Supplementary Materials for Review

These are the supplementary materials for review for the article submission "Collaborative cheating in hierarchical teams: Effects of incentive structure and leader behavior on subordinate behavior and perceptions of leaders".

Overview

This supplement contains a comparison of preregistered hypotheses for Study 1 and Study 2 with the presented results in the main manuscript, a more detailed description of the methods in Study 1, additional exploratory results from Studies 1 and 2, as well as tables for all reported and exploratory models presented in the manuscript (Tables S1 – S9), followed by supplementary Figures (Figures S1 – S3). Tables and figures that are not directly discussed in the manuscript are accompanied by a short description.

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 and readers to explore the additionally offered materials and analyses, including extensive robustness analyses, including various **plots**, and **exploratory analyses** of aspects not discussed in the main text.

The online supplementary materials (data, complete analysis code, full experiment code and materials) for both studies are presented here: https://osf.io/p2esr/. In addition, we provide the verbatim instructions for both studies in a separate file, for easier access.

Table of Content

Study 1

Preregistration Note for Study 1

The preregistration can be found here: https://osf.io/nsz5d/. Due to limitations of space, as well as to keep the article focused, we do not report all pre-registered tests in the main article. However, we are presenting all pre-registered analyses in the online supplement (see the document: https://osf.io/3h9nt/).

The three hypotheses laid out in the main article map onto the preregistered hypotheses as follows:

Main Article	Preregistration
Hypothesis 1	Hypothesis 1
Hypothesis 2	Hypothesis 2
Hypothesis 3	Hypothesis 6

The other preregistered hypotheses (Hypotheses 3,4,5, and 7) are thus not reported in the main article. Here, we want to briefly highlight that we do not find support for Preregistration Hypothesis 3 (positing an interaction between checks and cheats for evaluation rounds only), and Preregistration Hypothesis 4 (positing a moderating impact of Moral Foundations Authority on updating Leader evaluations).

We find support for Preregistration Hypothesis 5 (expecting stronger updating of Leader evaluations in the Low Check to High Check condition than in the High Check to Low Check condition) and Preregistration Hypothesis 7 (expecting the amount of money sent to the Leader in the trust game to be negatively related with Leader evaluations after the trust game). For the exact tests and models, see the online supplement linked above.

Extended Methods Study 1

The following describes the methods of Study 1 in more detail. Please note that all verbatim instructions (including code to recreate Study 1, and Study 2) are also provided on the OSF repository.

The general structure of the experimental task followed two stages, the first consisted of a rely-or-verify game (adapted from Levine & Schweitzer, 2015), and the second consisted of a trust game (Berg et al., 1995). The rely-or-verify game in this study 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). The die rolls were pseudo randomized, such that for each half of the game (10 rounds) participants rolled exactly one 1, two 2s, three 3s, two 4s, one 5, and one 6. The order of these die rolls was randomized within each game half. We chose this procedure to navigate two trade-offs: Firstly, the die results were slightly skewed to increase cheating (as lower die rolls are more likely to be inflated), without making participants skeptical about die rolls' being unfair. 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.

Having rolled the die, the Player was able to report any number (i.e., either the number they actually rolled, or an inflated number) to the Leader, who had been described to them as the Leader of the dyad, responsible for checking results. The Leader 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. Four different scenarios are possible (see Figure 1).

