Empirical Research Paper

Collaborative Cheating in Hierarchical Teams: Effects of Incentive Structure and Leader Behavior on Subordinate Behavior and Perceptions of Leaders

Personality and Social Psychology Bulletin
1–18

DOI: 10.1177/01461672221090859
journals.sagepub.com/home/pspb
SAGE

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

Abstract

What facilitates collaborative cheating in hierarchical teams, and what are its outcomes for those engaged? In two preregistered studies ($N = 724$), we investigated how subordinates are influenced by leaders signaling a willingness to engage in collaborative cheating, and how subordinates perceive such leaders. Participants performed a task in which they could either report their performance honestly, or cheat for financial gain. Each participant was assigned a leader who could choose to check the report's veracity. In Study 1, 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 investments in a subsequent trust game. Study 2 revealed that these relationship benefits specifically arise for collaborative cheating, compared to competitive cheating (at the leader's expense). We conclude that collaborative cheating in subordinate–leader dyads strengthens in-group bonds, bringing people closer together and cultivating trust.

Keywords

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

Received July 15, 2021; revision accepted March 10, 2022

Introduction

Collaboration is undoubtedly a key requirement of a flourishing society. Yet, collaboration can also serve immoral aims, unraveling the social fabric it otherwise helps to construct. One prominent example is the 2015 Volkswagen Dieselgate scandal, in which the company's top-management turned a blind eye to a fraudulent method devised to seemingly lower emissions (Goodman, 2015). The willful ignorance of leaders toward dishonest behavior by their subordinates constitutes a primary case of so-called collaborative corruption. This phenomenon is a perilous one; it is both widespread and can lead to higher degrees of cheating compared to individual settings (Weisel & Shalvi, 2015).

The frequency of collaborative corruption between leaders and subordinates is echoed in many real-world examples of corruption (e.g., Dieselgate, ENRON, Cum-ex, Operation Car Wash), showing that corruption frequently involves leaders and subordinates collaborating toward dishonest goals, with leaders turning a blind eye to or promoting corrupt activities (den Nieuwenboer et al., 2017; Ernst & Young, 2020; Pinto et al., 2008). While a large literature explores the role-modeling effect of leader behavior on subordinates' behavior (Gächter & Renner, 2018; Lemoine et al., 2019; Mayer et al., 2012; Trevino et al., 2014), the exploration of collaborative corruption in hierarchical settings is still relatively novel. In particular, although some research has examined progenitors of willful ignorance (e.g., Pittarello et al., 2015), the subordinate side of such forms of collaborative corruption is less examined. We therefore concentrate on how subordinates who engage in a new hierarchical collaboration perceive their potential co-conspiring leaders, and how these perceptions may explain the robustness of collaborative corruption.

Whereas cheating is usually conceived of as a competitive act, in which one person is hurting another, collaborative cheating requires individuals to engage in dishonest conduct at the expense of a third party. The collaborative nature of

[1]Aarhus University, Denmark
[2]Boston College, Chestnut Hill, MA, USA
[3]Duke University, Durham, NC, USA

Corresponding Author:
Simon Tobias Karg, Department of Political Science, Aarhus University, Bartholins Allé 7, Aarhus 8000, Denmark.
Email: simonkarg@ps.au.dk

cheating in this case affects behavior in crucial ways: research finds that dishonesty increases when more people benefit from it (Gino et al., 2013; Soraperra et al., 2017). Thus, people may attempt to justify their cheating as benefiting their immediate partner while discounting harm brought upon more distant others (Weisel & Shalvi, 2015). Indeed, the longer individuals cooperate with a partner in a potentially corrupt relationship, the greater the chances are for corruption to occur (Abbink, 2004)—especially when relationships are reciprocal (Song & Zhong, 2015; Thielmann et al., 2021), or when individuals know each other well (Akbari et al., 2020; Irlenbusch et al., 2020).

Despite this growing empirical evidence on the importance of the relationship between coconspirators, little is known about the actual kinds of relationships collaborative corruption establishes. In other words, how do coconspirators come to perceive each other? Do people trust their partners, or are they wary of them? Collaborative corruption presents a remarkable ambiguity for the people engaging in it. On one hand, knowledge that someone behaves dishonestly could translate to generally negative character evaluations. Therefore, one might expect people to form negative evaluations of coconspirators, even if they currently benefit from their dishonesty (Gross et al., 2018), as they recognize that their coconspirator may one day turn against them (Heintz et al., 2016). On the other hand, successful collaborative efforts create common ground and enhance ingroup trust (Rousseau & McLean Parks, 1993; Zelmer, 2003). From this perspective, if collaborative corruption is similar to other kinds of collaboration, engaging in it will yield positive effects, establishing close relationships and strong trust (Cole & Teboul, 2004). 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 coconspirators as competent, they might judge them to be immoral (Bhattacharjee et al., 2013). And, although they may trust them in one context, they might not in another.

This article contributes to the growing literature on collaborative corruption by investigating subordinate–leader dyads. We conduct two studies examining three main outcomes of collaborative cheating in hierarchical settings: subordinate cheating behavior, character evaluations of leaders, and subsequent trust. We do so using a repeated economic game—the rely-or-verify game—that models a hierarchical, dyadic setting of a leader and a subordinate, in which short-term outcomes for the dyad are dependent on either *reported* or *actual* performance. In this game, a player (i.e., the subordinate) engages in a task and conveys information about their performance to their leader, who then decides whether to verify (check) or to rely on (not check) that information. We operationalize player cheating as the reporting of performances that are higher than actual performance. Moreover, we implement specific incentive structures that can cleanly operationalize cheating to either be collaborative (beneficial to both leader and subordinate) or noncollaborative (cheating at the cost of the leader). Importantly, if the leader checks, any potential cheating by the subordinate is detected, with costly consequences to the subordinate, and, if cheating is collaborative, to the leader as well.

After several iterations of the game, players have the opportunity to invest in their leaders in a trust game, affording a test of how previously established collaborative cheating translates to a new context. In both studies, all participants were assigned to the role of player and believed their leader to be a human partner, whereas in reality leader responses were preprogrammed.

Study 1 manipulates leader checking rate, which affects whether cheating will be profitable or not. Study 2 manipulates the payoff structure—whether leaders financially benefit (collaborative cheating) or suffer from the players' cheating (competitive cheating). Examples of leaders benefiting from subordinate cheating might be found in organizations employing bonus schemes tied to performance reports or generated income from cheated customers. Leaders suffering from subordinate cheating may occur via decreased efficiency, or the loss of customers. We focus on the comparison between these two cases, as opposed to potential scenarios in which the leader is neutral or indifferent to the player's cheating, as the latter lack the critical *relationship* component of primary interest.

Next, we outline our hypotheses regarding how checking rate and payoff structures will affect cheating, leader evaluations, and subsequent trust (see Table 1).

Cheating

Unethical leaders can bring about unethical behavior in their subordinates (Fehr et al., 2019; Lemoine et al., 2019; Trevino et al., 2014). Yet, there is little empirical evidence for the impact of being checked on cheating behavior, and the scant existing evidence presents a mixed picture. Gino et al. (2009) found that, even when made very clear that cheating is undetectable, participants still refrained from maximum dishonesty (see also Abeler et al., 2019; Gerlach et al., 2019). In another study, Gamliel and Peer (2013) found that a small chance of being checked actually increased cheating compared to a 0% chance. However, a study by Thielmann and Hilbig (2018) attempting to replicate this effect found that higher chances of being checked led to lower degrees of cheating. Study 1 offers another test of this relationship (Hypothesis 1), and extends our knowledge of the degree to which social influence affects dishonesty in collaborative cheating settings. Whereas earlier work implemented an ostensibly probabilistic mechanism for being checked, our studies highlight the role of the leader in checking. Players may thus attribute intentions to the leader, and anticipate their leader's actions accordingly. From this perspective, checking serves as a signal that may decrease or increase dishonesty, depending on context.

Table 1. Overview of Hypotheses in Both Studies.

Study	Manipulation	Number	Hypothesis (preregistered)	Support
1	Leader checking shifts from high to low frequency versus leader checking shifts from low to high frequency	1	Higher rates of checking will be negatively related to cheating	Yes
		2	Checking will have a negative effect on evaluations	Yes
		3	Players will send more money to a leader who shifts from high to low checking than to a leader who shifts from low to high checking	Yes
2	Cheating is collaborative (benefits leader) versus competitive (hurts leader)	1A	Competitive contexts will exhibit less cheating than collaborative contexts	Yes
		1B	The difference in cheating between contexts will decrease over time	Yes
		2A	Checking will have a negative effect on evaluations	Yes (except morality)
		2B	Checked cheats will have a more negative impact on evaluations than checked noncheats	Yes (except morality)
		2C	The interaction between checking and cheating will be stronger in collaborative than in competitive contexts	Yes (except morality)
		3	Players will send more money to a leader they evaluated as trustworthy	Yes

Note. Preregistered hypotheses for Studies 1 and 2, and whether they found support or not. Colors highlight different dependent variables: gray = cheating, blue = leader evaluations, yellow = trust game behavior.

