.

But why do people regard their partners in crime positively, and trust them (to the point of investing in them) even in new contexts? One potential mechanism originates from work on social identity theory (Turner & Tajfel, 1986), and moral disengagement (Bandura, 1999; Shu, Gino, & Bazerman, 2011). Studies have shown that people are quick to form in-group biases and display in-group-favoritism, leading to moral disengagement (Branscombe, Ellemers, Spears, & Doosje, 1999; Mazar & Aggarwal, 2011; Mazar, Amir, & Ariely, 2008). In the case of collaborative corruption, cheating benefits not only the self but collaborators, which allows for an effective way of rationalizing immoral behavior as beneficial for the group (Fehr et al., 2019). In fact, people recognize special obligations to close others like family, and failing to adhere to these obligations is seen as especially immoral (McManus, Kleiman-Weiner, & Young, 2020). A Player who cares more about fairness may therefore demand honest behavior, whereas a Player who cares more about loyalty may defend collaborative cheating so long as it benefits the in-group (Geva, 2006; Graham et al., 2011). Consistent with this interpretation is the moderating role of fairness concerns on Leader evaluations: in our study, the effect of incorrect non-checks on perceived trustworthiness was particularly strong for participants assigning less value to fairness norms. Future work could explore this link more directly, by for instance examining how Leaders' checking is perceived when Player cheating is costly rather than beneficial to the Leader.

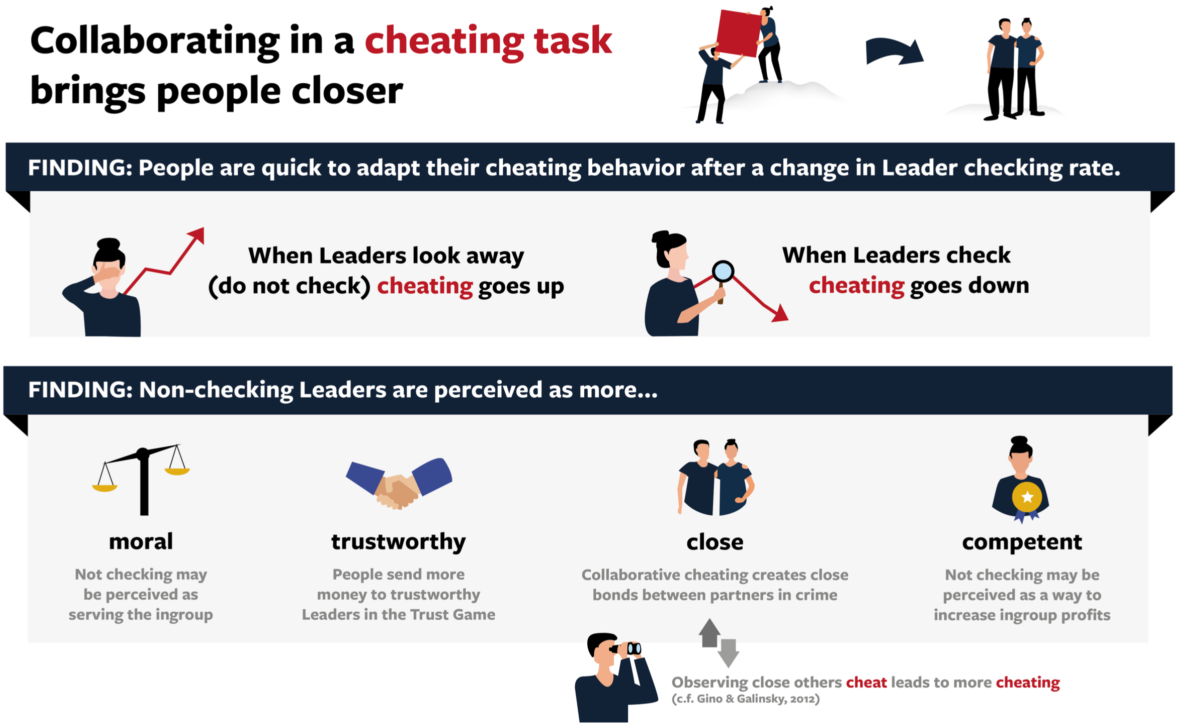

Figure 6: Summary of main findings.