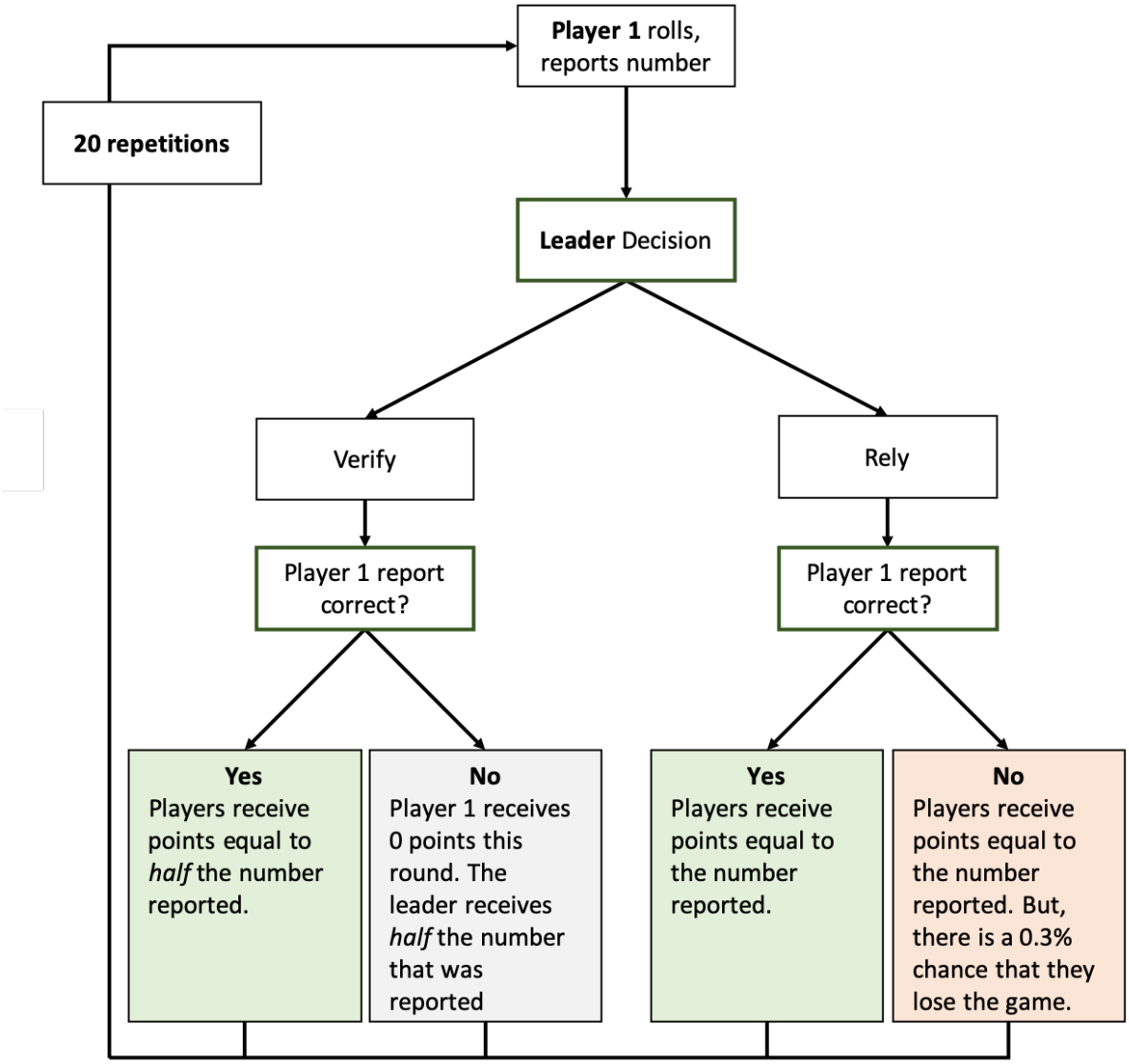

Figure 1: Structure of the rely-or-verify game as presented to the study participants

If the Leader checked and the report was correct, both the Player and the Leader received 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 checked and the report was incorrect, the Player received 0 points, whereas the Leader received half the points reported

(e.g. if the Player reported 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 did not check, and the Player's report was correct, both agents received the number of points reported. Finally, if the Leader did not check an inflated report, both agents still received the number of points reported. However, 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. Following the logic of Abbink et al. (2002), 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 were informed that the chance of losing the game at 0.3%, we set the actual probability of losing the game to 0. This approach was taken to ensure that participants did not actually lose the game, since studying the effects of losing the game was not our aim.

In addition to the die roll task, each Player evaluated 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 were allowed to start the game, they answered six comprehension check questions, which ensured they understood the basic payoff scheme, and that they were always playing with the same Leader. All instructions presented to the participants can be found in the online supplementary materials.

Leader behavior was manipulated in two conditions. In the High Check to Low Check 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 Check to High Check condition, such that the Leader checked with

only a 10% chance in the first 10 rounds, and with a 90% chance in the last 10 rounds. In both conditions, the Leader never checked reports of a 2 nor a 1, so as to not raise suspicion in participants that they were playing a computer.

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

Subsequently, Players played a one-shot trust game (Berg et al., 1995) with their Leader. Each participant acted as the sender, thus allowing them to send any amount of points they had earned in the previous game to their former Leader, which would be tripled. Players were instructed 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 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 learned about the outcome of the trust game. In this study, the Leader never chose 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 of Leader 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-instantiated all points sent to the Leader.

Finally, participants went through a funnel debrief, including questions probing participants for their perception of the purpose of the study, whether they had 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.

Additional demographics information

Most participants (39.9%) had finished a BA/BS degree, and 88.3% had at least some college education. Participants reported an average of 16.14 (SD = 11.64) years of employment; 62.3% of the participants reported previous experience working in a supervisory role. Participants were paid $0.50 for completing the study, plus $0.05 for each point they earned during the study (average pay: $3.21, SD = $0.26).

Literature

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

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

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. https://doi.org/10.1037/t13084-000

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

Graham, J., Nosek, B. A., Haidt, J., Iyer, R., Koleva, S., & Ditto, P. H. (2011). Mapping the moral domain. *Journal of Personality and Social Psychology*, *101*(2), 366-385. https://doi.org/10.1037/a0021847

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

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

Supplementary Results for Study 1

Table S1: Full model results

	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 S1: 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.