In contexts that allow for collaborative cheating, that is, when leaders benefit from player misreports, checking is not financially incentivized and may therefore be interpreted as the leader caring about honest reporting, or being risk-averse to potential downstream costs of detection. Repeated checking thus implies that a leader does not want to collaborate in cheating. Repeated *non*checking presents a more complex picture (meaning that repeated checking is a stronger signal of unwillingness to cheat than repeated nonchecking is one of willingness to cheat). For the specific context of our study, where leaders do not have other tasks to attend to, and only have one player to work with, repeated nonchecking may be interpreted as one of two conflicting signals. First, nonchecking can be interpreted as the leader believing that the player's report is correct. Alternatively, repeated nonchecking by the leader, even when faced with "too good to be true" reports, may be interpreted as the leader's willingness to engage in collaborative cheating, or at least not caring about ensuring honesty in the task. Regardless, repeated nonchecking should generally increase player cheating.

If cheating is a competitive act that is to the detriment of the leader (as we investigate in Study 2), however, players may be less likely to cheat than in collaborative settings because they recognize that leaders are financially motivated to check. Nevertheless, as players learn the actual rate of checking, they will likely adapt their behavior, such that differences in cheating in collaborative vs. competitive contexts may diminish over time if checking rates are in fact similar (Hypothesis H1A and H1B).

Character Evaluations

Character evaluations can be decomposed into several dimensions. For the current studies, we focus on five dimensions: morality, trustworthiness, competence, risk aversion, and experienced psychological closeness.

First, moral character judgments are fundamental to person perception (Goodwin et al., 2014); it is important to investigate whether participants actually perceive their coconspirators as immoral, or instead morally disengage (Bandura, 1999; Fehr et al., 2019). Second, we highlight trustworthiness, though in some cases, it may be subsumed under general moral character evaluations (Lapsley & Lasky, 2001). We suggest that general moral and specific trustworthiness judgments may diverge in collaborative corruption settings: someone may judge their coconspirator as trustworthy in a particular collaborative context, but not as overall moral, given their dishonest behavior. Third, competence is

another primary dimension of person perception, involving assessments of skill, intelligence, or talent (Fiske et al., 2007). The present context allows us to test whether dishonest players will see leaders who check their performance, thus uncovering their lies, as competent or incompetent. Fourth, we measure perceptions of risk-aversion to gauge whether participants believe leader checking is due to a leader's being risk averse, or concerned about honesty. In this way, we can account for an alternative explanation of leader checking. Finally, while perceived closeness is not typically part of character evaluation, we include it here because it is an important indicator of the collaborative spirit of a group (Cole & Teboul, 2004). Closeness is also tied to moral behavior in several ways: others' immoral actions lead to social distancing (Skitka et al., 2005), while close others' immoral behavior specifically can lead to increased immoral behavior (Gino & Galinsky, 2012), and people who are close are likelier to engage in corruption (Akbari et al., 2020; Irlenbusch et al., 2020).

As with cheating, we expect character evaluations in our task to diverge considerably given the nature of the interaction. Collaborative corruption presents an inherent ambiguity regarding how coconspirators should evaluate each other, and character judgments may therefore become decoupled in meaningful ways (Bhattacharjee et al., 2013): that is, a coconspirator may be seen as competent but not moral. Yet, character judgments may also vary as a function of how participants prioritize competing moral concerns, such as loyalty and fairness (Graham et al., 2011; Hildreth & Anderson, 2018). We therefore also measure how participants value loyalty, authority, and fairness, which are fundamental to the rely-or-verify game (Graham et al., 2011). In particular, loyalty and authority, which are generally considered to be group binding norms (Graham et al., 2009), may be positively related to collaborative cheating and to positive leader evaluations. In general, because checking results hurts players financially, we predict a negative effect of leader checking on character evaluations. Moreover, we expect that players will judge their leader according to their own evaluation of the situation: when cheating is collaborative, and participants believe that cheating is acceptable, they will positively evaluate a leader who agrees with them (Hypothesis 2).

In the context of competitive cheating, however, we predict that patterns of evaluations differ, as checking behavior is now financially warranted from the leader's perspective. Thus, correct checks may be perceived as signs of competence. In sum, even though checks should still be perceived negatively overall, the degree to how negatively they will be experienced can be expected to differ based on whether the player cheated or not (Hypothesis 2A, 2B, and 2C).

Trust in a New Context

Finally, we investigate how collaborative cheating translates to trust behavior in a new context—an adapted trust game. Trusting a partner in the trust game makes most sense if one is convinced that the other person will honor the investment by sharing the profits. Behavior in the trust game has real and direct monetary consequences, thus serving as a useful behavioral measure of the actual established trust in the prior relationship. Positive effects of being a coconspirator in the previous game are expected to carry over to the trust game. In particular, investments should be higher for leaders who checked less often, and in settings in which cheating is collaborative rather than competitive (Hypothesis 3).

General Notes on Methods and Analysis

Both studies were preregistered (Study 1: https://osf.io/nsz5d/ and Study 2: https://osf.io/kxg7n). Unless otherwise noted, all hypotheses, measures, exclusions, and statistical analyses follow the preregistered procedure. Given space limitations, all additional analyses and collected measures are provided in the online materials (https://osf.io/p2esr/), which include a wider description of participant demographics, an extensive set of preregistered and exploratory robustness checks (e.g., models without control variables, as well as models including various personality traits), as well as a clear documentation of deviations from the preregistered protocol. All data analyses were performed using R (R Core Team, 2014). Logistic mixed effects (LoME) and linear mixed effects (LME) models were both built using the lme4 (Bates et al., 2015) and lmerTest (Kuznetsova et al., 2017) packages. Unless otherwise noted, all mixed effects models specify random intercepts on the participant level, and random slopes for round number. Bayesian models were fit using brms (Bürkner, 2017).

Study 1

In Study 1, we explore how the frequency of checking impacts players' cheating, evaluations of their leader, as well as how much they trust them in a new context. In two between-participants conditions, we manipulate checking rate such that in the high-to-low condition, leaders start out checking often, and then check less in the second half of the game. In the low-to-high condition, this pattern is reversed (see Table 1). Adopting this complex experimental setup carries several benefits. First, asking participants to evaluate and update their evaluations of their leader repeatedly allows us to test both how different rates of checking impact early evaluations and cheating behavior as a relationship between leaders and players begins to form, as well as the development of these variables, and their change given a change in the leader's strategy. Furthermore, we believe the iterative nature of the task to be critical for establishing even limited notions of relationships. Finally, leveraging repeated measures designs increases statistical power, accounting better for different sources of variance (Lindstrom & Bates, 1990;

Meteyard & Davies, 2020). Please note that to reduce complexity, analyses of updating evaluations are presented in the Supplement.

Method

Participants

We recruited 350 participants from Amazon Mechanical Turk (mTurk). Participants had to live in the United States, and have a >95% acceptance rate. We removed 42 participants because they failed the attention check, or made more than three attempts at the comprehension check. This left 308 participants (mean age = 38.06, *SD* = 11.38, 138 females, 168 males, and two participants specifying other), with 155 in the high-to-low condition, and 153 in the low-to-high condition. This satisfied our preregistered sample size goal of 150 participants per condition. Because we employed a novel paradigm with unknown effect sizes, and because 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, to allow testing for relatively small differences between our conditions (e.g., 80% power for a *t* test with $d = 0.32$).

Experimental Task and Procedure

The experiment was implemented in oTree (Chen et al., 2016), and structured in two stages. The first consisted of a rely-or-verify game (adapted from Levine & Schweitzer, 2015), and the second consisted of a trust game. The rely-or-verify game was 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 performed 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 was computerized, players did not roll physical dice, but rather clicked a button which triggered a video of a die being rolled (Kocher et al., 2018).

Having rolled the die, the player was able to report any number to the leader, who then decided 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 were calculated (see Figure 1 and Table 2).

In addition to the die roll task, each player evaluated their leader every five rounds, including a baseline evaluation before the game. This approach allowed us to track the updating of character evaluations, as players learned more about the behavior of their leader (Kim et al., 2020). Players were informed that the leader would not see these evaluations.

In general, leaders were programmed to never verify reports of 1s or 2s, as these reports were not likely to be dishonest. Moreover, leader behavior was manipulated in two conditions. In the high-to-low condition, the leader checked player reports with a 90% chance in the first 10 rounds. Then, in the next 10 rounds, the leader checked with only a 10% chance. This pattern of checking was reversed in the low-to-high condition.

Subsequently, players played a one-shot trust game with their leader. Each participant acted as the sender, thus allowing them to send any number of points they had earned in the previous game to their former leader, which would be tripled. Players were informed that their former leader may choose to send some points back but was under no obligation to do so.

After deciding how much money to send, participants filled out personality measures, moral foundations subscales for fairness, authority, as well as in-group/loyalty (Graham et al., 2011), the DOSPERT financial risk-taking subscale (Blais & Weber, 2006), which we measure to control for the potential impact of participants' own risk preferences on cheating behavior, and demographics. Finally, participants went through a funnel debrief. The experiment lasted around 22 min (see Supplement for more detailed methods).

