Being Bad to Look Good: Competence Reputational Stakes Can *Increase* Unethical Behavior

Cory J Clark
University of Pennsylvania

Daniel Keighley
Durham University

Milica Vasiljevic
Durham University

In press at Evolutionary Behavioral Sciences as of 5.12.22.

Corresponding author: Cory Clark; cjclark@sas.upenn.edu

Abstract

Two studies (total n = 1,245) explored the influence of (1) receiving public *vs.* private performance feedback, (2) competing on a team *vs.* solo, and (3) individual differences in team competition participation on cheating behavior. Participants were given opportunities to cheat in an online trivia competition and self-reported their cheating behavior. Meta-analyses of Studies 1 and 2 revealed that participants who believed their performance feedback would be public cheated more than those who believed their performance feedback would be private, and individuals who regularly participate in team competition cheated more than those who do not. We found no evidence that experimentally manipulating team competition (*vs.* solo competition) influenced cheating. Our findings suggest that people will put their moral reputations at risk in order to protect their competence reputations by engaging in unethical behavior that signals (false) competence to others.

Keywords: ethical behavior, cheating, status, reputation, signaling

Public Significance Statement

The present research found that providing public performance feedback (as opposed to private feedback) increased the tendency for people to cheat in a competence task and that those who regularly participate in team competitions were more prone to cheating. These findings suggest that competence reputational stakes can increase dishonest behavior aimed at enhancing one's apparent value to the social group and that those most attracted to group competition may be most likely to signal competence falsely.

1. “A bad conscience is easier to cope with than a bad reputation.” (Nietzsche, 1887/1974, p. 102)

People serve as their own Public Relations agents, managing their reputations and public image (Fehr, 2004; Sperber & Baumard, 2012; Tennie et al., 2010). A great deal of research has shown that reputational stakes, such as public visibility of behavior, reduce selfish and antisocial behavior (e.g., Gächter & Fehr, 1999; Engelmann & Fischbacher, 2008; Van Vugt & Hardy, 2010). The present work tests whether reputational stakes can also *increase* unethical behavior. Specifically, people may be willing to engage in unethical behavior—inevitably putting their moral reputations at risk—in order to preserve their public *competence* reputations.

Humans evolved to detect, exclude, and punish costly members of the social group, those who free ride and take advantage of others (Axelrod & Hamilton, 1981; Cosmides & Tooby, 1992; Fehr et al., 2002; Kurzban & Leary, 2001; Nowak & Sigmund, 1998; Trivers, 1971) because this benefits individuals and social groups by sustaining high levels of cooperation (Fehr & Fischbacher, 2004; Milinski et al., 2001). Consequently, being detected as a free rider oneself can be costly, and so people are deeply concerned with maintaining positive social reputations (e.g., Vonasch et al., 2018; Wu et al., 2016a, 2016b), even among children as young as age five (Engelmann et al., 2012).

Much of this work focuses on moral reputation and cooperation, such as contributions to public goods, finding that people are more cooperative and prosocial when their behavior will be known to others and they have a chance to gain social approval or build a positive moral reputation (e.g., Delmas & Lessem, 2014; Gächter & Fehr, 1999; Engelmann & Fischbacher, 2008). But people may seek to manage other aspects of their reputations to obtain social benefits

and avoid costs (e.g., Anderson & Kilduff, 2009a; Clark et al., 2019). One of the best ways to attain social status is to provide benefits to others (Durkee et al., 2020). Thus, individuals can pursue status in social groups by enhancing their apparent value to the social group (Winegard et al., 2020), for example, by signaling commitment to the group (Clark & Winegard, 2020), displaying generosity (Hardy & Van Vugt, 2006), or demonstrating competence (Anderson & Kilduff, 2009b; Henrich & Gil-White, 2001). But creating the *illusion* of adding value to the group, for example, by faking commitment or competence, can also be effective (Anderson & Kilduff, 2009a). If people award status and benefits to those who add group value, they might also award status and benefits to those whom they *believe* add value. This creates an incentive for people to signal value to others falsely.

In the real world, people must balance various tradeoffs by weighing the probabilities and magnitudes of different risks and rewards (Clark et al., 2022). A demonstration of bravery might increase one's value to the social group, but also puts one at risk of harm. Or signaling commitment to a particular social group might signal antipathy toward another group. One dilemma people frequently face is whether to cheat (Becker, 1974), and people are generally more willing to cheat as incentives increase, particularly when there is little risk of getting caught (Kajackaite & Gneezy, 2017). But cheating and other unethical behavior always involve *some* risk of getting caught. It can never be known for sure that other people have no way of discovering the cheating behavior (or even whether one's conscience could lead to a later admission of guilt, especially if questioned). Consequently, unethical behavior always carries some risk of damaging one's moral reputation.

Organizational behavior research often uses economic games with financial incentives to explore cheating behavior, but people can be incentivized to cheat for reputational reasons as

well. Although many sports, games, and competitions are played for money (e.g., professional sports, poker tournaments, chess tournaments), many are not, and yet cheating is still a regular problem (Kamis et al., 2016), suggesting that there is more at stake than financial gain. Cheating may allow people to falsely display competence to others (Schwieren & Weichselbaumer, 2010), which can contribute to higher status. Indeed, a series of studies found that risks of losses to status increased cheating behavior (Pettit et al., 2016). At first this might seem puzzling—surely cheating creates risks of harming one's good moral reputation—but people might be willing to put their moral reputations at a small risk in order to obtain reputational gains by increasing their apparent competence (or avoid reputational losses by revealing incompetence).