Table S2: Predicting player cheating based on checking in previous round

	Cheating		
Predictors	*Odds Ratios*	*95% CI*	*p*
(Intercept)	0.46	0.18 – 1.16	0.098
Checked last round	**0.50**	**0.38 – 0.67**	**<0.001**
Actually rolled number	**0.39**	**0.31 – 0.49**	**<0.001**
Round	**0.94**	**0.89 – 1.00**	**0.033**
MF: Authority	0.70	0.43 – 1.16	0.168
MF: Ingroup/Loyalty	0.75	0.46 – 1.23	0.254
MF: Fairness	**0.66**	**0.45 – 0.96**	**0.028**
Dospert: Investment	1.04	0.69 – 1.55	0.866
Dospert Betting	**1.58**	**1.05 – 2.37**	**0.028**
Random Effects			
σ^2	3.29		
$\tau_{00\ participant.code}$	13.34		
$\tau_{11\ participant.code.round}$	0.02		
$\tau_{11\ participant.code.actually\_rolled\_number}$	0.17		
ρ_{01}	-0.27		
	-0.74		
ICC	0.71		
$N_{\ participant.code}$	308		
Observations	5852		
Marginal R^2 / Conditional R^2	0.181 / 0.766		

Table 2: Logistic mixed effects model modeling player cheating based on having been checked in the prior round ("checked last round" – coded 1 if yes, 0 if not). Estimates are odds ratio. This analysis uses data for round 2 – 20 (first round observations are dropped as "checked last round" does not exist for round 1.

Table S3: Robustness Analysis of models using a subset of only cheaters.

Table S2 presents a robustness analysis for models reported in Table S1. 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 S3: Robustness analysis of models using a subset of only cheaters.

	Dependent variable:				
	Cheating *generalized linear mixed-effects* (1)	Morality *linear mixed-effects* (2)	Trustworthiness *linear mixed-effects* (3)	Competence *linear mixed-effects* (4)	Closeness *linear mixed-effects* (5)
Intercept	**1.19*****	**0.89*****	0.17	0.23	**-0.92*****
	(0.77, 1.61)	(0.57, 1.21)	(-0.17, 0.52)	(-0.15, 0.61)	(-1.30, -0.55)
Condition	0.09	0.17	**0.61***	**0.79****	**0.83****
	(-0.38, 0.55)	(-0.29, 0.63)	(0.11, 1.10)	(0.25, 1.33)	(0.29, 1.37)
Cheating		-0.58	-0.25	0.16	-0.29
		(-1.21, 0.06)	(-1.00, 0.50)	(-0.61, 0.93)	(-0.98, 0.40)
Round	**14.13*****	-0.61	**5.76***	**4.00**	**8.60****
	(6.07, 22.19)	(-5.30, 4.08)	(0.06, 11.47)	(-2.03, 10.03)	(3.16, 14.04)
Round^2	**14.41*****	**3.82***	1.01	**4.35***	**3.69***
	(6.20, 22.62)	(0.76, 6.88)	(-2.67, 4.70)	(0.64, 8.07)	(0.45, 6.93)
Round^3	-6.49	-0.59	**-4.28***	**-4.20***	**-4.06****
	(-14.60, 1.63)	(-3.44, 2.26)	(-7.72, -0.84)	(-7.66, -0.74)	(-7.07, -1.06)
Actual number rolled	**-1.01*****				
	(-1.12, -0.91)				
MF: Authority	0.05	-0.15	-0.27	0.02	-0.26
	(-0.26, 0.35)	(-0.42, 0.12)	(-0.56, 0.01)	(-0.30, 0.34)	(-0.58, 0.07)
MF: Ingroup/Loyalty	-0.18	0.24	**0.38****	0.18	**0.50****
	(-0.48, 0.11)	(-0.03, 0.50)	(0.10, 0.66)	(-0.13, 0.49)	(0.19, 0.82)
MF: Fairness	-0.20	0.05	0.03	0.09	-0.01
	(-0.42, 0.02)	(-0.14, 0.24)	(-0.18, 0.23)	(-0.14, 0.31)	(-0.24, 0.22)
DOSPERT: investment	**-0.29***				
	(-0.54, -0.03)				
DOSPERT: betting	**0.33****				
	(0.09, 0.57)				
Condition * Round	**-45.46*****	-5.79	**-22.53*****	**-18.92*****	**-19.98*****
	(-57.48, -33.44)	(-12.46, 0.87)	(-30.61, -14.46)	(-27.48, -10.37)	(-27.70, -12.27)
Condition * Round^2	**-26.90*****	**-8.34*****	**-5.46***	**-10.22*****	**-8.89*****
	(-38.73, -15.06)	(-12.73, -3.95)	(-10.75, -0.17)	(-15.55, -4.89)	(-13.53, -4.25)
Condition * Round^3	**15.52****	2.22	**11.96*****	**7.74****	**11.22*****
	(3.89, 27.14)	(-1.96, 6.40)	(6.92, 17.00)	(2.67, 12.80)	(6.82, 15.62)
Cheating * Round		0.30	-1.70	9.26	9.85
		(-14.07, 14.67)	(-18.86, 15.46)	(-8.67, 27.19)	(-6.13, 25.84)
Cheating * Round2		**-11.30***	0.64	-8.17	-7.65
		(-22.22, -0.39)	(-12.46, 13.73)	(-21.44, 5.10)	(-19.24, 3.94)
Cheating * Round3		-7.20	0.94	1.09	-2.52
		(-17.31, 2.92)	(-11.25, 13.13)	(-11.18, 13.36)	(-13.19, 8.15)
Condition * Cheating		0.30	-0.11	-0.71	0.43
		(-0.63, 1.23)	(-1.21, 0.99)	(-1.84, 0.43)	(-0.59, 1.45)
Condition * Cheating * Round^1		-4.11	0.24	1.09	-22.54
		(-24.51, 16.29)	(-24.35, 24.82)	(-24.57, 26.75)	(-45.44, 0.36)
Condition * Cheating * Round^2		**16.71***	5.18	**19.00***	10.55
		(1.65, 31.77)	(-12.94, 23.29)	(0.68, 37.32)	(-5.44, 26.54)
Condition * Cheating * Round^3		10.14	4.07	-0.69	3.53
		(-3.87, 24.16)	(-12.82, 20.97)	(-17.69, 16.30)	(-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 S3: 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 2 in the main article.