Measures

Cheating was captured by 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 (Kocher et al., 2018). In addition, we also calculated mean cheating scores for each evaluation period, that is, the five rounds of the game prior to an evaluation. Thus, for each player, we calculated mean cheating in Rounds 1 to 5, 6 to 10, 11 to 15, and 16 to 20. This way, we could investigate the influence of cheating in an evaluation period on subsequent evaluations.

Evaluations of leader morality, trustworthiness, competence, closeness, and risk aversion were made on single-item 7-point Likert-type scale questions, ranging from −3 to 3, with the endpoints labeled "Not at all," and "Extremely" (e.g., "how trustworthy is your leader as a person?").

For the trust game, the amount of money participants sent was measured as the percentage of the maximum amount a participant could have sent (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. $\alpha = .50$), in-group/loyalty (std. $\alpha = .57$), and authority (std. $\alpha = .72$) (see Supplement for robustness analyses using exploratory factor model scores).

Regarding the DOSPERT financial risk-taking subscale, we had preregistered a parallel analysis, due to recent

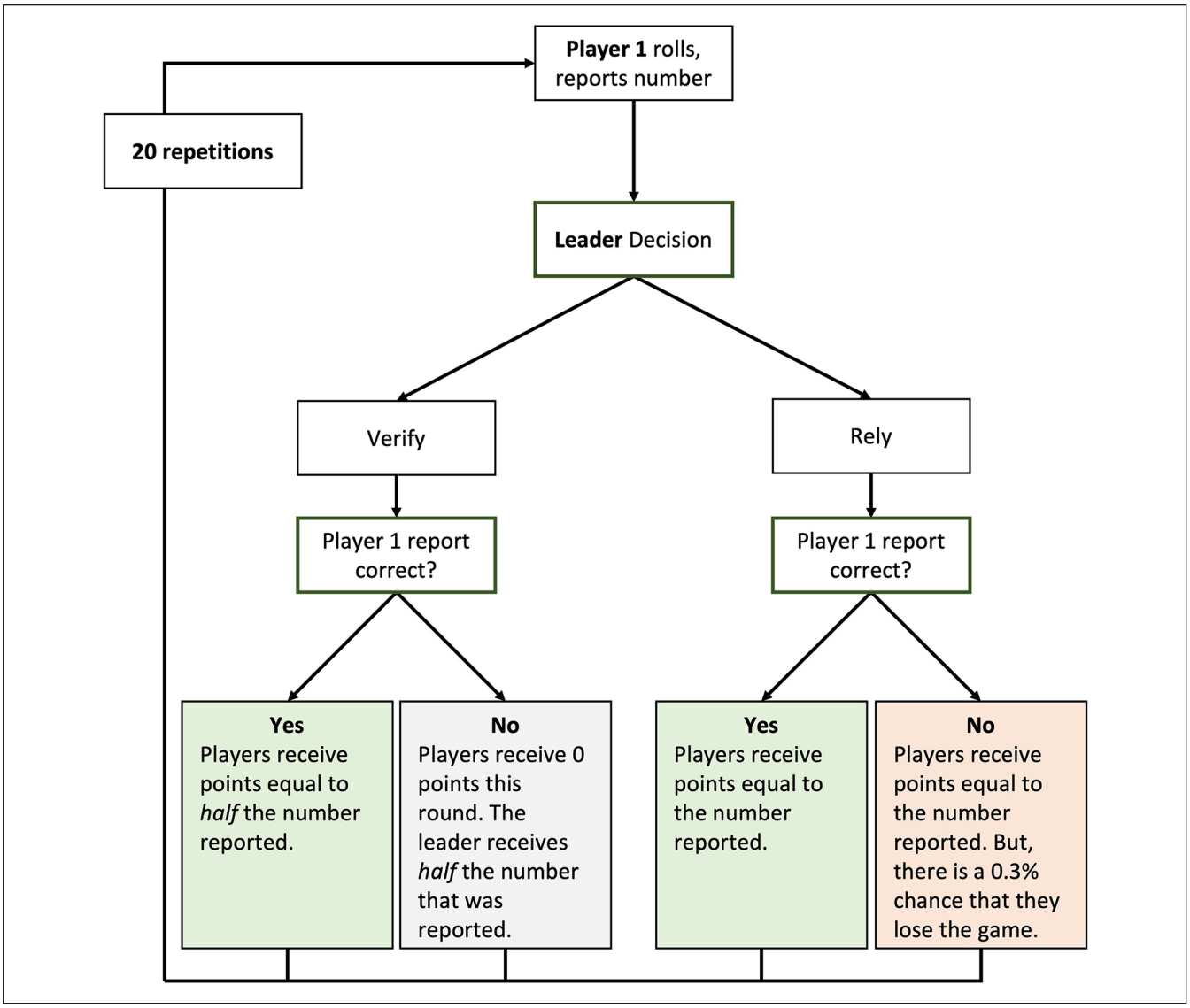

Figure 1. Structure of the rely-or-verify game as presented to the study participants.
Note. Participants were rewarded with points on each round. These points accumulated over the course of the game, and were converted to USD at the end of the study (1 point = US$0.05). Note that losing the game means the game ends in the current round, and all players (player and leader) lose all points they made thus far.

Table 2. Payoffs for Players and Leaders Given Different Behavior.

Subordinate behavior	Leader behavior	Player pay	Leader pay
No cheat	Verify	Reported number/2	Reported number/2
Cheat	Verify	0	Reported number/2
No cheat	Rely	Reported number	Reported number
Cheat	Rely	Reported number*	Reported number*

Note. Payoffs for players and leaders were determined by the interaction of players and leaders behavior. Payoffs were given in points, which were converted to USD at the end of the Study (1 point = US$0.05).
*In this case, there was a 0.3% chance that the team would lose the game, meaning that they would 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 (this low percentage losing chance is also often employed in bribery games, for example, Abbink et al., 2002; Bodenschatz & Irlenbusch, 2019; cf. also Cullinan & Sutton, 2002). Importantly, even though participants were informed that the chance of losing the game at 0.3%, we set the actual probability of losing the game to 0.

Table 3. Overview of Descriptive Statistics for Study 1.

Behavior	*M* (%)	*SD* (%)
Overall cheating across conditions	11.67	32.10
Cheating given the rolled number		
1	25.32	43.52
2	16.64	37.26
3	10.98	31.28
4	5.60	23.00
5	2.27	14.92
Overall leader verifying (computer determined)	37.22	48.34
Amount of caught cheats	40.83	49.19
Overall amount sent in trust game	36.96	36.73

Note. Main descriptive statistics for Study 1. As these values (except for money sent in trust game, see below) did not differ significantly between conditions, we collapse them here. *SD* = standard deviation.

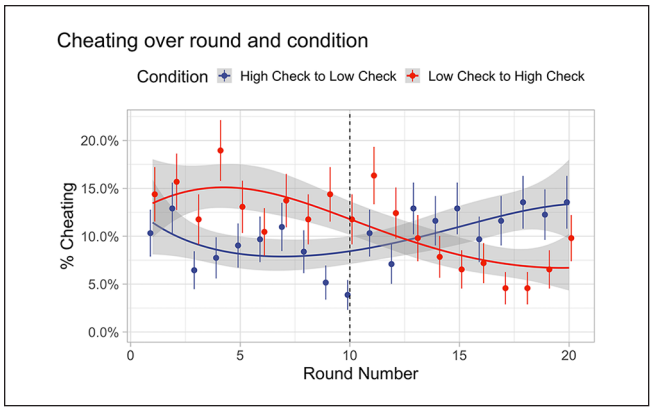

Figure 2. Likelihood of cheating by round and condition in Study 1.
Note. Lines represent b-spline curves modeling player cheating occurring during the game, with three degrees of freedom. The dotted line at round 10 indicates the point when leader checking behavior switched. Gray ribbons indicate 95% confidence intervals for spline model. Whiskers represent standard error around mean estimates..

discussions about the DOSPERT's factor structure (Highhouse et al., 2017). This analysis indicated a two-factor solution. A subsequent exploratory factor analysis (EFA) revealed two factors, one related to betting (Items 1, 3, and 5 on the financial risk-taking scale; std. $\alpha = .90$), and the other related to investment (Items 2, 4, and 6; std. $\alpha = .81$), for which we calculated sum scores.

Finally, we also asked players to evaluate their own perceived control, as well as their leader's control over the outcome in the game ("How much control did you [your leader] have over the outcome of this task?," both rated on a 7-point Likert-type scale). Furthermore, we asked participants to evaluate their own, as well as the estimated social standing of their leader, using the social ladder measure (Adler et al., 2000). Both perceived control and perceived standing in society were used as indirect manipulation checks, allowing us to gauge whether players indeed saw their leaders to be more in control, and to be of somewhat higher social status, as one might expect of leaders.

Results

Descriptive Measures and Manipulation Checks

Starting with 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 (49%) 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 data set or only cheaters were comparable (see Table S3, Figure S3, and Supplement). A further description of basic properties of behavior can be found in Table 3.