Conclusion

When people collaborate successfully, they form strong bonds of trust and friendship. As we show, such collaborative ties develop not only for positive collaborations, but also for corrupt collaborations. In the present work, leaders who functioned as partners in crime, turning a blind eye to cheating, were evaluated positively, whereas leaders who spoiled opportunities to cheat by checking reports were evaluated negatively. These findings suggest that collaborative corruption, as in the Volkswagen case, may be especially robust and requires strong outside regulation to be curtailed.

Acknowledgements

We thank Joshua Hirschfeld-Kroen, Kevin Jiang, Gordon Kraft-Todd, Justin Martin, Ryan McManus, and BoKyung Park for helpful feedback and comments, Savhannah Schulz for creative development of Figure 6, and the John Templeton Foundation as well as the Boston College Virtue Project for financial support.

Literature

Bandura, A. (1999). Moral disengagement in the perpetration of inhumanities. *Personality and Social Psychology Review, 3*(3), 193-209. doi:10.1207/s15327957pspr0303_3

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Usinglme4. *Journal of Statistical Software, 67*(1). doi:10.18637/jss.v067.i01

Berg, J., Dickhaut, J., & Mccabe, K. (1995). Trust, Reciprocity, and Social-History. *Games and Economic Behavior, 10*(1), 122-142. doi:DOI 10.1006/game.1995.1027

Bhattacharjee, A., Berman, J. Z., & Reed, A. (2013). Tip of the Hat, Wag of the Finger: How Moral Decoupling Enables Consumers to Admire and Admonish. *Journal of Consumer Research, 39*(6), 1167-1184. doi:10.1086/667786

Blais, A. R., & Weber, E. U. (2006). A Domain-Specific Risk-Taking (DOSPERT) scale for adult populations. *Judgment and Decision Making Journal, 1*(1), 33-47. Retrieved from <Go to ISI>://WOS:000203681500004

Branscombe, N. R., Ellemers, N., Spears, R., & Doosje, B. (1999). The context and content of social identity threat. In R. Spears, N. Ellemers, & B. Doosje (Eds.), *Social identity: Context, commitment, content* (pp. 35-58): Wiley.

Bürkner, P.-C. (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software, 80*(1). doi:10.18637/jss.v080.i01

Chen, D. L., Schonger, M., & Wickens, C. (2016). oTree—An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance, 9*, 88-97. doi:https://doi.org/10.1016/j.jbef.2015.12.001

Fehr, R., Welsh, D., Yam, K. C., Baer, M., Wei, W., & Vaulont, M. (2019). The role of moral decoupling in the causes and consequences of unethical pro-organizational behavior. *Organizational Behavior and Human Decision Processes, 153*, 27-40. doi:10.1016/j.obhdp.2019.05.007

Fischbacher, U., & Föllmi-Heusi, F. J. J. o. t. E. E. A. (2013). Lies in disguise—an experimental study on cheating. *11*(3), 525-547.

Fiske, S. T., Cuddy, A. J., & Glick, P. (2007). Universal dimensions of social cognition: warmth and competence. *Trends Cogn Sci, 11*(2), 77-83. doi:10.1016/j.tics.2006.11.005

Gershman, S. J. (2018). How to never be wrong. *Psychon Bull Rev*. doi:10.3758/s13423-018-1488-8

Geva, A. (2006). A Typology of Moral Problems in Business: A Framework for Ethical Management. *Journal of Business Ethics, 69*(2), 133-147. doi:10.1007/s10551-006-9072-y

Gino, F., & Galinsky, A. D. (2012). Vicarious dishonesty: When psychological closeness creates distance from one's moral compass. *Organizational Behavior and Human Decision Processes, 119*(1), 15-26. doi:10.1016/j.obhdp.2012.03.011