Table S4: Differential impact of checking type on different evaluation dimensions

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, 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 S1; detailed results of these interaction models are presented in Table S3).

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 S4). 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 (see Model 6 in Table S4). 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 S3).

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 (compared to morality). These results are illustrated in Figure S1.

Figure S1: Differences in updating evaluations based on check type

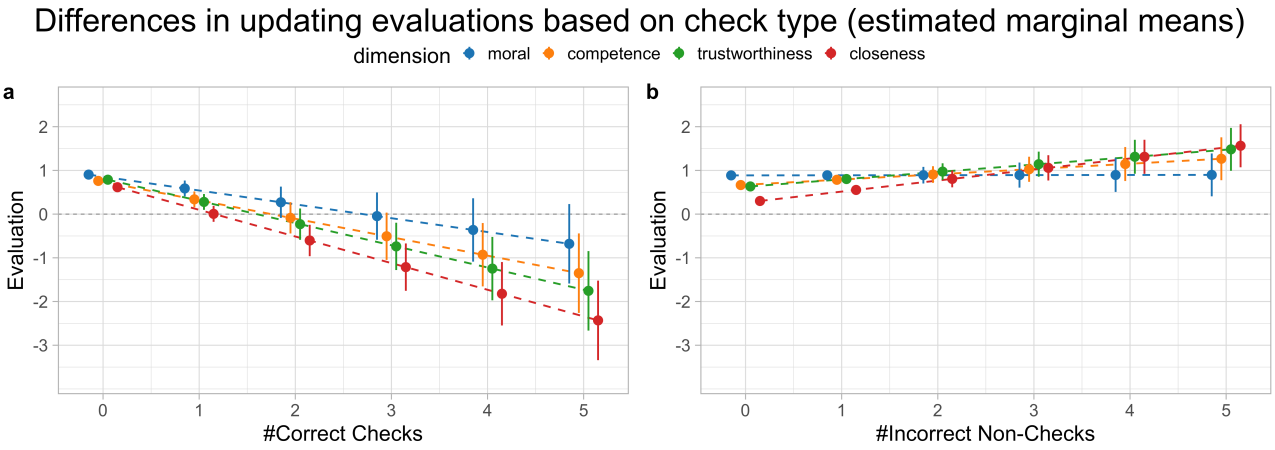

Figure S1: Estimated marginal means from linear mixed effects models predicting Leader evaluations based on the number of a specific check type (correct check, or incorrect non-check) in the five rounds leading up to an evaluation (i.e. rounds 1-5,6-10,11-15,16-20). Each plot represents a separate model. 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 S3).

Table S4 presents model results comparing the impact of different types of checks (i.e. correct check, incorrect check, correct non-check, incorrect non-check) 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 S1.

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 S4

	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)
No. of checks of specified 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 * No. of checks†	**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 * No. of checks †	**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 * No. of checks †	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 S4: Linear mixed effects models modelling the differential effect (interactions) of check type on the different evaluation dimensions.

Table S5: check type analysis

	Check type: correct check				*Check type: incorrect non-check*			
	Dependent variable							
	Morality	Trustworthiness	Competence	Closeness	Morality	Trustworthiness	Competence	Closeness
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(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)
No. of checks	**-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)
No. of checks * 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)
No. of checks * 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)
No. of checks * 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 S5: Results for linear mixed effects models predicting Leader evaluations (morality, trustworthiness, competence & closeness) based on number of correct checks (models 1-4), and number of incorrect non-checks (models 5-8) in a game period (the 5 rounds leading up to an evaluation), controlling for participant personality, and game period. Models include random intercepts for participants, and random slopes for check type.

Table S6: How does a change in checking behavior influence evaluations?

In this analysis, we explore whether a Leader switching from checking to non-checking will be perceived more positively than a Leader switching from non-checking to checking. This analysis was originally one of the main foci of Study 1 (e.g., preregistered hypothesis 4 & 5, see online supplementary), but moved here for streamlining the manuscript.