As manipulation checks, 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, participants did perceive their leader as generally more powerful than themselves.

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 an LoME model specifying an interaction between condition and the cubic polynomial of round number, controlling for the number participants had actually rolled (as this affects the potential gain a participant can achieve by cheating), as well as participants' scores on moral foundations and financial risk aversion.

In line with our expectations, this model showed significant interactions for round number and condition, such that participants in the low-to-high condition cheated more in the first half of the game than participants in the high-to-low condition. This pattern reversed following the switch in leader behavior (see Figure 2). 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 ($b = -1.03$, 95% confidence interval [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 ($b = 0.54$ [0.18, 0.91], $p < .001$). Full model results can be found in Table S1.

In sum, Hypothesis 1 finds support: the more a leader checked, the less participants cheated. This was also

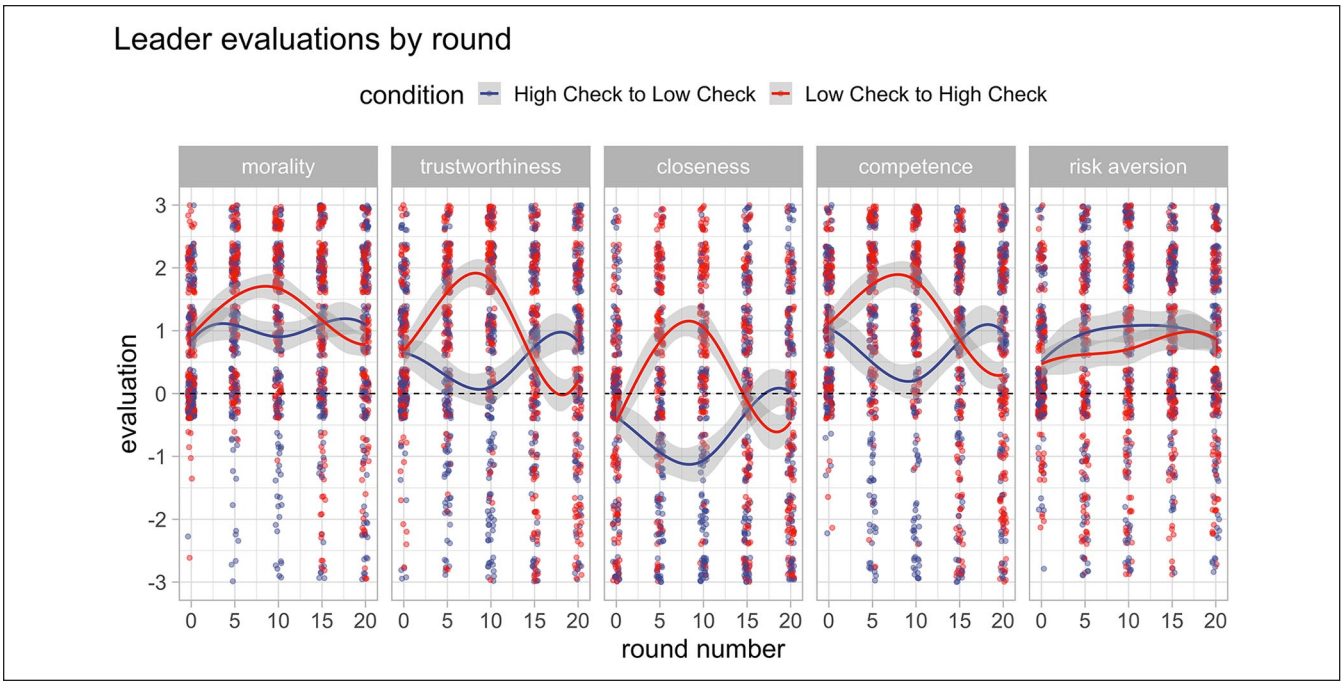

Figure 3. Leader evaluations by round number and condition.
Note. Leader evaluations occurred every five rounds, with a baseline evaluation at round 0. Lines represent b-splines with knots at round 10, and three degrees of freedom. Gray ribbons indicate 95% confidence intervals for spline model.

supported by a follow-up, and arguably more direct, analysis that modeled the effect of having been checked in the prior round on cheating in the current round, showing a significant and negative effect of checking on cheating (odds ratio = 0.50 [0.38, 0.67], $p < .001$; in other words, having been checked in the prior round decreased the odds of cheating by 0.50; see Table S2 for full model results).

Interestingly, there was no long-term impact of being checked: when a leader stopped checking, cheating increased quickly to similar degrees when leaders started out not checking. Thus, there was no overall difference in the amount of cheating between conditions (Wilcox $W = 11568$, $p = .69$, $d = 0.064$).

How Does Checking Influence Leader Evaluations?

To analyze Hypothesis 2, that checking has adverse effects on leader evaluations, we built LME models for each evaluation dimension: morality, trustworthiness, closeness, competence, and risk aversion. 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 five rounds leading up to an evaluation), further controlling for personality (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.

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 ($b = -29.37$ [−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 ($b = -0.54$ [−1.07, −0.01], $p = .045$). Yet, those participants that cheated often evaluated leaders who switched to nonchecking in the high-to-low condition as more moral than leaders who switched to checking in the low-to-high condition ($b = 14.55$ [−0.42, 29.53], $p = .057$). However, these effects are only marginally significant, and await further investigation.

Because of the complex relationship emerging between checking and cheating, we analyzed their relationship in an exploratory way. Specifically, we analyzed whether the type of checking matters for evaluations, that is, whether a check or a noncheck was correct or not. This analysis finds further support for the notion that collaborative cheating can enhance evaluations of coconspirators compared to people refusing to engage in such practices, and that this effect is especially strong for closeness and trustworthiness ratings (compared to morality; cf. Supplement, Tables S4 and S5, and Figure S1).

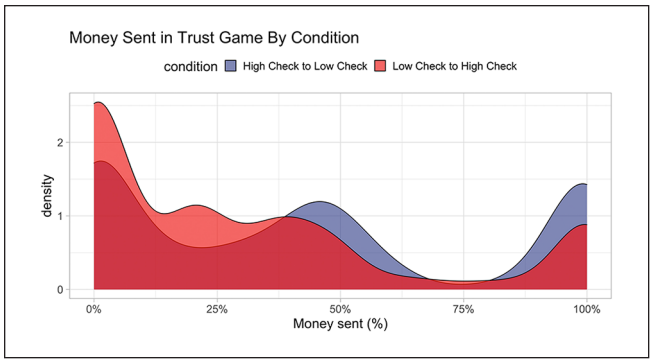

Figure 4. Density plot of money sent by players to their leaders in the trust game, by condition.
Note. Contributions are measured as percentage of the maximum possible amount.

Finally, we turn to Hypothesis 3, which predicted that participants in the high-to-low condition would send the leader more money in the trust game than participants in the low-to-high condition. To model behavior in the trust game, we analyzed the percent of the endowment 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 0% (28.6% of participants) or 100% (16.2 % of participants) of the points they had earned during the previous game to their partner (see Figure 4). ZOIB models are ideally suited for modeling 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 modeled these 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).

The model revealed an effect of condition both for mean and conditional one inflation, such that participants in the high-to-low condition entrusted overall more money to the leader ($b_{mean} = -0.08$, 95% *credible interval* = [−0.16, −0.01]), and were also more likely to send everything rather than nothing ($b_{coi} = -0.15$ [−0.29, −0.02], see also Figure 4). Furthermore, trustworthiness evaluations predicted conditional one inflation, such that participants rating the leader as more trustworthy were more likely to send everything rather than nothing ($b_{coi} = 0.09$ [0.04, 0.15]). These results support Hypothesis 3.

In addition, participants who scored higher on Authority gave less money to the leader ($b_{mean} = -0.08$ [−0.13, −0.03]). Further predictors of conditional one inflation were the amount of cheating in the previous game ($b_{coi} = 0.46$ [0.01, 0.75]), leader risk aversion judgments ($b_{coi} = 0.09$ [0.03, 0.15]), and participant investment risk-taking propensity ($b_{coi} = 0.16$ [0.05, 0.27]). Full model results are presented in Table S7.

In sum, participants sent more money to a leader that stopped rather than started checking, falling in line with Hypothesis 3. Furthermore, 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.

Study 2

Study 1 investigated how manipulating the degree of leader checking affects collaborative cheating, leader evaluations, and subsequent trust. 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 nonchecking of cheating enhanced trustworthiness ratings and perceived psychological closeness, but had no negative impact on morality and competence ratings. This 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.

Study 2 has two aims. First, we seek to conceptually replicate the core findings of Study 1. As we had analyzed the precise interplay of checks and cheats on character evaluations only in an exploratory analysis, Study 2 entails a conceptual replication of Study 1, by testing the link between cheating, checking, and leader evaluations more precisely.