Goodman, L. M. (2015). Why Volkswagen Cheated. *Newsweek Magazine*. Retrieved from https://www.newsweek.com/2015/12/25/why-volkswagen-cheated-404891.html

Goodwin, G. P., Piazza, J., & Rozin, P. (2014). Moral character predominates in person perception and evaluation. *J Pers Soc Psychol, 106*(1), 148-168. doi:10.1037/a0034726

Graham, J., Nosek, B. A., Haidt, J., Iyer, R., Koleva, S., & Ditto, P. H. (2011). Mapping the moral domain. *J Pers Soc Psychol, 101*(2), 366-385. doi:10.1037/a0021847

Gross, J., Leib, M., Offerman, T., & Shalvi, S. (2018). Ethical Free Riding: When Honest People Find Dishonest Partners. *Psychological Science, 29*(12), 1956-1968. doi:10.1177/0956797618796480

Highhouse, S., Nye, C. D., Zhang, D. C., & Rada, T. B. (2017). Structure of the Dospert: Is There Evidence for a General Risk Factor? *Journal of Behavioral Decision Making, 30*(2), 400-406. doi:10.1002/bdm.1953

Kocher, M. G., Schudy, S., & Spantig, L. (2018). I Lie? We Lie! Why? Experimental Evidence on a Dishonesty Shift in Groups. *Management Science, 64*(9), 3995-4008. doi:10.1287/mnsc.2017.2800

Kuznetsova, A., Brockhoff, P., & Christensen, R. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software, 82*, 1-26. doi:10.18637/jss.v082.i13

Lapsley, D. K., & Lasky, B. (2001). Prototypic Moral Character. *Identity, 1*(4), 345-363. doi:10.1207/s1532706xid0104_03

Levine, E. E., & Schweitzer, M. E. (2015). Prosocial lies: When deception breeds trust. *Organizational Behavior and Human Decision Processes, 126*, 88-106. doi:10.1016/j.obhdp.2014.10.007

Magee, L. (1998). Nonlocal Behavior in Polynomial Regressions. *The American Statistician, 52*(1), 20-22. doi:10.1080/00031305.1998.10480531

Mazar, N., & Aggarwal, P. (2011). Greasing the palm: can collectivism promote bribery? *Psychol Sci, 22*(7), 843-848. doi:10.1177/0956797611412389

Mazar, N., Amir, O., & Ariely, D. (2008). The Dishonesty of Honest People: A Theory of Self-Concept Maintenance. *Journal of Marketing Research, 45*(6), 633-644. doi:DOI 10.1509/jmkr.45.6.633

McManus, R. M., Kleiman-Weiner, M., & Young, L. (2020). What We Owe to Family: The Impact of Special Obligations on Moral Judgment. *Psychol Sci, 31*(3), 227-242. doi:10.1177/0956797619900321

Ospina, R., & Ferrari, S. L. P. (2012). A general class of zero-or-one inflated beta regression models. *Computational Statistics & Data Analysis, 56*(6), 1609-1623. doi:10.1016/j.csda.2011.10.005

R Core Team. (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2013. In.

Reeder, G. D., & Brewer, M. B. (1979). A schematic model of dispositional attribution in interpersonal perception. *Psychological Review, 86*(1), 61-79. doi:10.1037/0033-295x.86.1.61

Rousseau, D. M., & McLean Parks, J. (1993). The contracts of individuals and organizations. *Research in Organizational Behavior, 15*, 1-1.

Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review, 5*(4), 296-320. doi:Doi 10.1207/S15327957pspr0504_2

Shalvi, S. (2016). Behavioural economics: Corruption corrupts. *Nature, 531*(7595), 456-457. doi:10.1038/nature17307

Shu, L. L., Gino, F., & Bazerman, M. H. (2011). Dishonest deed, clear conscience: when cheating leads to moral disengagement and motivated forgetting. *Pers Soc Psychol Bull, 37*(3), 330-349. doi:10.1177/0146167211398138

Soraperra, I., Weisel, O., Zultan, R. i., Kochavi, S., Leib, M., Shalev, H., & Shalvi, S. (2017). The bad consequences of teamwork. *Economics Letters, 160*, 12-15. doi:10.1016/j.econlet.2017.08.011

Thielmann, I., & Hilbig, B. E. (2018). Daring dishonesty: On the role of sanctions for (un)ethical behavior. *Journal of Experimental Social Psychology, 79*, 71-77. doi:10.1016/j.jesp.2018.06.009

Turner, J. C., & Tajfel, H. (1986). The social identity theory of intergroup behavior. *Psychology of intergroup relations, 5*, 7-24.

Weisel, O., & Shalvi, S. (2015). The collaborative roots of corruption. *Proc Natl Acad Sci U S A, 112*(34), 10651-10656. doi:10.1073/pnas.1423035112

Zelmer, J. (2003). Linear Public Goods Experiments: A Meta-Analysis. *Experimental Economics*(6), 299-310. doi:10.1023/A:1026277420119

Supplementary Materials for Review

These are the supplementary materials for review for the article submission "Collaborative cheating brings people closer: How people evaluate and interact with partners in crime".

Overview

This supplement contains the following

1. Interaction models for incorrect checks and correct non-checks (Table S1)
2. Comparison Models for non-cheaters (Table S2)
3. Graph comparing trajectories of Leader evaluations for cheaters and non-cheaters (Figure S1)

Please note: These supplementary materials present only a subset of the full analyses and robustness checks available in the full online supplementary materials, which we provide in an effort to adhere with open science best practices. We invite reviewers to explore the additionally offered materials.

The full online supplementary materials (data, complete analysis code, full experiment code and materials) are presented in anonymized form here:

https://osf.io/p2esr/?view_only=ef604a56ad9d4639ae5e670271b9d76c

1. Differential impact of checking type on different evaluation dimensions

Table S1 presents model results comparing the impact of different types of checks (i.e. correct check, incorrect check, correct noncheck, incorrect noncheck) on evaluation dimensions (i.e. morality, trustworthiness, competence, closeness; treating closeness evaluations as the baseline). These models underlie the marginal means plots shown in Figure 4 of the main article.

To analyze the differential impact of checking type on the different evaluation dimensions, we use linear mixed effects models, treating the different Leader evaluations (morality, trustworthiness, competence, closeness) as repeated measures. For the random effects structure, we are specifying random intercepts for evaluation rounds nested within participants, adding the particular check type as a random slope (however forcing random slopes to be uncorrelated with intercepts in order to counter singular fit issues).