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 prediction, 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-to-High condition than in the High-to-Low 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 = 0.053; see Table S6). 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.

Table S6: Final (end of task) evaluations

	Dependent variable:				
	Morality	Trustworthiness	Competence	Closeness	Risk Aversion
	(1)	(2)	(3)	(4)	(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)
Overall cheating rate	**-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)
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 S6: 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.

Table S7: Contributions to the trust game, ZOIB model estimates (Study 1)

Distribution parameter	Term	Estimate	95% CI low	95% CI high
mean	Intercept	0.38	0.31	0.45
	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	Intercept	3.44	2.23	5.02
	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	Intercept	0.47	0.37	0.57
	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	Intercept	0.29	0.16	0.44
	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 S7: 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.

Study 2

Preregistration Note for Study 2

The original preregistration for Study 2 can be found here: https://osf.io/kxg7n As with Study 1, we could not report all preregistered Hypotheses in the main text. In particular, Preregistered Hypothesis 2 (*We predict that participants in the competing payoffs condition will evaluate their leaders as more close, competent, moral, and trustworthy in the final round of the rely-or-verify game than participants in the aligned payoffs condition*) is not reported in the main text. We note that we could not find support for Hypothesis 2 (see "End of task evaluations by condition" below, as well as full model results in the supplement: https://osf.io/ej46r/

The three hypotheses laid out in the main article map onto the preregistered hypotheses as follows:

Main Article	**Preregistration**
Hypothesis 1A, B	Hypothesis 1A, B
Hypothesis 2A, B, C	Hypothesis 3A, B, C
Hypothesis 3	Hypothesis 4

Additional demographics information

Most participants (47.1%) had finished a BA/BS degree, and 91.8% had at least some college education. Participants reported an average of 15.43 (SD = 10.81) years of employment; 63.0% of the participants reported previous experience working in a supervisory role.

To keep incentives similar to Study 1, participants were paid $0.50 for completing the study, plus 5 cents for each point they earned during the study. However, a minimum payment of $2.75 was set to ensure fair payment. Due to the change in round numbers from 20 to only 10 rounds in Study 2, only few participants who cheated to large degree exceeded the minimum pay (average pay: $2.77, SD = $0.094).

Supplementary Results for Study 2

Table S8: LME models for Leader Evaluations (Study 2)

	Dependent variable:			
	Morality	Trustworthiness	Competence	Closeness
	(1)	(2)	(3)	(4)
Intercept	**0.77*****	**0.84*****	**0.93*****	**0.60*****
	(0.67, 0.87)	(0.74, 0.94)	(0.82, 1.03)	(0.47, 0.73)
Previous Evaluation	**0.42*****	**0.49*****	**0.45*****	**0.48*****
	(0.39, 0.45)	(0.46, 0.51)	(0.43, 0.48)	(0.46, 0.50)
Competing Payoffs	0.10	0.03	-0.01	**-0.19***
	(-0.02, 0.23)	(-0.09, 0.15)	(-0.13, 0.12)	(-0.36, -0.03)
Cheat	**-0.12****	0.09	**0.17*****	**0.19*****
	(-0.20, -0.04)	(-0.01, 0.18)	(0.07, 0.27)	(0.09, 0.28)
Check	-0.10	**-0.74*****	**-0.81*****	**-0.91*****
	(-0.21, 0.01)	(-0.88, -0.61)	(-0.95, -0.67)	(-1.05, -0.78)
Round Number	-0.01	**-0.01****	**-0.01***	**-0.01***
	(-0.02, 0.002)	(-0.02, -0.004)	(-0.02, -0.002)	(-0.02, -0.001)
MF: Fairness	**0.15*****	**0.12*****	**0.10****	**0.11****
	(0.09, 0.21)	(0.06, 0.18)	(0.04, 0.15)	(0.03, 0.19)
MF: Traditional	**0.15*****	**0.11*****	**0.11****	**0.09***
	(0.08, 0.21)	(0.04, 0.17)	(0.04, 0.17)	(0.01, 0.18)
MF: Ingroup	0.06	**0.09****	**0.10****	**0.15****
	(-0.005, 0.13)	(0.02, 0.15)	(0.03, 0.16)	(0.06, 0.24)
DOSPERT: betting	-0.004	-0.01	-0.04	0.002
	(-0.07, 0.07)	(-0.07, 0.06)	(-0.10, 0.03)	(-0.09, 0.09)
DOSPERT: investment	0.02	0.03	0.01	0.08
	(-0.05, 0.08)	(-0.03, 0.09)	(-0.05, 0.08)	(-0.003, 0.17)
Competing Payoffs * Cheat	0.01	**-0.17***	**-0.50*****	**-0.42*****
	(-0.11, 0.12)	(-0.31, -0.03)	(-0.65, -0.36)	(-0.57, -0.28)
Competing Payoffs * Check	**-0.24****	-0.02	**0.31****	0.17
	(-0.39, -0.09)	(-0.20, 0.16)	(0.12, 0.49)	(-0.02, 0.35)
Cheat * Check	0.15	**-0.63*****	**-0.96*****	**-0.99*****
	(-0.003, 0.31)	(-0.82, -0.44)	(-1.15, -0.76)	(-1.18, -0.80)
Competing Payoffs * Check * Cheat	-0.03	**0.77*****	**1.60*****	**1.18*****
	(-0.26, 0.20)	(0.49, 1.04)	(1.31, 1.89)	(0.90, 1.46)
Observations	4,160	4,160	4,160	4,160
Akaike Inf. Crit.	9,334.55	10,698.82	11,010.82	11,072.46
Note				*p<0.05; **p<0.01; ***p<0.001