Additionally, people may be more motivated to preserve their competence reputations when working as part of a group. Humans engage in a wide variety of behaviors geared toward attaining and maintaining status among ingroup members in relation to outgroup members. For example, people are particularly generous when evaluating the ideas and behavior of ingroup members (e.g., Christenson & Kriner, 2017; Claassen & Ensley, 2016; Cohen, 2003; Hawkins & Nosek, 2012; Kahan et al., 2012), and people seek out and are credulous toward information that supports ingroup beliefs (e.g., Campbell & Kay, 2014; DeMarree et al., 2017; Ditto et al., 2019a, 2019b; Frimer et al., 2017; Gampa et al., 2019; Kahan et al., 2017; Lord et al., 1979; Stroud 2008, 2010; Taber & Lodge, 2006). Although people may wish to be viewed as highly competent across contexts, group contexts may amplify these desires and lead people to pay higher costs or take larger risks to preserve their competence reputations. In the present research, we hypothesized that competence reputational stakes might increase cheating behavior *more* in group contexts compared to individual contexts. Consistent with this idea, past research has found that team-based compensation schemes increase cheating behavior (Conrads et al., 2013).

A naturalistic setting that includes these precise conditions (competence reputational stakes in group settings), team sports, is one domain in which cheating behavior has been studied extensively (e.g., Šukys, 2013; Zimniuch, 2009). These investigations often consider economic (Preston & Szymanski, 2003) and motivational climate factors (e.g., Palou et al., 2013; Zaksaite, 2012) as well as characteristics of athletes themselves (Zaksaite, 2012). It is often reported that the competitiveness and performance incentives are key environmental features that promote cheating behavior in sports, but to our knowledge, no research has ever compared cheating behavior between those who opt in to team competition and those who do not. It is possible that the reputational stakes and groupishness are primary drivers of cheating in team competitions, but it is also possible that those attracted to team competitions are more prone to cheating. As an additional exploratory measure, we tested the relationship between regular participation in team competition and cheating behavior. Because we identified no prior work exploring this question, we had no *a priori* hypotheses, but we thought two alternative hypotheses were both plausible: (1) those who regularly participate in team competition might take such competitions more seriously and be more respectful of the norms and thus cheat less, or (2) those who regularly participate in team competition might care more about winning and displaying competence to others and so might be more likely to cheat.

1.1 The Present Research

Whereas past research has demonstrated how *moral reputational stakes often decrease antisocial behavior*, the present work tests whether *competence reputational stakes can increase unethical behavior*. As suggested by Friedrich Nietzsche in the epigraph, “A bad conscience is easier to cope with than a bad reputation.” People may be willing to engage in unethical behavior to preserve other aspects of their reputation in group contexts.

In two studies, we explored the influence of (1) receiving public performance feedback *vs.* receiving private performance feedback and (2) competing on a team *vs.* competing solo on cheating behavior in a competence task. In both studies, participants were given the opportunity to cheat in an online trivia competition, in which they could easily cheat by looking up answers online. To rule out the possibility that the experimental manipulations influenced *actual* performance, a second comparison round of trivia was time-restricted, eliminating the ability to cheat. As a second measure of cheating behavior, participants also self-reported the number of trivia questions they cheated on at the conclusion of the studies. Because the public performance condition puts participants' competence reputations at risk, we hypothesized that participants would be more likely to cheat in the public feedback condition than the private feedback condition. We also expected that this effect might be larger in the Team competition condition than the Solo condition. Because group contexts may be particularly likely to arouse concerns about social status, people might have a stronger desire to enhance their apparent competency to ostensible teammates than to other individual competitors. We also explored the relationship between cheating and one individual difference variable, regular participation in team competition, but we had no *a priori* hypotheses regarding the relationship between participation in team competitions and cheating

2. Open Science Statement

No participants were excluded from either study except those who requested their data be withdrawn after the debriefing (in accordance with ethics requirements). Data were not analyzed until the target samples were reached for each study, and no data were collected after that point. There are no undisclosed manipulations or dependent variables. All data that have ever been

collected to test the present hypotheses are reported in the present paper—there are no file drawer studies. Upon acceptance for publication, all data and syntax will be made publicly available on the Open Science Framework, along with the supplemental materials containing the Qualtrics surveys for Studies 1 and 2. Ethics approval for both studies was obtained from the authors' institution at the time the data were collected.

3. Study 1

3.1 Method

3.1.1 Participants. We recruited 450 English-speaking participants on Prolific Academic, but because of concurrent sign ups, 451 ended up participating. Sample size was determined by a target sample of at least 100 participants per experimental condition and funds available to pay. In the debriefing, participants were given the option to withdraw their data by typing "I would like to withdraw my data." Seven participants chose to do so, resulting in a final sample of 444 participants (M_{age} = 32.47 years, *SD* = 12.13; 208 female). This sample size gave us the ability to detect a small to medium effect size with 95% power (Faul et al., 2009).

3.1.2 Procedure. In a 2 (reputation manipulation: public *vs.* private) x 2 (team manipulation: team *vs.* solo) experimental design, participants were told that they would be competing in two rounds of trivia.

Participants were first randomly assigned to be in either the Team or Solo conditions. In the Team condition, participants were told that they would be competing as part of a team and that all members of the top scoring team would receive an additional $1.00 bonus. Team participants were told that there would be five teams total and that the team score would be the average score of all team members. Participants were then given the option to choose one of five teams to join, including *The Rebels*, *The Warriors*, *The Bosses*, *The All Stars*, and *The Legends*.

They were then welcomed to whatever team they joined in piped text (e.g., "Welcome to The Warriors.").

In the Solo condition, no teams were mentioned, and instead, participants were told that those who scored in the top 20% on the trivia tasks would receive an additional $1.00 bonus payment. Thus, in both the Team and Solo conditions, there was a one in five chance of winning the bonus payment.