Table S1

	Dependent variable:			
	Leader evaluations modelled using different check types:			
	correct check	incorrect check	correct noncheck	incorrect noncheck
	(1)	(2)	(3)	(4)
Intercept	0.506***	1.147***	-0.873***	0.240***
	(0.396, 0.615)	(1.038, 1.257)	(-1.022, -0.724)	(0.126, 0.353)
Prior Evaluation	0.710***	0.694***	0.601***	0.723***
	(0.688, 0.731)	(0.675, 0.713)	(0.578, 0.624)	(0.701, 0.746)
Competence	0.145**	0.091	0.458***	0.209***
	(0.039, 0.250)	(-0.017, 0.198)	(0.300, 0.616)	(0.098, 0.320)
Morality	0.289***	-0.059	1.093***	0.411***
	(0.182, 0.395)	(-0.168, 0.049)	(0.935, 1.252)	(0.299, 0.523)
Trustworthiness	0.172**	0.123*	0.398***	0.207***
	(0.067, 0.277)	(0.016, 0.230)	(0.241, 0.555)	(0.097, 0.316)
Check Type†	-0.609***	-0.438***	0.410***	0.282***
	(-0.794, -0.424)	(-0.490, -0.387)	(0.364, 0.456)	(0.162, 0.401)
MF: Fairness	-0.006	-0.004	-0.002	0.038
	(-0.045, 0.033)	(-0.042, 0.034)	(-0.053, 0.049)	(-0.002, 0.078)
MF: Authority	0.064*	0.088***	0.020	0.057*
	(0.015, 0.114)	(0.039, 0.138)	(-0.047, 0.087)	(0.006, 0.107)
MF: Ingroup/Loyalty	0.059*	0.057*	0.132***	0.075**
	(0.011, 0.108)	(0.007, 0.106)	(0.066, 0.198)	(0.025, 0.125)
Dospert: Betting	0.0003	-0.031	-0.0001	-0.044*
	(-0.039, 0.040)	(-0.071, 0.008)	(-0.053, 0.053)	(-0.085, -0.003)
Dospert: Investment	0.029	0.031	-0.008	0.027
	(-0.010, 0.068)	(-0.009, 0.071)	(-0.062, 0.047)	(-0.013, 0.067)
Round Number	-0.153***	-0.181***	-0.115***	-0.126***
	(-0.186, -0.121)	(-0.213, -0.150)	(-0.152, -0.078)	(-0.160, -0.093)
Competence * Check Type†	0.187*	0.068**	-0.051*	-0.130*
	(0.019, 0.355)	(0.022, 0.115)	(-0.099, -0.003)	(-0.257, -0.003)
Morality * Check Type†	0.293***	0.262***	-0.211***	-0.248***
	(0.124, 0.461)	(0.216, 0.309)	(-0.259, -0.164)	(-0.374, -0.121)
Trustworthiness * Check Type†	0.101	0.052*	-0.035	-0.084
	(-0.067, 0.269)	(0.005, 0.098)	(-0.083, 0.012)	(-0.211, 0.043)
Observations	4,928	4,928	4,928	4,928
Log Likelihood	-8,133.341	-7,265.040	-7,694.247	-8,289.398
Akaike Inf. Crit.	16,302.680	14,566.080	15,424.490	16,614.790
Note:				*p<0.05; **p<0.01; ***p<0.001

†Check type is different for each model: Model (1) uses correct checks, Model (2) uses incorrect checks, Model (3) uses correct nonchecks, and Model (4) uses incorrect nonchecks.

Table S1: Linear mixed effects models modelling the differential effect (interactions) of check type on the different evaluation dimensions.

2. Robustness Analysis of models using a subset of only cheaters.

Table S2 presents a robustness analysis for models reported in Table 1 in the main article. The robustness analysis uses only data from participants who cheated at least once (N = 151).

(*Note*: A visual comparison of these model estimates is presented in the additional online materials, see online supplementary analysis 3.1.6, 3.2.1.1.7, 3.2.1.2.7, 3.2.1.3.7, and 3.2.1.4.7).