Second, we aim to extend our findings by examining how changing the context of cheating to be collaborative vs. competitive (while keeping leader checking at a fixed rate) affects behavior and evaluations. As detailed in the introduction, we deem the factor of whether cheating is collaborative versus competitive to be highly important in shaping behavior. Comparing how these two contexts influence behavior and the formation of relationships is crucial for better understanding the particular nature of collaborative corruption. We thus manipulate the payoff structure of the game in two between-participants conditions (aligned vs. competing payoffs), which change the overall context of the game to allow for collaborative versus competitive cheating. While inflating outcomes in the aligned payoffs condition also benefits the leader, this behavior is to the detriment of the leader in the competing payoffs condition. Importantly, whereas

Table 4. Payoffs for Players and Leaders for Different Conditions.

		Aligned payoffs		Competing payoffs	
Player behavior	Leader behavior	Player pay	Leader pay	Player pay	Leader pay[a]
No cheat	Verify	Reported number/2	Reported number/2	Reported number/2	Reported number/2
Cheat	Verify	0	0	0	Actual number/2
No cheat	Rely	Reported number	Reported number	Reported number	Reported Number
Cheat	Rely	Reported number	Reported number	Reported number	0

Note. Payoffs for players and leaders, depending on condition, and behavior of leaders and players, in points. The different payoff structures were designed such they would only diverge in cases when the player cheated.
[a]Leaders received an additional bonus point after each round, no matter the outcome, to offset an asymmetry in total pay that would have otherwise resulted in leaders earning considerably less than players.

leaders do not have a financial reason to check players' reports when payoffs are aligned (similar to Study 1), they do when payoffs diverge. In relation to these different payoff structures, we again investigate cheating behavior, leader evaluations, as well as subsequent trust (see Table 1 for hypotheses).

Method

Participants

We recruited 461 participants from mTurk, who were living in the United States and had a $>$ 95% acceptance rate. Forty-five participants were excluded as they either submitted more than 10 incorrect answers in the comprehension check, or failed an attention check. This left 416 participants (mean age = 37.56, SD = 11.26, 171 females, 240 males, and five participants specifying other), with 210 participants in the competing payoffs condition, and 206 in the aligned payoffs condition. This satisfied our preregistered sample size goal of 200 participants per condition. Sample size was calculated to achieve 80% power for a moderate (b = 0.5) three-way interaction specified in Hypothesis 2C. We used the simglm package (LeBeau, 2019) to simulate data, using effect sizes that were informed from a pilot study (see preregistration).

Experimental Procedure

Like Study 1, this study was implemented in oTree (Chen et al., 2016), and followed a similar two-part structure: a rely-or-verify game followed by a one-shot trust game. All participants were instructed about the design and payoff options in the study, and were required to pass a number of comprehension checks to begin the rely-or-verify game.

Studies 1 and 2 differ in three ways: First, while leaders were again played by the computer, leader checking rate was not manipulated in this study. Instead, the rate was fixed at 25% chance for die roll reports higher than 3. Reports of 3s, 2s, or 1s were never checked, to enhance believability.

Second, the rely-or-verify game lasted for only 10 rounds (not 20 rounds), and participants had the opportunity to evaluate their leader after every round. This approach allowed us to investigate the change of leader evaluations in a more fine-grained manner in Study 2, increasing power as a result.

Third, Study 2 manipulated payoffs in two conditions that were randomly assigned between participants (see Table 4). In the aligned payoffs condition, payoffs were similar as in Study 1, with the notable differences being that both leaders and players now received the same payoff every round (including leaders verifying a cheated report), and we removed the chance for outside detection (i.e., losing the game) from participant instructions. This was done to make the competing payoffs condition more comparable, and to completely align player and leader payoff. Notably, a consequence of this change is the perceived incentive structure for leader checking in the aligned condition: leader checking is never financially warranted in this condition.

In the competing payoffs condition, payoffs for leaders and players diverged whenever a player cheated (see Table 4). With this, leaders were incentivized to verify whenever they suspected a cheat, but to rely when they presumed no cheat. Importantly, players (participants) were informed that the leaders would not learn about their payout until after the completion of the entire study. Thus, leaders would not be able to infer that a player had cheated in the previous round unless they checked.

Measures

Study 2 employed similar measures as Study 1. In particular, we recorded actual and reported die roll results, and leader checks for each round. Leader evaluations occurred after each round on the same dimensions as in Study 1 (morality, trustworthiness, competence, closeness, and risk aversion).

For personality measures, we again collected DOSPERT (Blais & Weber, 2006) scores for financial risk-taking, and moral foundation scores for in-group/loyalty, authority, and fairness (Graham et al., 2011). As in Study 1, both the DOSPERT and moral foundation scales exhibited poor reliability measures. Thus, following preregistered protocol, we computed separate financial risk-taking scores for betting

Table 5. Overview of Descriptive Statistics for Study 2.

Behavior	Across conditions	Aligned incentives	Competing incentives
Mean cheating over all rounds	34.56% (47.56%)	42.24% (49.41%)	26.78% (44.29%)
Cheating given the rolled number			
1	50.87% (50.02%)	60.13% (49.02%)	41.48% (49.32%)
2	46.31% (49.89%)	53.45% (49.93%)	39.08% (48.85%)
3	32.21% (46.75%)	41.38% (49.3%)	22.93% (42.08%)
4	20.07% (40.07%)	26.72% (44.3%)	13.32% (34.01%)
5	12.15% (32.7%)	16.81% (37.48%)	7.42% (26.27%)
Leader verifying rate (computer determined)	16.71% (12.02%)	17.48% (12.6%)	15.92% (11.39%)
Caught cheats	24.03% (42.74%)	23.36% (42.34%)	25.09% (43.39%)
Contributions in trust game	44.63% (34.78%)	47.39% (34.68%)	41.81% (34.75%)

Note. Main descriptive statistics for Study 2. Values are mean percentages. Standard deviations in parentheses.

and investment frames, and computed factor scores for the moral foundations items, which indicated a slightly different factor structure than reported in the original study by Graham et al. (2011). We term these factors traditionalism (mostly authority items), in-group, and fairness (see Supplement for more detail on these factors, and additional robustness analyses).

In addition to these variables, after the two tasks, we asked participants to rate the perceived ethicality of cheating in the rely-or-verify task ("How immoral do you think it is to report an inflated die roll in this study?" on a 7-point Likert-type scale, with endpoints labeled "completely moral"—"completely immoral"). As in Study 1, we also asked participants to indicate their perception of how much control they and their leader had over the outcome of the rely-or-verify game, as well as their and their leader's perceived status in society.

Results

Descriptive Measures and Manipulation Checks

An overview of descriptive statistics is presented in Table 5. Looking at the general distribution in player cheating and leader checking, we find that participants cheated in 34.56% of rounds in which cheating was possible, and got checked in 16.71% of rounds. In this study, 74.2% of participants cheated at least once.

Inspecting the perceived morality of cheating in the rely-or-verify task, we find that participants assigned moderate immorality to cheating ($M = 4.23$, $SD = 1.81$). Importantly, the perceived immorality of cheating did not differ between conditions, $t(413.95) = -1.21$, $p = .23$, $d = -0.12$, even though cheating in the competing payoffs condition resulted in a clear victim, namely, the leader, whereas cheating in the aligned payoffs condition did not negatively impact the leader.

Similar to Study 1, players perceived themselves to be less in control over the outcome of the rely-or-verify game than their leaders, $t(415) = 7.59$, $p < .001$, $d = 0.37$, and as generally lower status, $t(415) = -5.99$, $p < .001$, $d = -0.29$. There were no significant differences between perceived status or control between conditions.

Differences in Cheating Between Conditions

Recall our two hypotheses regarding player cheating. First, we hypothesized a main effect of condition, such that participants would cheat less in the competing payoffs condition than in the aligned payoffs condition. Second, we hypothesized that this difference in cheating between conditions would decrease as the game progressed. To investigate these hypotheses, we built an LoME model predicting cheating based on the interaction of round number and condition, controlling for what players had actually rolled, whether players had been checked in the previous round, and personality measures. The model specified random slopes for the actually rolled number.

Supporting our hypotheses, we observed a main effect of condition, such that participants cheated less in the competing payoffs condition (*log odds* $= -2.00$ [-2.74, -1.27], $p < .001$). This main effect was also subject to the moderating influence of round number, in that the difference between conditions decreased over time (*log odds* $= 0.12$ [0.042, 0.21]), $p = .0031$). In addition, an exploratory analysis indicated that, at the end of the game, participants in the aligned payoffs condition still cheated more than participants in the competing payoffs condition (*log odds* $= -0.63$ [-1.09, -0.16], $p = .0085$). Keeping in mind the above result that the absolute number of cheaters (i.e., people who cheat at least once) is the same in both conditions, this implies that cheaters cheat more often in the aligned than the competing payoffs condition. Thus, both Hypothesis 1A and 1B are supported by our model (see also Figure 5; full model results in Supplement).