Participants were then randomly assigned to be in either the Public or Private reputation conditions. In the Public condition, participants were told the following (note, bold indicates the text for the Team condition and italics indicates the text for the Solo condition):

> "Please select a username to keep track of your score on **your team scoreboard**/*the scoreboard*. Whatever name you select will be displayed to **your other team members**/*the other participants* on **your team scoreboard**/*the scoreboard*, which will display the scores of each **member of the team**/*participant* at the conclusion of the study. Your username must be 4-10 characters long. To protect your identity, we advise against using your full legal name."

They were then thanked by their username in piped text (e.g., "Thank you, crabbycakes.") and told the following:

> "Below, you can see the current **team rankings**/*participant rankings*. After you finish your two rounds of trivia, your ranking will be added to the **team scoreboard**/*scoreboard* and displayed to **your teammates**/*other participants*."

This text was followed by a scoreboard with 14 names, with the number 1 name shaded in gold, the number 2 name shaded in silver, and the number 3 name shaded in bronze. Those in

the Private condition simply proceeded to the trivia task without selecting a username or learning of the existence of a scoreboard.

Participants then proceeded to the first round of trivia. They were asked to please refrain from looking answers up and to simply do their best based on their current state of knowledge. The first round of trivia contained 10 multiple choice questions across a variety of topics (e.g., "What is the capital city of Florida?" and "Who discovered radium?", see Supplement for all trivia questions). This round was purposefully untimed so that participants could easily cheat by looking up answers on their phones or computers while participating.

Participants then proceeded to a second round of trivia, which contained 10 more multiple choice questions across a variety of topics. However, this time participants were told:

> "This round will be timed. You will have only 6 seconds per question. After 6 seconds have passed, the question will automatically advance, and if you have not answered, it will be counted as incorrect."

In pre-testing, 6 seconds was deemed just enough time to read the question and select an answer, but not enough time to look up an answer before the automatic advancement. If any differences were found on the Round 1 (time unrestricted) trivia between experimental conditions or between those who do or do not regularly participate in team competition, we would not know whether those differences were the result of cheating or of differential performance. By removing the possibility of cheating in Round 2, we could rule out the differential performance alternative explanation *if* differences were observed in Round 1, when cheating was possible, but not Round 2, when cheating was impossible. This enabled us to infer whether differences in Round 1 Trivia scores indeed reflect cheating behavior without relying on honesty in a self-reported cheating measure.

Participants then reported some demographics including age (options: 18-97), sex (options: male, female, other, prefer not to say), work status (options: employed, unemployed, prefer not to say), religious participation (options: strictly adheres to religious practices, loosely affiliated with a religion, non-religious, prefer not to say), political ideology (options: extreme left, center left, center, center right, extreme right, prefer not to say), and level of education (6-point scale from "high school or less" to "doctoral degree"). All demographic variables were scored starting with a value of 1 at the lowest level and then increased in increments of 1 at each level based on the order presented in the text above. Participants who responded "prefer not to say" or "other" on demographic questions were treated as missing values in the relevant demographic analyses. Demographics were collected as routine practice for purposes of knowing the characteristics of our sample. We conducted exploratory analyses testing for demographic differences in cheating behavior, but we had no hypotheses regarding these analyses. They are reported only for purposes of providing more information on predictors (or non-predictors) of cheating behavior and are not central to the present research.

Participants were asked if they fully read all study instructions and materials and were assured they would not be penalized for their answer. Only three participants reported that they did not, and these were not excluded. Regular team competitor status was determined by asking participants whether they regularly compete as a member of a sports team at any ability level (options: yes, no).

As a second measure of cheating, the last question prior to the debriefing was:

> "We understand that because this study took place online, it is possible participants could have looked up some of the answers during the trivia game (e.g., on Google). Did you look any answers up before submitting your responses, and if so, approximately how

many did you look up? Please be honest, you will not be penalized in any way for how you respond to this question."

Response options were on a 6-point scale, including "No, I looked up 0 answers," "Yes, I looked up 1 answer," "Yes, I looked up 2-3 answers," "Yes, I looked up 4-5 answers," "Yes, I looked up 6-7 answers," and "Yes, I looked up 8 or more answers." Validating our two measures of cheating, correlational analyses revealed that self-reported cheating was positively related to Round 1 Trivia scores (in which cheating was possible), $r = .23$, $p < .001$, but unrelated to Round 2 Trivia scores (in which cheating was impossible), $r = -.05$, $p = .312$.

3.2 Results

3.2.1 Cheating. We conducted two 2 (team manipulation: team, solo) x 2 (reputation manipulation: public, private) x 2 (team competitor status: yes, no) Univariate Analysis of Variances (ANOVAs) on Round 1 Trivia Scores and self-reported cheating behavior. As can be seen in Table 1 and Figure 1, the only significant main effects for both dependent variables were for regular team competitor status. Those who regularly participate in team competition scored higher in Round 1 Trivia (when cheating was possible) and self-reported more cheating behavior than those who do not.

The Reputation manipulation was trending in the hypothesized direction, with participants in the Public condition scoring slightly (although not statistically significantly) higher on Round 1 Trivia and self-reporting very slightly higher cheating behavior than those in the Private condition.