Table S2

	Dependent variable:				
	Cheating	Morality	Trustworthiness	Competence	Closeness
	generalized linear mixed-effects	*linear mixed-effects*	*linear mixed-effects*	*linear mixed-effects*	*linear mixed-effects*
	(1)	(2)	(3)	(4)	(5)
Intercept	**1.19***** (0.77, 1.61)	**0.89***** (0.57, 1.21)	0.17 (-0.17, 0.52)	0.23 (-0.15, 0.61)	**-0.92***** (-1.30, -0.55)
Condition	0.09 (-0.38, 0.55)	0.17 (-0.29, 0.63)	**0.61*** (0.11, 1.10)	**0.79**** (0.25, 1.33)	**0.83**** (0.29, 1.37)
Cheating		-0.58 (-1.21, 0.06)	-0.25 (-1.00, 0.50)	0.16 (-0.61, 0.93)	-0.29 (-0.98, 0.40)
Round	**14.13***** (6.07, 22.19)	-0.61 (-5.30, 4.08)	**5.76*** (0.06, 11.47)	**4.00** (-2.03, 10.03)	**8.60**** (3.16, 14.04)
Round^2	**14.41***** (6.20, 22.62)	**3.82*** (0.76, 6.88)	1.01 (-2.67, 4.70)	**4.35*** (0.64, 8.07)	**3.69*** (0.45, 6.93)
Round^3	-6.49 (-14.60, 1.63)	-0.59 (-3.44, 2.26)	**-4.28*** (-7.72, -0.84)	**-4.20*** (-7.66, -0.74)	**-4.06**** (-7.07, -1.06)
Actual number rolled	**-1.01***** (-1.12, -0.91)				
MF: Authority	0.05 (-0.26, 0.35)	-0.15 (-0.42, 0.12)	-0.27 (-0.56, 0.01)	0.02 (-0.30, 0.34)	-0.26 (-0.58, 0.07)
MF: Ingroup/Loyalty	-0.18 (-0.48, 0.11)	0.24 (-0.03, 0.50)	**0.38**** (0.10, 0.66)	0.18 (-0.13, 0.49)	**0.50**** (0.19, 0.82)
MF: Fairness	-0.20 (-0.42, 0.02)	0.05 (-0.14, 0.24)	0.03 (-0.18, 0.23)	0.09 (-0.14, 0.31)	-0.01 (-0.24, 0.22)
DOSPERT: investment	**-0.29*** (-0.54, -0.03)				
DOSPERT: betting	**0.33**** (0.09, 0.57)				
Condition * Round	**-45.46***** (-57.48, -33.44)	-5.79 (-12.46, 0.87)	**-22.53***** (-30.61, -14.46)	**-18.92***** (-27.48, -10.37)	**-19.98***** (-27.70, -12.27)
Condition * Round^2	**-26.90***** (-38.73, -15.06)	**-8.34***** (-12.73, -3.95)	**-5.46*** (-10.75, -0.17)	**-10.22***** (-15.55, -4.89)	**-8.89***** (-13.53, -4.25)
Condition * Round^3	**15.52**** (3.89, 27.14)	2.22 (-1.96, 6.40)	**11.96***** (6.92, 17.00)	**7.74**** (2.67, 12.80)	**11.22***** (6.82, 15.62)
Cheating * Round		0.30 (-14.07, 14.67)	-1.70 (-18.86, 15.46)	9.26 (-8.67, 27.19)	9.85 (-6.13, 25.84)
Cheating * Round2		**-11.30*** (-22.22, -0.39)	0.64 (-12.46, 13.73)	-8.17 (-21.44, 5.10)	-7.65 (-19.24, 3.94)
Cheating * Round3		-7.20 (-17.31, 2.92)	0.94 (-11.25, 13.13)	1.09 (-11.18, 13.36)	-2.52 (-13.19, 8.15)
Condition * Cheating		0.30 (-0.63, 1.23)	-0.11 (-1.21, 0.99)	-0.71 (-1.84, 0.43)	0.43 (-0.59, 1.45)
Condition * Cheating * Round^1		-4.11 (-24.51, 16.29)	0.24 (-24.35, 24.82)	1.09 (-24.57, 26.75)	-22.54 (-45.44, 0.36)
Condition * Cheating * Round^2		**16.71*** (1.65, 31.77)	5.18 (-12.94, 23.29)	**19.00*** (0.68, 37.32)	10.55 (-5.44, 26.54)
Condition * Cheating * Round^3		10.14 (-3.87, 24.16)	4.07 (-12.82, 20.97)	-0.69 (-17.69, 16.30)	3.53 (-11.25, 18.31)
Observations	3,020	604	604	604	604
Akaike Inf. Crit.	2,349.49	1,901.73	2,086.43	2,137.70	2,032.39
Note:					*p<0.05; **p<0.01; ***p<0.001

Table S2: Model results for Player cheating and Leader evaluations when using only data from participants who cheated at least once. Model results here compare to Table 1 in the main article.