Leader Evaluations

Turning to leader evaluations, we replicated the negative effect of leader checking on evaluations from Study 1

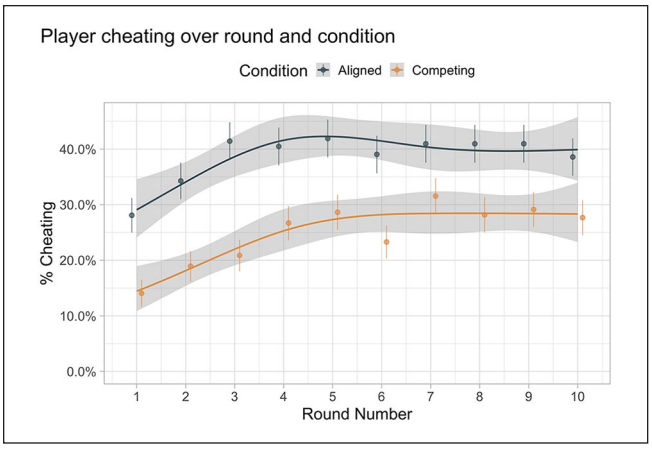

Figure 5. Percentage of players cheating by round and condition.
Note. Smoothed lines represent natural splines with three degrees of freedom. Whiskers represent standard error around mean estimates.

(Hypothesis 2A). Leader checks led to lower evaluations of competence, trustworthiness, and closeness, but not morality, which showed a much smaller trend in the same direction ($b = -0.10$ [−0.21, 0.01], $p = .064$. Full results are given in Table S8).

Similarly, we found support for Hypothesis 2B: independent of condition, checked cheats led to more negative evaluations than checked noncheats, for all dimensions except morality. Participants judged leaders as *more* moral following checked cheats versus checked noncheats ($b = 0.15$ [−0.003, 0.31], $p = .055$). When figuring in condition differences, participants judged leader checking as more moral in the aligned versus the competing payoffs condition ($b = -0.24$ [−0.39, −0.09], $p = .0018$). Thus, participants were more likely to interpret leader checking as a sign of the leader's moral conviction in the aligned payoffs condition than in the competing payoffs condition.

The two-way interaction between checks and cheats of Hypothesis 2B was further qualified by condition in the predicted three-way interaction specified in Hypothesis 2C. More specifically, the interaction between checking and cheating was stronger in the aligned than in the competing payoffs condition. In other words, participants who cheated and were caught evaluated their leaders worse in the aligned payoffs condition in terms of trustworthiness, closeness, and competence, than participants in the competing payoffs condition. Thus, the collaborative nature of cheating in the aligned payoffs condition led participants to evaluate leaders who checked their reports (when they had cheated) harsher than participants in the competing payoffs condition, who had cheated to the detriment of the leader. Participants were therefore highly attuned to the underlying motivations for leaders to check, adjusting their evaluations accordingly (see Figure 6 illustrating the three-way interactions, cf. Table S8).

Trust Game Behavior

Finally, we investigated player behavior in the trust game following the rely-or-verify game, testing Hypothesis 3 that players' evaluations of their leader's trustworthiness would positively predict money sent to the leader. To do so, we used a ZOIB model predicting money sent by a players' final trustworthiness evaluation of their leader, their degree of cheating during the rely-or-verify game, condition, and personality scores.

In line with Hypothesis 3, participants who rated the leader as more trustworthy invested more money ($b_{coi} = 0.08$, 95% *credible interval* = [0.03, 0.15]). Similarly, this analysis showed a mean effect of condition ($b_{mean} = -0.07$ [−0.13, −0.02]), and a conditional one inflation of condition ($b_{coi} = -0.21$ [−0.39, −0.02]), with participants in the competing payoffs condition sending less money. Another factor influencing trust game behavior was the degree to which players had cheated, which was negatively related to money sent ($b_{mean} = -0.20$ [−0.30, −0.11]; $b_{coi} = -0.39$ [−0.65, −0.09]). Notably, this effect is the opposite of what we found in Study 1, where the more a player had cheated in the rely-or-verify game, the more they invested in the trust game (see S9 for full model results, and explorations of this result).

General Discussion

In two studies, we find support for the idea that collaborative cheating in hierarchical teams leads to close relations among coconspirators. In settings that allow for collaborative corruption, letting cheating go undetected not only increased the frequency of dishonest behavior, but also had positive effects on how co-conspiring leaders were evaluated in terms of competence, trustworthiness, and experienced closeness. This trustworthiness bonus translated to behavior in new settings, leading participants to entrust their coconspirators with more money (see Figure 7 for a summary of our findings).

Regarding cheating behavior, we found that people were highly sensitive to the amount of checking, in line with prior research (Thielmann & Hilbig, 2018). Extending previous results, we found that participants were strongly affected by the leader's incentive either to check them or turn a blind eye. When cheating was to the leader's detriment, participants anticipated that leaders would be motivated to check them. Consequently, they cheated less compared to when cheating was collaborative, and only gradually adapted their behavior to the actual checking rate, and not to the full extent. This finding is important for understanding the emergence of collaborative corruption: when people realize that a certain situation allows for collaborative rather than competitive cheating, they will be more prone to engage in dishonest behavior.

One potential extension is to study how subordinates perceive leader behavior in a more "neutral" setting, where a leader's pay is entirely independent from subordinate

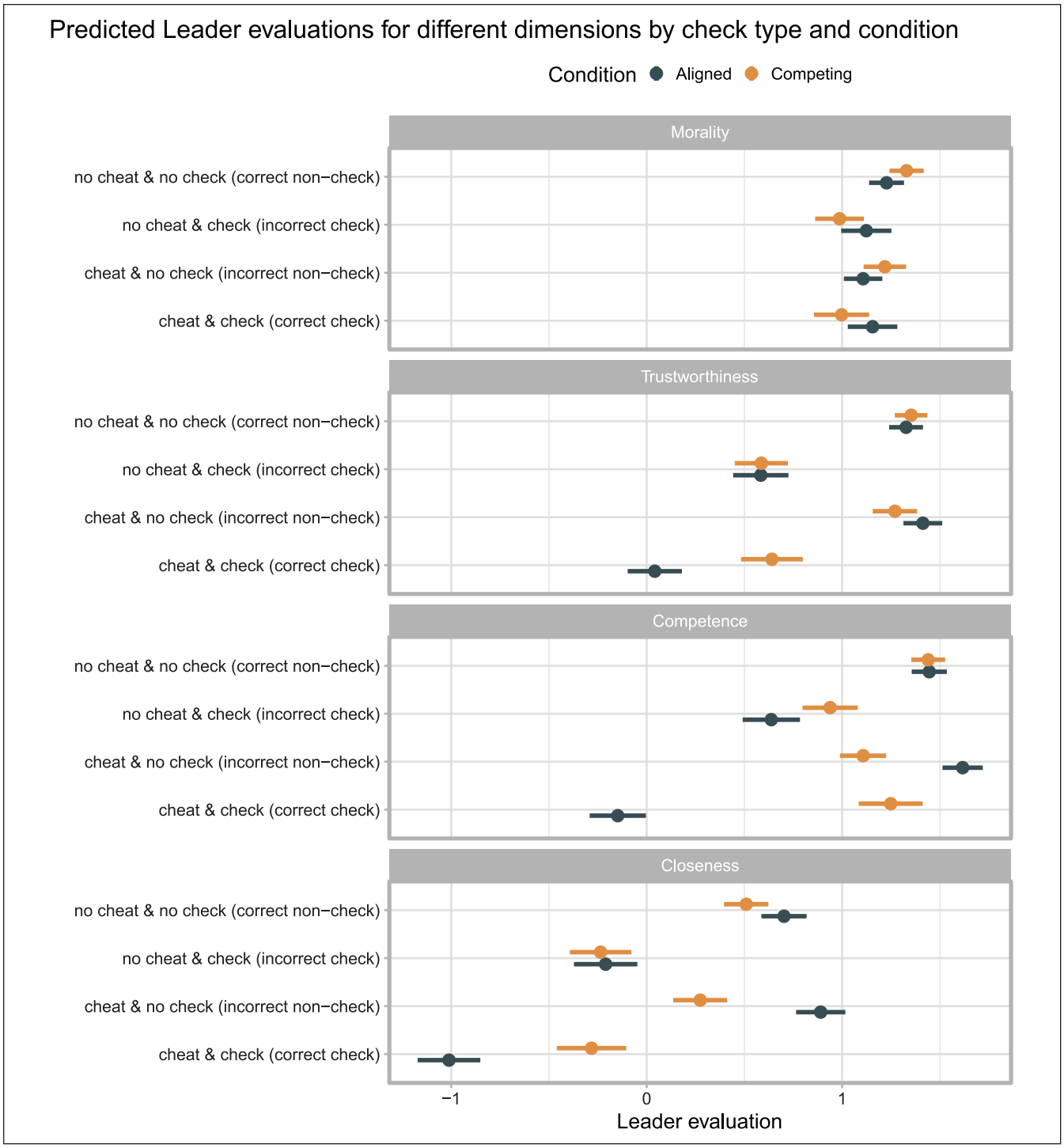

Figure 6. Predicted leader evaluations based on condition and check type in a given round (marginal effects).
Note. Evaluations were made on a 7-point Likert-type scale ranging from −3 to 3. This plot illustrates the three-way interactions for leader evaluations based on condition, cheating, and checking behavior (compare especially the differences between conditions in cases when the participant cheated, and the leader either checked or did not check). Plots are facetted by evaluation dimension. Dots represent estimated evaluations as a function of the interaction between checking, cheating, and condition, whiskers represent standard error (see Supplement for model details).

cheating (as an independent auditor's would be). We note, however, that it is not clear that subordinates would perceive such a leader as neutral in this setting. Instead, subordinates may construe a collaborative or anti-collaborative stance directly from the leader's verifying behavior. This points to an exciting area of future research, integrating recent insights

Figure 7. Summary of main findings.

of situation perception and specifically interdependence construal (e.g., Columbus & Molho, 2022) into the study of collaborative cheating.