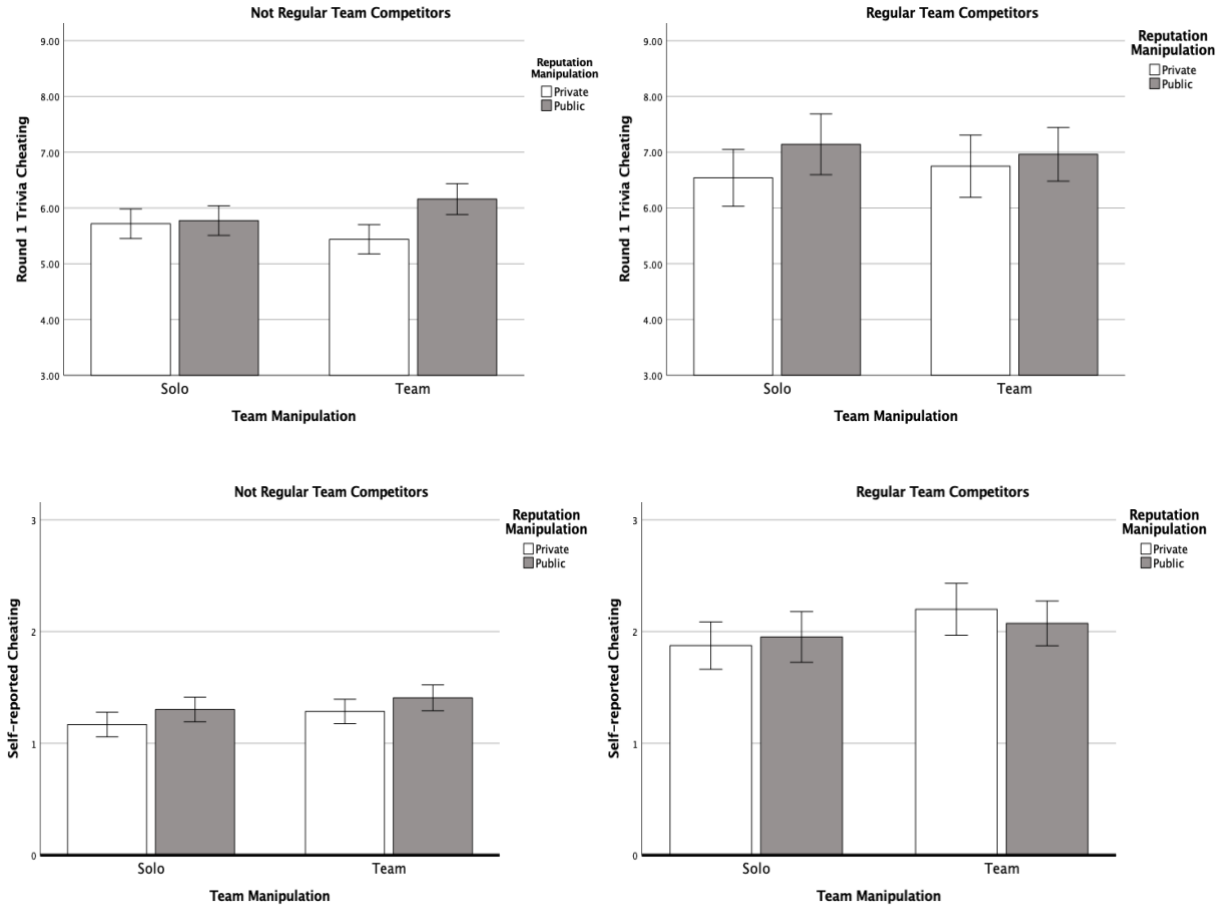
The Team manipulation was also trending in the hypothesized direction, with participants in the Team condition self-reporting slightly (although not statistically significantly) higher cheating behavior than those in the Solo condition and scoring very slightly higher on Round 1

Trivia. However, the latter difference was very near zero. And there was little to no evidence for any interactions with the Team manipulation.

Table 1. *Main effects for team manipulation, reputation manipulation, team competitor status, and all interactions in Study 1*

	Round 1 Trivia			Self-reported Cheating		
	F	*p*	η_p^2	*F*	*p*	η_p^2
Team Manipulation	0.01	.909	.000	1.85	.174	.004
Reputation Manipulation	1.82	.178	.004	0.18	.672	.000
Competitor	13.32	<.001	.030	35.84	<.001	.076
Team x Reputation	0.06	.815	.000	0.20	.659	.000
Team x Competitor	0.00	.948	.000	0.21	.646	.000
Reputation x Competitor	0.00	.975	.000	0.39	.534	.001
Team x Reputation x Competitor	0.80	.372	.002	0.15	.698	.000

Figure 1. *Means and standard errors by team manipulation (x-axis), reputation manipulation (colored bars), and team competitor status (columns) for Round 1 Trivia (top row) and Self-reported Cheating (bottom row) in Study 1*



Note. Y-axis was truncated for ease of visualization. Top row Round 1 Trivia scores could have ranged up to 10 and bottom row Self-reported Cheating could have ranged up to 6.

3.2.2 Round 2 Trivia. We also conducted a 2x2x2 ANOVA on Round 2 Trivia scores (in which cheating was impossible). The only significant effect was for the individual difference, team competitor status, $F(1, 434) = 15.41$, $p < .001$, $\eta_p^2 = .034$, with team competitors performing *worse* on Round 2 Trivia ($M = 2.33$, $SD = 1.48$) than non-team competitors ($M = 3.07$, $SD = 1.60$). These results can increase our confidence that team competitors' better performance in Round 1 was indeed due to cheating (consistent with their self-reported cheating) and not better trivia performance, because they actually performed worse than non-team competitors when cheating was impossible.