3. Graph comparing trajectories of Leader evaluations for cheaters and non-cheaters (Figure S1)

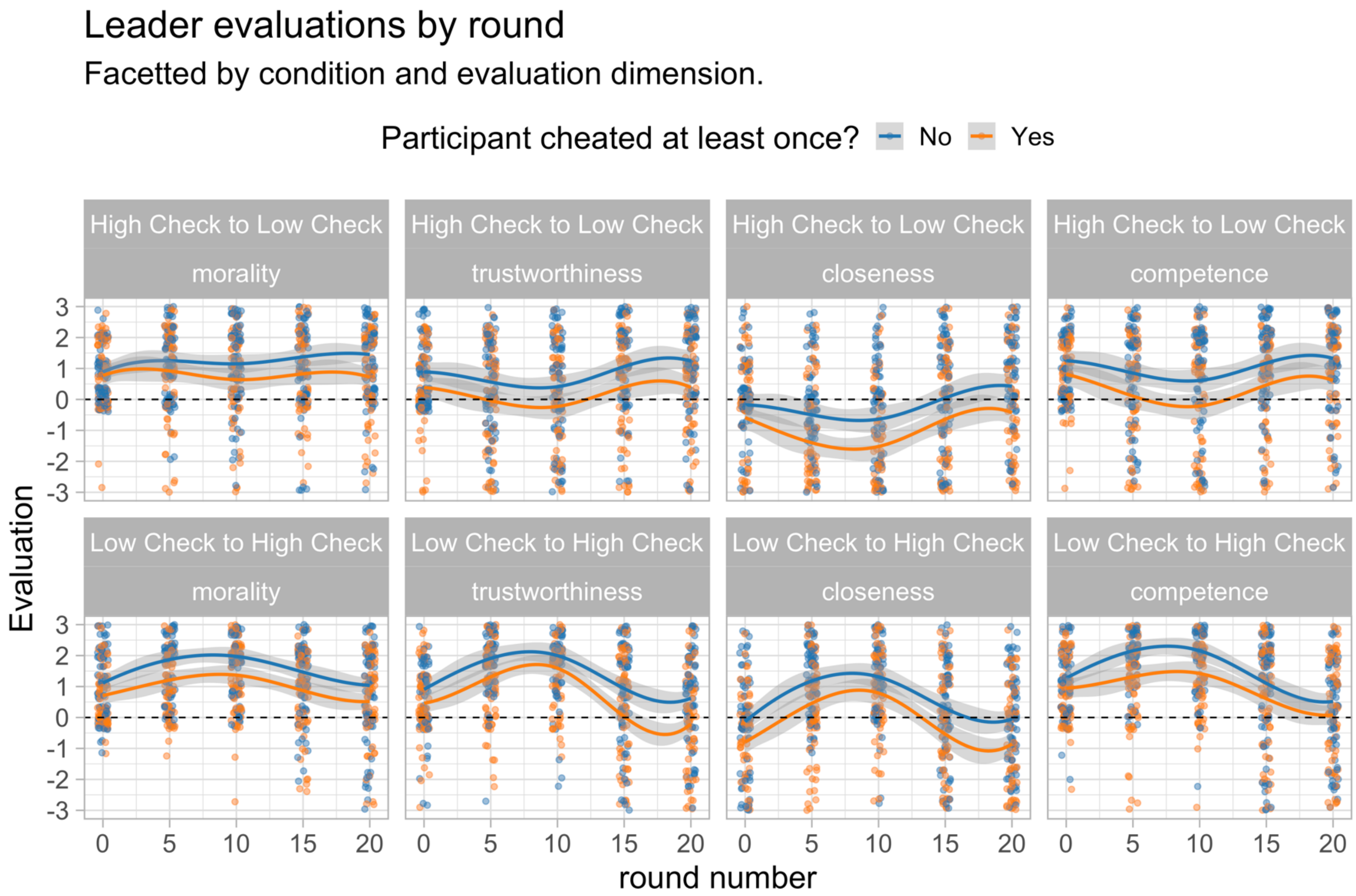

Figure S1: Leader evaluations over round, comparing participant who cheated at least once with participants who never cheated. Facetted by condition and evaluation dimension. Lines are b-splines with 3 degrees of freedom, and a knot at round 10.