Looking at evaluations of leaders as potential coconspirators, our study suggests that, at least in the context of rely-or-verify situations, collaborative corruption is similar to other types of collaborations, and has largely positive effects on person perception (Zelmer, 2003). In fact, letting cheating go unchecked was perceived as particularly positive (and, conversely, checking cheats was seen as especially negative) when the benefits of the cheating were shared. This effect also held for competence evaluations, indicating that participants did not believe that leaders were placing undeserved trust in them, but instead identified ignoring dishonest reports as signs of their leaders' competence.

Zooming in on the various evaluation dimensions, we make several observations. First, morality evaluations showed the lowest volatility, with participants largely maintaining their judgments in response to the interaction of cheating and checking. Likewise, morality evaluations did not exhibit the same sensitivity to personal payoff structures as the other dimensions of interest did. One possible explanation for this pattern may be that participants saw checking behavior as less diagnostic of the leader's overall morality than their trustworthiness. Interpreted this way, participants may have decoupled their more holistic moral character judgment of the leader from their more task-specific judgments of competence or even trustworthiness (Bhattacharjee et al., 2013). This divergence between morality and trustworthiness deserves additional exploration. One question that

remains is how people conceive of and maintain context-specific trustworthiness judgments, especially as trust established from emergent collaborative corruption in the initial game translated to the subsequent trust game. Future studies could thus compare trust built upon "positive" collaborations to trust established via collaborative corruption.

Relatedly, psychological closeness not only aligned with trustworthiness evaluations but also showed the strongest sensitivity toward leader checking behavior, and was also strongly affected by whether cheating was collaborative or competitive. Perceived closeness is a fundamental factor for moral psychology: people will often engage in bad behavior if they observe close others' engaging in bad behavior first (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 subjective experience of closeness, with especially pernicious effects. In a cycle of socially reinforced immorality, people who cheat together feel closer to each other, and are then even more influenced by their partner's bad behavior. Future work could probe this cycle by manipulating perceived closeness via group membership. An open question is how checking behavior and payoff structures affect not only emergent collaborations as in the present work, but also how collaborations might unfold when people who already know and like each other come into a context that allows for collaborative corruption.

Finally, the relationship building effects of emergent collaborative corruption between coconspirators is underscored by participants' behavior in the trust game: positive evaluations of nonchecking leaders translated to trust behavior in a new, noncorrupt kind of collaboration, where collaborating does not cause third-party harm, and where participants were made aware of the trust game as a new and risky situation (also indicated by the fact that many participants chose to send nothing). Our findings suggest that trust built on corrupt collaborations can carry over to new contexts. Nevertheless, future research should consider the different ways in which cheating collaboratively extends beyond immediate contexts and underwrites phenomena, such as "honor among thieves."

One important question this research raises is how in-group, localized morals square against more holistic norms, such as those of fairness (also sometimes referred to as group binding and individualizing norms (Graham et al., 2009, 2011). From this perspective, one primary conflict in collaborative corruption settings is that between loyalty and honesty. As has been documented, loyalty concerns (i.e., performing actions that benefit the in-group) can trump fairness concerns toward wider society (e.g., Hildreth & Anderson, 2018; Waytz et al., 2013). In the present work, we find limited evidence for this connection, as participants placing more value on in-group/loyalty and authority also generally evaluated their leader more positively (see Supplement). Yet, we did not explicitly probe any feelings of commitment and in-group formation that might be established as participants completed the task. Future work should therefore more explicitly explore the potential role that perceived obligation plays in the development of collaborative corruption. One interesting direction is to vary whom the cheating is harming. In our experiments, an outside party (i.e., the funders of the research project) suffered the costs of cheating. However, in some forms of collaborative corruption, other members of the same organization are victims of dishonesty. These latter constellations might lead to interesting conflicts in obligations toward different kinds or levels of in-groups, and could affect leader evaluations.

One general limitation of our studies is that we focused only on evaluations of leaders by subordinates. Thus, future work should test whether relationship-building qualities also emerge for the other direction, and whether this effect generalizes to other contexts. Regarding the directionality of the effect, prior research has shown that leaders and subordinates evaluate each other positively so long as they share moral values (Egorov et al., 2020), suggesting that the present findings may generalize to some extent. Future work may uncover other differences, however. For instance, leaders may be less inclined to establish relationships with unethical subordinates than subordinates with unethical leaders, as leaders are less dependent on subordinates.

As for nonhierarchical relationships (e.g., same-status employees engaging in dishonest conduct), we predict that our results will generalize to these contexts as well, at least to some degree. The relational aspect of collaborative cheating has been identified as a core element in its theoretical and empirical account (Weisel & Shalvi, 2015), and is thus not unique to our specific implementation. We expect any collaboration in which people work together toward a shared goal, in a corrupt or noncorrupt manner, to generally yield positive relationship-building outcomes (Cole & Teboul, 2004).

Still, our studies characterize collaboration in a conceptually "thin" way, lacking many features of real-world, "thick" sense collaboration, like spoken or written communication, joint action, or the co-creation of goals. Collaboration in the rely-or-verify game requires only minimal coordination between players and leaders, such that players cheat and leaders do not check. Even more, because our studies rely on simulated leader behavior, they do not capture genuine interactions between two humans. Instead, with our design, we chose to concentrate on specific, theoretically interesting patterns of interaction (e.g., what happens when leaders decide to drastically change their behavior). It is thus important that future studies recruit participants for both leader and player roles to study dyadic and group-level adaption of behavior as they play out in more naturalistic settings. On a more general level, we believe that incorporating "thick" collaboration is an important frontier in the study of collaborative corruption.

Conclusion

When people collaborate successfully, they form strong bonds of trust. We show that close ties can also develop for

corrupt collaborations. In the present work, leaders who functioned as coconspirators, turning a blind eye to cheating, were evaluated positively (i.e., closer and more trustworthy), whereas leaders who spoiled opportunities to cheat by checking reports were evaluated negatively. These findings suggest that collaborative corruption may be especially robust and require strong external regulation to be curtailed.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article. They thank Joshua Hirschfeld-Kroen, Kevin Jiang, Gordon Kraft-Todd, Justin Martin, Ryan McManus, and BoKyung Park for their helpful feedback and comments, Savhannah Schulz for creative development of Figure 7, and the John Templeton Foundation as well as the Boston College Virtue Project for financial support.

ORCID iD


Simon Tobias Karg https://orcid.org/0000-0002-7153-7169

Supplemental Material

Supplemental material is available online with this article.

References

Abbink, K. (2004). Staff rotation as an anti-corruption policy: An experimental study. *European Journal of Political Economy*, *20*(4), 887–906. https://doi.org/10.1016/j.ejpoleco.2003.10.008

Abbink, K., Irlenbusch, B., & Renner, E. (2002). An experimental bribery game. *Journal of Law, Economics, and Organization*, *18*(2), 428–454. https://doi.org/10.1093/jleo/18.2.428

Abeler, J., Nosenzo, D., & Raymond, C. (2019). Preferences for truth-telling. *Econometrica*, *87*(4), 1115–1153. https://doi.org/10.3982/ecta14673

Adler, N. E., Epel, E. S., Castellazzo, G., & Ickovics, J. R. (2000). Relationship of subjective and objective social status with psychological and physiological functioning: Preliminary data in healthy white women. *Health Psychology*, *19*(6), 586–592. https://doi.org/10.1037/0278-6133.19.6.586

Akbari, M., Bahrami-Rad, D., Kimbrough, E. O., Romero, P. P., & Alhosseini, S. (2020). An experimental study of kin and ethnic favoritism. *Economic Inquiry*, *58*(4), 1795–1812. https://doi.org/10.1111/ecin.12917

Bandura, A. (1999). Moral disengagement in the perpetration of inhumanities. *Personality and Social Psychology Review*, *3*(3), 193–209. https://doi.org/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), 1–48. https://doi.org/10.18637/jss.v067.i01

Bhattacharjee, A., Berman, J. Z., & Reed, A. (2013, April). 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. https://doi.org/10.1086/667786

Blais, A. R., & Weber, E. U. (2006, July). A Domain-Specific Risk-Taking (DOSPERT) scale for adult populations. *Judgment and Decision Making Journal*, *1*(1), 33–47. https://doi.org/10.1037/t13084-000

Bodenschatz, A., & Irlenbusch, B. (2019). Do two bribe less than one?–An experimental study on the four-eyes-principle. *Applied Economics Letters*, *26*(3), 191–195.