3.2.3 Demographic predictors of cheating and trivia performance. Older age was unrelated to Round 1 Trivia performance and self-reported cheating, *r*s (442) < .10, *ps* > .100, and negatively related to Round 2 Trivia performance, *r* (442) = -.21, *p*<.001. Increasingly lower religiosity (higher on scale) was associated with worse Round 1 Trivia performance, *r* (442) = -.19, *p*<.001, less self-reported cheating, *r* (442) = -.26, *p*<.001, and better Round 2 Trivia performance, *r* (442) = .23, *p*<.001. More right-wing ideology was associated with slightly better Round 1 Trivia performance, *r* (442) = .08, *p*<.001, more self-reported cheating, *r* (442) = .11, *p* = .027, and worse Round 2 Trivia performance, *r* (442) = -.28, *p*<.001. More education was associated with better Round 1 Trivia performance, *r* (442) = .27, *p*<.001, more self-reported cheating, *r* = .19, *p*<.001, and worse Round 2 Trivia performance, *r* (442) = -.16, *p* = .001. And men self-reported more cheating, *p* =.007, and slightly outperformed women in Round 1 Trivia, *p* =.052, but not Round 2 Trivia, *p* =.320. These patterns suggest higher religiosity, more right-wing ideology, higher level of education, and being male were all associated with more cheating.

3.3 Discussion

The results of Study 1 provided very little support for our hypothesis that people would be more likely to cheat when their performance would be displayed publicly than when it would only be known privately. The means for both cheating measures were in the hypothesized direction, but these differences were not statistically significant.

We also did not (conceptually) replicate the finding that team incentives increase cheating relative to individual incentives (Conrads et al., 2013), although our methods differ from Conrads and colleagues' study in numerous ways. Again, the means were in the hypothesized direction, but differences were very small and not statistically significant.

The most robust finding was that regular team competitors were more likely to cheat than those who do not regularly compete as part of a team. This suggests that the kinds of people attracted to team competition are more willing to engage in unethical behavior to inflate their own competence.

Because the results were trending in the hypothesized direction, but not statistically significant, we sought to replicate Study 1 with a larger and higher quality sample in Study 2 to increase confidence either that (1) the small effects are in fact null, or (2) the small effects are simply that—small effects.

4. Study 2

Study 2 replicated Study 1 with a larger and a higher quality sample by using CloudResearch options to block duplicate IP addresses, restrict participation to those in the United States (to reduce the likelihood that participants did not have a mastery of the survey language, English), block suspicious geocode locations and verify worker country location, and include only CloudResearch approved participants (participants who have shown prior evidence of attentiveness).

4.1 Method

4.1.1 Participants. We recruited 800 participants on CloudResearch, but because of concurrent sign ups, 803 ended up participating (M_{age} = 39.35 years, *SD* = 12.75; 355 female). Sample size was determined by a target sample as large as possible within financial constraints. Participants were again given the option to withdraw their data at the end, but none chose to do so. This sample size gave us the ability to detect smaller effect sizes than in Study 1, but still in the small-medium range (Faul et al., 2009).

4.1.2 Procedure. Procedures were identical to Study 1, with the exceptions that some of the team name options, instructions, and trivia questions were revised in unimportant ways (see supplement), and "prefer not to say" options were removed from the demographics (though participants had the ability to skip questions). Only 1 participant reported that they did not fully read the instructions, and they were not excluded. Again validating our measures of cheating, correlational analyses revealed that self-reported cheating was positively related to Round 1 Trivia scores (in which cheating was possible), $r = .31$, $p < .001$, and actually somewhat negatively related to Round 2 Trivia scores (in which cheating was impossible), $r = -.07$, $p = .043$.

4.2 Results

4.2.1 Cheating. As can be seen in Table 2 and Figure 2, with this larger and higher quality sample, consistent with our initial hypothesis, there were statistically significant effects of the reputation manipulation on both Round 1 Trivia scores and self-reported cheating, such that those in the Public condition scored higher than those in the Private condition and also self-reported more cheating. Replicating the results of Study 1, regular team competitors scored higher in Round 1 Trivia (when cheating was possible) and also self-reported more cheating than