Table S8: Results for linear mixed effects models predicting Leader evaluations (morality, trustworthiness, competence & closeness) of Study 2. All models include random slopes for round number.

Table S9: Contributions to the trust game, ZOIB model estimates (Study 2)

Distribution Parameter	Term	Estimate	95% CI low	95% CI high
mean	Intercept	0.51	0.45	0.58
	Condition*	**-0.07**	**-0.13**	**-0.02**
	Cheating	**-0.20**	**-0.30**	**-0.11**
	Leader Trustworthiness	0.01	-0.01	0.03
	Leader Risk aversion	-0.00	-0.02	0.01
	MF: Fairness	0.02	-0.01	0.04
	MF: Authority	0.03	-0.00	0.05
	MF: Ingroup / Loyalty	-0.01	-0.04	0.02
	DOSPERT: investment	**0.04**	**0.01**	**0.06**
	DOSPERT: betting	0.01	-0.02	0.04
phi	Intercept	3.73	2.34	5.58
	Condition*	0.00	-0.06	0.06
	Cheating	0.09	-0.01	0.19
	Leader Trustworthiness	0.01	-0.02	0.04
	Leader Risk aversion	0.01	-0.01	0.03
	MF: Fairness	-0.02	-0.05	0.01
	MF: Authority	**-0.07**	**-0.11**	**-0.03**
	MF: Ingroup / Loyalty	**0.06**	**0.02**	**0.09**
	DOSPERT: investment	-0.07	-0.12	-0.03
	DOSPERT: betting	-0.00	-0.04	0.03
zoi	Intercept	0.31	0.21	0.42
	Condition*	0.03	-0.06	0.13
	Cheating	0.15	-0.04	0.34
	Leader Trustworthiness	-0.02	-0.06	0.01
	Leader Risk aversion	0.02	-0.01	0.05
	MF: Fairness	-0.02	-0.07	0.03
	MF: Authority	**-0.05**	**-0.09**	**-0.01**
	MF: Ingroup / Loyalty	-0.02	-0.07	0.03
	DOSPERT: investment	-0.02	-0.06	0.03
	DOSPERT: betting	0.03	-0.02	0.09
coi	Intercept	0.66	0.44	0.84
	Condition*	**-0.21**	**-0.39**	**-0.02**
	Cheating	**-0.39**	**-0.65**	**-0.09**
	Leader Trustworthiness	**0.08**	**0.03**	**0.15**
	Leader Risk aversion	-0.01	-0.06	0.04
	MF: Fairness	0.03	-0.05	0.12
	MF: Authority	**-0.11**	**-0.22**	**-0.01**
	MF: Ingroup / Loyalty	-0.03	-0.14	0.06
	DOSPERT: investment	**0.11**	**0.03**	**0.20**
	DOSPERT: betting	-0.10	-0.22	0.01

Table S9: 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.

Trust game behavior

We explored whether the degree of cheating, or the trustworthiness ratings differently impacted behavior depending on condition. To this end, we ran an exploratory analysis focusing on these interactions, again using a ZOIB model. This model suggested that the amount of cheating differently predicted investment depending on condition, with high cheaters giving less money in the competing payoffs condition than in the aligned payoffs condition (b_{coi} = -0.40, 95% CrI [-0.64, 0.00]), indicating that the negative relationship between cheating and subsequent trust can be mainly attributed to participants in the competing payoffs condition, and that participants in the aligned payoff condition behaved more like participants in Study 1, which used that kind of incentive structure (full model results in online supplement). Similarly, we found an interaction between condition and trustworthiness ratings, such that trustworthiness ratings were more positively related to money sent in the aligned payoffs condition than for participants in the competing payoff conditions (b_{mean} = -0.04, 95% CrI [-0.08, -0.00]). Notably, both interaction estimates contain 0 in their credible intervals, and must be regarded as weak evidence.

Replicating effects of check type on evaluation from Study 1

With the specific aim of replicating results from Study 1 (cf. Table S4 above), we also explored the relationship of check type (i.e., correct checks, incorrect checks, correct non-checks, or incorrect non-checks) and Leader evaluations in the aligned payoffs condition. Using correct non-checks as the baseline, we examined how other types of checks differently impacted Leader evaluations.