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Chen, D. L., Schonger, M., & Wickens, C. (2016, March 1). oTree—An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, *9*, 88–97. https://doi.org/10.1016/j.jbef.2015.12.001

Cole, T., & Teboul, J. C. B. (2004). Non-zero-sum collaboration, reciprocity, and the preference for similarity: Developing an adaptive model of close relational functioning. *Personal Relationships*, *11*(2), 135–160. https://doi.org/10.1111/j.1475-6811.2004.00075.x

Columbus, S., & Molho, C. (2022, February 1). Subjective interdependence and prosocial behaviour. *Current Opinion in Psychology*, *43*, 226–231. https://doi.org/10.1016/j.copsyc.2021.07.022

Cullinan, C. P., & Sutton, S. G. (2002). Defrauding the public interest: A critical examination of reengineered audit processes and the likelihood of detecting fraud. *Critical Perspectives on Accounting*, *13*(3), 297–310. https://doi.org/10.1006/cpac.2001.0527

den Nieuwenboer, N. A., Cunha, J. V. d., & Treviño, L. K. (2017). Middle managers and corruptive routine translation: The social production of deceptive performance. *Organization Science*, *28*(5), 781–803. https://doi.org/10.1287/orsc.2017.1153

Egorov, M., Kalshoven, K., Pircher Verdorfer, A., & Peus, C. (2020). It's a match: Moralization and the effects of moral foundations congruence on ethical and unethical leadership perception. *Journal of Business Ethics*, *167*, 707–723. https://doi.org/10.1007/s10551-019-04178-9

Ernst & Young. (2020). *Global integrity report*. https://www.ey.com/en_gl/global-integrity-report

Fehr, R., Fulmer, A., & Keng-Highberger, F. T. (2019). How do employees react to leaders' unethical behavior? The role of moral disengagement. *Personnel Psychology*, *73*(1), 73–93. https://doi.org/10.1111/peps.12366

Fischbacher, U., & Föllmi-Heusi, F. (2013). Lies in disguise—An experimental study on cheating. *Journal of the European Economic Association*, *11*(3), 525–547. https://doi.org/10.1111/jeea.12014

Fiske, S. T., Cuddy, A. J., & Glick, P. (2007, February). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Science*, *11*(2), 77–83. https://doi.org/10.1016/j.tics.2006.11.005

Gächter, S., & Renner, E. (2018). Leaders as role models and "belief managers" in social dilemmas. *Journal of Economic Behavior & Organization*, *154*, 321–334. https://doi.org/10.1016/j.jebo.2018.08.001

Gamliel, E., & Peer, E. (2013). Explicit risk of getting caught does not affect unethical behavior. *Journal of Applied Social Psychology*, *43*(6), 1281–1288. https://doi.org/10.1111/jasp.12091

Gerlach, P., Teodorescu, K., & Hertwig, R. (2019, January). The truth about lies: A meta-analysis on dishonest behavior. *Psychological Bulletin*, *145*(1), 1–44. https://doi.org/10.1037/bul0000174

Gino, F., Ayal, S., & Ariely, D. (2009). Contagion and differentiation in unethical behavior: The effect of one bad apple on the barrel. *Psychological Science*, *20*(3), 393–398. https://doi.org/10.1111/j.1467-9280.2009.02306.x

Gino, F., Ayal, S., & Ariely, D. (2013). Self-serving altruism? The lure of unethical actions that benefit others. *Journal of Economic Behavior & Organization*, *93*, 285–292. https://doi.org/10.1016/j.jebo.2013.04.005

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. https://doi.org/10.1016/j.obhdp.2012.03.011

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

Goodwin, G. P., Piazza, J., & Rozin, P. (2014, January). Moral character predominates in person perception and evaluation. *Journal of Personality and Social Psychology*, *106*(1), 148–168. https://doi.org/10.1037/a0034726

Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology*, *96*(5), 1029–1046. https://doi.org/10.1037/a0015141

Graham, J., Nosek, B. A., Haidt, J., Iyer, R., Koleva, S., & Ditto, P. H. (2011, August). Mapping the moral domain. *Journal of Personality and Social Psychology*, *101*(2), 366–385. https://doi.org/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. https://doi.org/10.1177/0956797618796480

Heintz, C., Karabegovic, M., & Molnar, A. (2016). The co-evolution of honesty and strategic vigilance. *Frontiers in Psychology*, *7*, 1503. https://doi.org/10.3389/fpsyg.2016.01503

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. https://doi.org/10.1002/bdm.1953

Hildreth, J. A. D., & Anderson, C. (2018). Does loyalty trump honesty? Moral judgments of loyalty-driven deceit. *Journal of Experimental Social Psychology*, *79*, 87–94. https://doi.org/10.1016/j.jesp.2018.06.001

Irlenbusch, B., Mussweiler, T., Saxler, D. J., Shalvi, S., & Weiss, A. (2020). Similarity increases collaborative cheating. *Journal of Economic Behavior & Organization*, *178*, 148–173. https://doi.org/10.1016/j.jebo.2020.06.022

Kim, M., Park, B., & Young, L. (2020). The psychology of motivated versus rational impression updating. *Trends in Cognitive Science*, *24*(2), 101–111. https://doi.org/10.1016/j.tics.2019.12.001

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. https://doi.org/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. https://doi.org/10.18637/jss.v082.i13

Lapsley, D. K., & Lasky, B. (2001). Prototypic moral character. *Identity*, *1*(4), 345–363. https://doi.org/10.1207/s1532706xid0104_03

LeBeau, B. (2019). *simglm: Simulate models based on the generalized linear model* (Version 0.7.4) [R package]. http://CRAN.R-project.org/package=simglm

Lemoine, G. J., Hartnell, C. A., & Leroy, H. (2019). Taking stock of moral approaches to leadership: An integrative review of ethical, authentic, and servant leadership. *Academy of Management Annals*, *13*(1), 148–187. https://doi.org/10.5465/annals.2016.0121

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

Lindstrom, M. J., & Bates, D. M. (1990). Nonlinear mixed effects models for repeated measures data. *Biometrics*, *46*(3), 673–687. https://doi.org/10.2307/2532087

Mayer, D. M., Aquino, K., Greenbaum, R. L., & Kuenzi, M. (2012). Who displays ethical leadership, and why does it matter? An examination of antecedents and consequences of ethical leadership. *Academy of Management Journal*, *55*(1), 151–171. https://doi.org/10.5465/amj.2008.0276

Meteyard, L., & Davies, R. A. I. (2020). Best practice guidance for linear mixed-effects models in psychological science. *Journal of Memory and Language*, *112*, 104092. https://doi.org/10.1016/j.jml.2020.104092

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. https://doi.org/10.1016/j.csda.2011.10.005

Pinto, J., Leana, C. R., & Pil, F. K. (2008). Corrupt organizations or organizations of corrupt individuals? Two types of organization-level corruption. *Academy of Management Review*, *33*(3), 685–709. https://doi.org/10.5465/amr.2008.32465726

Pittarello, A., Leib, M., Gordon-Hecker, T., & Shalvi, S. (2015, June). Justifications shape ethical blind spots. *Psychological Science*, *26*(6), 794–804. https://doi.org/10.1177/0956797615571018

R Core Team. (2014). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. http://www.R-project.org/

Rousseau, D. M., & McLean Parks, J. (1993). The contracts of individuals and organizations. *Research in Organizational Behavior*, *15*, 1–43.

Skitka, L. J., Bauman, C. W., & Sargis, E. G. (2005, June). Moral conviction: Another contributor to attitude strength or something more? *Journal of Personality and Social Psychology*, *88*(6), 895–917. https://doi.org/10.1037/0022-3514.88.6.895

Song, F., & Zhong, C.-B. (2015). You scratch his back, he scratches mine and I'll scratch yours: Deception in simultaneous cyclic networks. *Journal of Economic Behavior & Organization*, *112*, 98–111. https://doi.org/10.1016/j.jebo.2015.01.009

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. https://doi.org/10.1016/j.econlet.2017.08.011

Thielmann, I., Bohm, R., & Hilbig, B. E. (2021). Buying unethical loyalty: A behavioral paradigm and empirical test. *Social Psychological and Personality Science*, *12*(3), 363–370. https://doi.org/10.1177/1948550620905218

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. https://doi.org/10.1016/j.jesp.2018.06.009

Trevino, L. K., den Nieuwenboer, N. A., & Kish-Gephart, J. J. (2014). (Un)ethical behavior in organizations. *Annual Review of Psychology*, *65*, 635–660. https://doi.org/10.1146/annurev-psych-113011-143745

Waytz, A., Dungan, J., & Young, L. (2013). The whistleblower's dilemma and the fairness–loyalty tradeoff. *Journal of Experimental Social Psychology*, *49*(6), 1027–1033. https://doi.org/10.1016/j.jesp.2013.07.002

Weisel, O., & Shalvi, S. (2015, August 25). The collaborative roots of corruption. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(34), 10651–10656. https://doi.org/10.1073/pnas.1423035112

Zelmer, J. (2003). Linear public goods experiments: A meta-analysis. *Experimental Economics*, *6*, 299–310. https://doi.org/10.1023/A:1026277420119