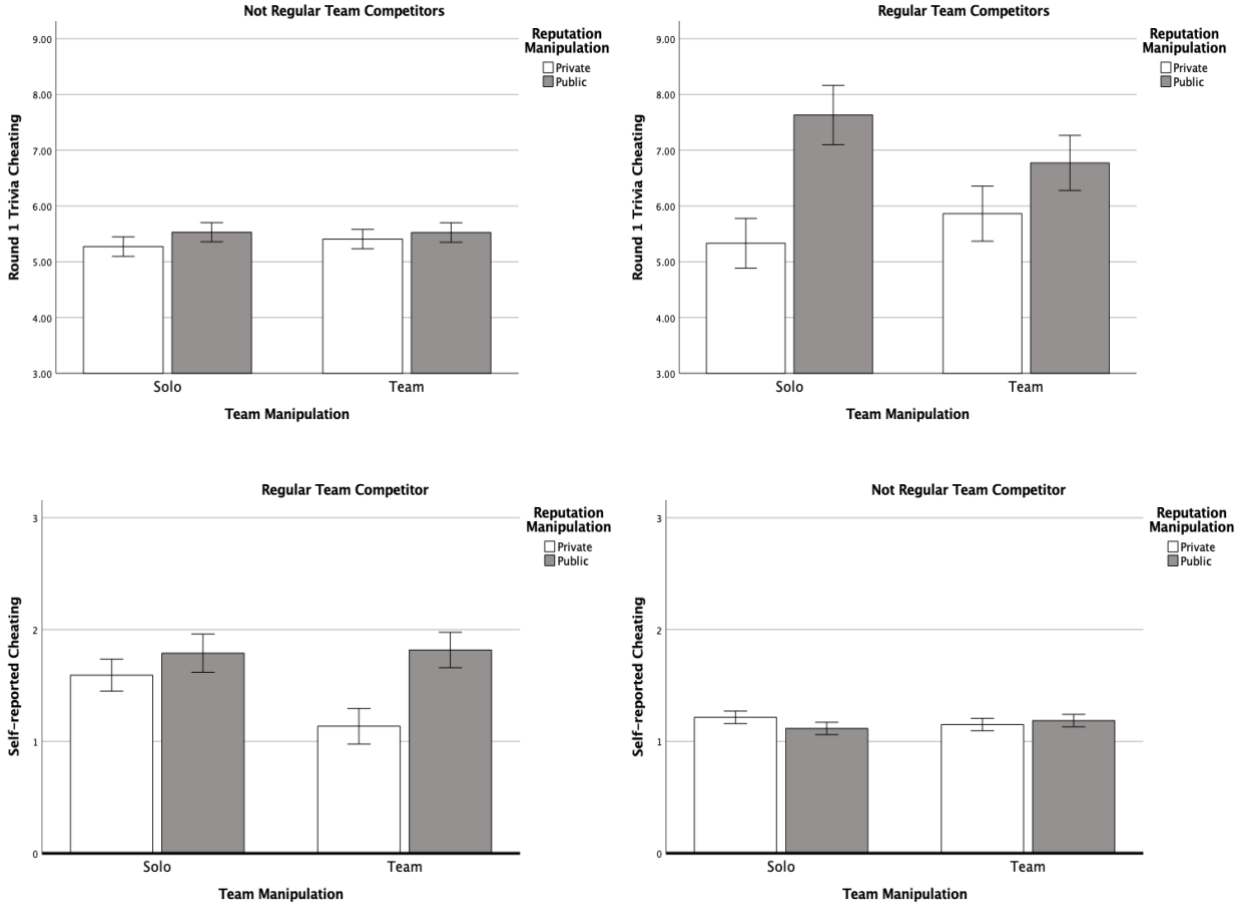
non-team competitors. Last, there were significant interactions between the reputation manipulation and team competitor status on both Round 1 Trivia scores and self-reported cheating such that the reputation manipulation had a larger effect among team competitors than non-team competitors. Meta-analyses of these effects are reported following the results of Study 2.

The Team manipulation continued to produce no significant differences, and thus we can be reasonably confident that this manipulation (at least as we executed it) had little to no influence on cheating behavior.

Table 2. *Main effects for team manipulation, reputation manipulation, team competitor status, and all interactions in Study 2*

	Round 1 Trivia			Self-reported Cheating		
	F	*p*	η_p^2	*F*	*p*	η_p^2
Leadership						
Team Manipulation	0.04	.849	.000	1.59	.208	.002
Reputation Manipulation	11.79	<.001	.015	5.91	.015	.007
Competitor	13.72	<.001	.017	24.76	<.001	.030
Team x Reputation	2.15	.143	.003	3.43	.064	.004
Team x Competitor	0.19	.660	.000	1.67	.197	.002
Reputation x Competitor	7.37	.007	.009	7.92	.005	.010
Team x Reputation x Competitor	1.43	.232	.002	1.09	.297	.001

Figure 2. *Means and standard errors by team manipulation (x-axis), reputation manipulation (colored bars), and team competitor status (columns) for Round 1 Trivia (top row) and Self-reported Cheating (bottom row) in Study 2*



Note. Y-axis was truncated for ease of visualization. Top row Round 1 Trivia scores could have ranged up to 10 and bottom row Self-reported Cheating could have ranged up to 6.

4.2.2 Round 2 Trivia. As in Study 1, regular team competitors performed *worse* on Round 2 Trivia (M = 3.23, SD = 1.68) than those who do not regularly participate in team competition (M = 4.04, SD = 1.73), $F(1, 795) = 17.77$, $p < .001$, $\eta_p^2 = .022$. And the Reputation manipulation had no influence on Round 2 Trivia, $F(1, 795) = 0.03$, $p = .869$, $\eta_p^2 = .000$. Thus, higher scores in Round 1 Trivia among those in the Public reputation condition and among regular team competitors likely do not reflect superior trivia ability but rather cheating (consistent with the self-reported cheating).

4.2.3 Demographic predictors of cheating and trivia performance. Older age was associated with slightly worse Round 1 Trivia performance, r (803) = -.08, p =.031, and unrelated to self-reported cheating and Round 2 Trivia performance, ps > .100. Increasingly lower religiosity (higher on scale) was unrelated to Round 1 Trivia performance, p > .100, and related to slightly less self-reported cheating, r (802) = -.08, p = .036, and better Round 2 Trivia performance, r (802) = .20, p<.001. More right-wing ideology was unrelated to Round 1 Trivia performance and self-reported cheating, ps > .100, and related to worse Round 2 Trivia performance, r (802) = -.12, p = .001. More education was associated with better Round 1 Trivia performance, r (803) = .11, p = .002, and Round 2 Trivia performance, r (803) = .10, p = .004, but unrelated to self-reported cheating, p > .100. And men performed better on Round 1 and Round 2 Trivia, ps < .016, but did not self-report more cheating, p = .736. Overall then, most of the demographic differences found in Study 1 were not replicated in Study 2 with a larger and higher quality sample, except that higher religiosity may have been associated with more cheating (as evidenced by higher self-reported cheating, worse performance on Round 2, and attenuated worse performance on Round 1).

5. Mini Metas and Summary Analyses

5.1 Method

5.1.1 Fixed effect Meta-analyses

As a last step, we conducted six mini meta-analyses on the main effect of the Reputation manipulation on Round 1 Trivia scores and Self-reported Cheating, the main effect of regular team competitor status on Round 1 Trivia scores and Self-reported Cheating, and the interaction between the Reputation manipulation and regular team competitor status on Round 1 Trivia scores and Self-reported Cheating. Mini meta-analyses are recommended in papers containing more than one study as a means of obtaining more stable estimates of true effect sizes and testing the statistical significance of these estimates (Goh et al., 2016).