This analysis revealed that, compared to correct non-checks, incorrect non-checks (letting cheating go undetected) were beneficial for closeness evaluations (beta = 0.18, 95% CI = [0.07, 0.30], p = .0011), as well as for competence evaluations (beta = 0.16, 95% CI = [0.05, 0.27], p = .0041), but detrimental for morality evaluations (beta = -0.13, 95% CI = [-0.22, -0.05], p = .0024), and had no differential effect for trustworthiness evaluations (beta = 0.08, 95 % CI = [-0.03, 0.018], p = .14). Letting cheating go undetected therefore was the strongest positive driver for closeness evaluations (see also Figure 5 in manuscript; full model results in online supplement – 4.2.6).

End of task evaluations by condition:

Similar as to Study 1 (cf. Table S6), we investigated whether participants in the competing payoffs condition would evaluate their Leaders as more close, competent, moral, and trustworthy in the final round of the rely-or-verify game than participants in the aligned payoffs condition. We built individual linear models for each evaluation dimension, using condition, the interaction of a how often a Player had cheated, and how often they got checked, as well as personality variables as predictors.

Counter to our initial prediction, the payoff scheme did not significantly affect final Leader evaluations, with the exception of closeness, which showed a negative main effect of condition (also counter to our hypothesis): participants in the competing payoffs condition evaluated their Leader as less close than participants in the aligned payoffs condition (beta = -0.39, 95% CI = [-0.76, -0.03], p = .034).

Thus, the overall differences in evaluations between conditions were not as hypothesized (see online supplement – 5.1 – for full model results). Even though checking was perceived to be more negative in the aligned than in the competing payoffs condition, the few checks that Players received in the aligned payoffs condition were not enough to lead to a net negative effect of condition.

Supplementary Figures

Figure S2: Graph comparing trajectories of Leader evaluations for cheaters and non-cheaters (Study 1)