We included one effect size for each of our two samples across Studies 1 and 2 for a total of two effect sizes and 1,245 participants per meta-analysis. We followed procedures outlined by Goh, Hall, and Rosenthal (2016). For all effects, the *F*-values from the ANOVAs were converted to *r* effect sizes, which were then Fisher's *Z* transformed to r_Zs. Because study procedures were virtually identical across samples, we conducted fixed effect meta-analyses, which weight by sample size using the formula: Weighted $\bar{r}_Z = \Sigma$ ([N-3] r_Z) / Σ (N-3). The $\bar{r}_Z$s were then converted back to *r* effect sizes for ease of presentation. Confidence intervals were computed with META-MAR V2.7.0 (Beheshti, 2020). To estimate statistical significance, we used the Stouffer's Z test, in which the *p* values for each effect size were converted to *Z*s, combined using the formula: $Z_{combined} = \Sigma Z$ / sqrt(k), and then converted back to *p*s.

5.1.2 Random effects Meta-analyses

At the request of a reviewer, we also conducted eight supplemental random-effects meta-analyses testing the overall effect of the Reputation manipulation and team competitor status as well as the Reputation manipulation within each level of team competitor status on Round 1

Trivia and Self-reported Cheating. We used the *M*s, *SD*s, and *n*s to compute *r* effect sizes and conducted single sample *t*-tests on these effect sizes. Random effects meta-analyses are appropriate when methods or samples vary between studies. Although our methods were virtually identical and both studies used samples of online adults, one sample was drawn from Prolific Academic and the other from Cloud Research, and so it is unclear whether the two studies should be treated as nearly identical or different. The gender compositions of the two samples were nearly identical, but the latter sample had a mean age ~7 years older. Given the ambiguity of whether fixed effects or random effects is more appropriate, we retained our original meta-analyses and supplemented them with this approach. However, random-effects meta-analyses are generally underpowered to detect statistical significance for mini-metas (Guolo & Varin, 2017; Seide et al., 2019)—indeed some scholars have argued that they are not appropriate when there are fewer than five studies (Jackson & Turner, 2017), and we have only two.

5.2 Results

5.2.1 Fixed effect Meta-analyses Results

There were statistically significant main effects for the Reputation manipulation on Round 1 Trivia scores, $r = .10$, $p = .0007$, 95% CI [.05, .16] and Self-reported Cheating, $r = .06$, $p = .043$, 95% CI [.01, .12]. There were significant main effects for regular team competitor status on Round 1 Trivia scores, $r = .15$, $p < .0001$, 95% CI [.09, .20], and Self-reported Cheating, $r = .21$, $p < .0001$, 95% CI [.16, .26]. And there was some evidence for a (small) interaction between the Reputation manipulation and regular team competitor status, with a marginal effect for Round 1 Trivia scores, $r = .06$, $p = .054$, 95% CI [.01, .12], and a significant effect for Self-reported Cheating, $r = .07$, $p = .015$, 95% CI [.02, .13].

5.2.1 Random effects Meta-analyses Results

In the random-effects analysis, the effect sizes for the Reputation manipulation and team competitor status were similar, however, because such analyses are inevitably underpowered (only two values per analysis), no effects were statistically significant. The effect for the Reputation manipulation was small on Round 1 Trivia scores, r = .08, p = .079, 95% CI [-.05, .21] and small to trivial on Self-reported Cheating, r =.04, p = .395, 95% CI [-.28, .35]. The effect for regular team competitor status on Round 1 Trivia scores was small, r = .15, p = .109, 95% CI [-.17, .46], and the effect for Self-reported Cheating was small to medium, r = .23, p = .153, 95% CI [-.47, .93]. That these effects were significant in the fixed effects approach and not significant in the random effects approach likely reflects two things: (1) the effects for the Reputation manipulation were very small and (2) random effects models are underpowered to detect statistical significance.

Four random effects mini metas within subgroups of team competitor status similarly showed no significant effects. The effects of the Reputation manipulation among non-team competitors on Round 1 Trivia scores, r = .06, p = .205, 95% CI [-.19, .31], and Self-reported Cheating, r = .05, p = .344, 95% CI [-.33, .43], were small to trivial, and the effects of the Reputation manipulation among team competitors on Round 1 Trivia scores, r = .20, p = .34, 95% CI [-1.27, 1.66], and Self-reported Cheating, r = .08, p = .500, 95% CI [-.88, 1.03], were small to medium.

6. General Discussion

The present work found that public performance feedback (relative to private performance feedback) led to small increases in cheating behavior on a competence task. Consistent with evolutionary theories of reputation management (e.g., Anderson & Kilduff,

2009a, 2009b; Fehr, 2004; Gächter & Fehr, 1999; Engelmann & Fischbacher, 2008; Henrich & Gil-White, 2001; Van Vugt & Hardy, 2010; Vonasch et al., 2018), these results suggest people may be willing to engage in unethical behavior to display false competence to others. Insofar as cheating always carries some reputational risk—one can never be 100% certain they will "get away" with cheating—these findings demonstrate that people make reputational tradeoffs, willing to put their moral reputations at risk to preserve their competence reputations.