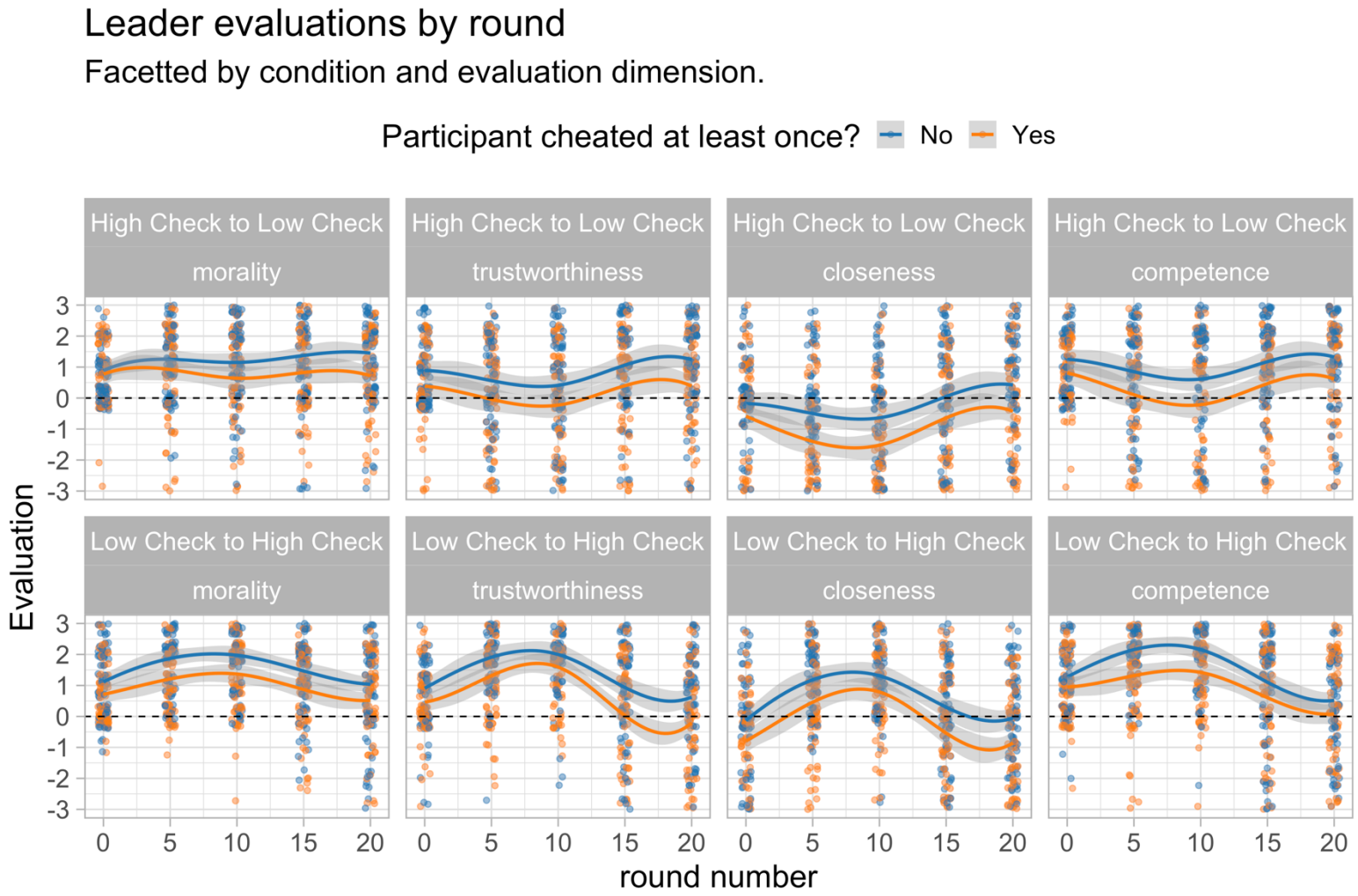

Figure S2: Leader evaluations over round, comparing participant who cheated at least once with participants who never cheated in Study 1. Facetted by condition and evaluation dimension. Lines are b-splines with 3 degrees of freedom, and a knot at round 10.

Figure S3: Density plot of money sent to Leader in trust game (Study 2)

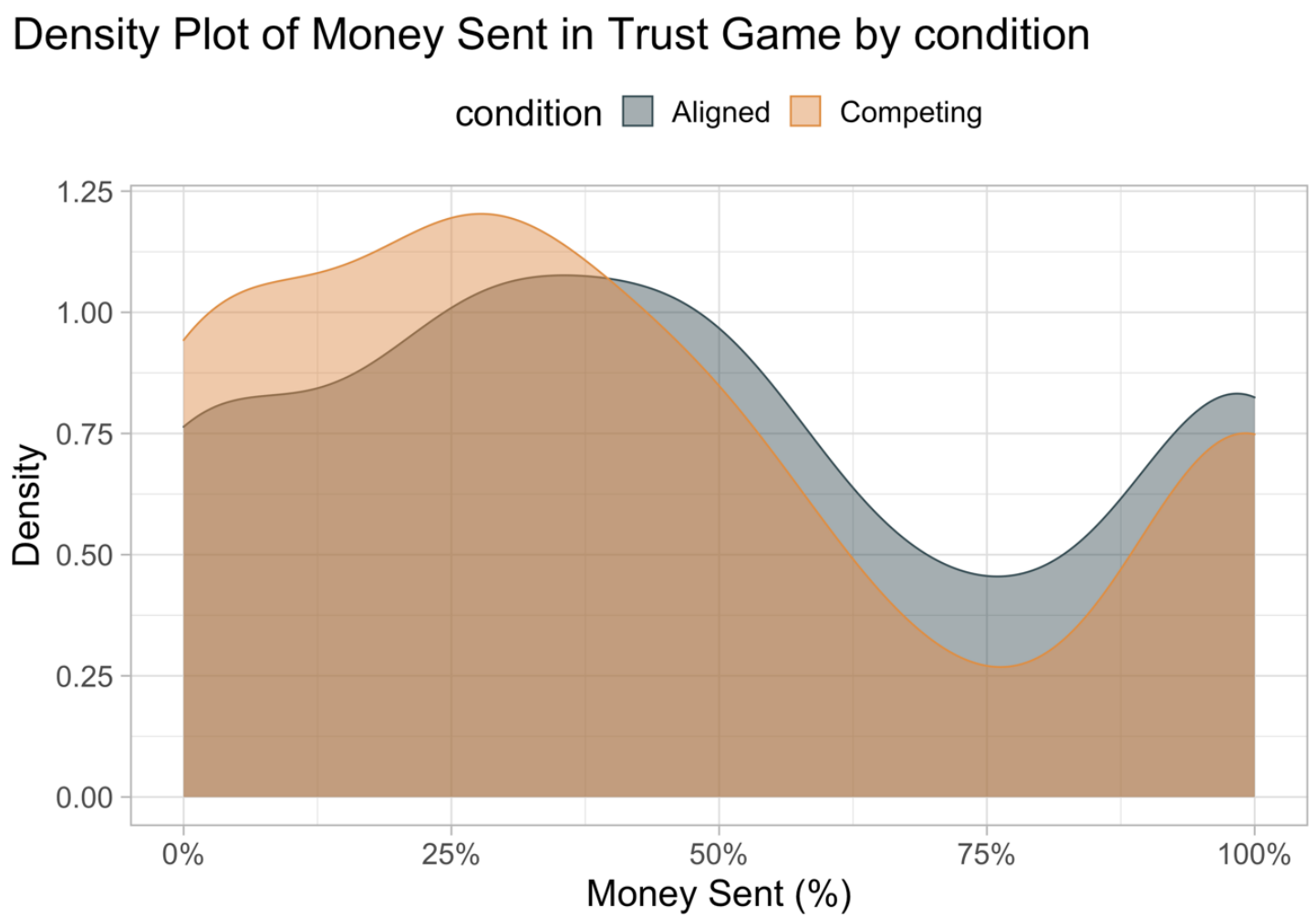

Figure S3: Density plot of the percentage of possible points sent to the Leader in the trust game by condition in Study 2.