The observed effect was small to very small. However, the "public" in our studies were online strangers whom participants would never meet in real life, and the only marker of participants' identities were online usernames made up for purposes of the study. Thus, it is possible that this effect would be larger or perhaps quite a bit larger when the stakes are larger and more real (e.g., Hilbig & Thielmann, 2017), with consequences for one's reputation among well-known others (Delton et al., 2011). Future work should explore similar effects in more naturalistic settings to better estimate the effect size in the real world.

We also observed a consistent unpredicted effect: that those who participate regularly in team competitions cheated more than those who do not. These results suggest that the very sorts of people who might be attracted to team competition are more likely to cheat in such competitions (Whitley, 1998). Future research should explore why this is the case and other personality variables that might relate to how people weight moral reputation *vs.* competence reputation tradeoffs. For example, those who regularly participate in team competitions may be more competitive or status-striving, or perhaps they are higher in self-monitoring or more prone to impression management. There was also some evidence that the reputation manipulation increased cheating behavior *more* among those who regularly participate in team competitions. Thus there may be important individual differences in how public performance feedback

incentivizes unethical behavior. The mechanisms underlying this phenomenon require exploration in future studies.

We found little to no support for the hypothesis that competing as part of a team (*vs.* alone) influenced cheating behavior (with effect sizes very near zero). If this is a true null effect, future work should explore why this is the case. It is possible that team incentives do not increase cheating behavior (potentially contradicting results found by Conrads and colleagues [2013]). But it is also possible that there are hidden individual difference moderators, such as group identification, or environmental moderators, such as size of incentives. Working on a team might incentivize some people to display competence to others but might also incentivize others to social loaf (e.g., Karau & Williams, 1993; Schippers, 2014; Stark et al., 2017), producing an overall null effect. Another possibility is that our team manipulation was not strong enough and that with a more powerful manipulation, there would have been more cheating in the Team condition. Indeed, participants in the Solo condition (at least the subset in the Public condition) were also in a group context in the sense that they believed their scores would be visible to other participants. Even though they were not competing as a member of a team, participants may still have felt as though their social value was being evaluated by their co-participants. Any public performance may elevate concerns about one's apparent social value as a potential group member. Future research should explore these multiple possibilities for our null effect.

One additional not hypothesized pattern of findings was that religiosity appeared to be weakly associated with more cheating. In Study 1, higher scores (less religious) were associated with worse Round 1 Trivia performance (indicative of less cheating), less self-reported cheating, and better Round 2 Trivia performance (in which cheating was impossible). In Study 2, religiosity was not related to Round 1 Trivia performance, but there was again a small relation

between lower religiosity and lower self-reported cheating and better Round 2 performance. If the non-religious are better at trivia in general (as suggested by Round 2 performance) but less likely to cheat (as suggested by self-reported cheating), these two patterns could attenuate a relationship between religiosity and Round 1 Trivia. Exploring the relation between religiosity and cheating was not the purpose of the present research, so future work should test this pattern with methods more purposefully designed to explore the association between religiosity and performance enhancing cheating behavior. Such findings may contribute to decades long debates surrounding the relationship between religiosity and moral behavior and suggest the possibility that in some contexts, religiosity may be associated with *more* immoral behavior, not less.

The present research was limited in several respects. Although nearly all participants self-reported being attentive, they may not have been honest (especially in a study designed to provide opportunities for cheating) or they may not have been sufficiently motivated to care about the outcome of the trivia competition. Whether these effects would be larger or smaller in real world face-to-face competitive interactions remains to be investigated in future research. Furthermore, at present little is known about the mechanisms driving these effects. There are many questions left to be answered regarding when and why and in whom we observe this competence/morality tradeoff, and we hope the present work stimulates this line of research.

One big challenge for encouraging human greatness is that by incentivizing high performance, high competence, and higher productivity, we inevitably incentivize *faking* high performance, high competence, and higher productivity. Perhaps not all people would put their moral reputations at risk to display false competence to others, but some people would. Thus it is critical to understand how people balance these tradeoffs so we can work toward bringing out the best in people without also bringing out the worst.

References

Anderson, C., & Kilduff, G. J. (2009a). The pursuit of status in social groups. *Current Directions in Psychological Science*, *18*, 295-298.

Anderson, C., & Kilduff, G. J. (2009). Why do dominant personalities attain influence in face-to-face groups? The competence-signaling effects of trait dominance. *Journal of Personality and Social Psychology*, *96*, 491-503.

Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. *Science, 211,* 1390-1396.

Becker, G.S. (1974). Crime and punishment: An economic approach. In G.S. Becker & W.M. Landes (Eds.), *Essays in the economics of crime and punishment* (pp. 1-54), New York: National Bureau of Economic Research.

Beheshti, A., (2020). META-MAR V2.7.0.

Campbell, T. H., & Kay, A. C. (2014). Solution aversion: On the relation between ideology and motivated disbelief. *Journal of Personality and Social Psychology*, *107*(5), 809-824.

Christenson, D. P., & Kriner, D. L. (2017). Constitutional qualms or politics as usual? The factors shaping public support for unilateral action. *American Journal of Political Science, 61*(2), 335-349.

Claassen, R. L., & Ensley, M. J. (2016). Motivated reasoning and yard-sign-stealing partisans: Mine is a likable rogue, yours is a degenerate criminal. *Political Behavior*, *38*(2), 317-335.

Clark, C. J., Costello, T., Mitchell, G., & Tetlock, P. E. (2022). Keep your enemies close: Adversarial collaborations will improve behavioral science. *Journal of Applied Research in Memory and Cognition*.

Clark, C. J., Liu, B. S., Winegard, B. M., & Ditto, P. H. (2019). Tribalism is human nature. *Current Directions in Psychological Science, 28,* 587-592.

Clark, C. J., & Winegard, B. M. (2020). Tribalism in war and peace: The nature and evolution of ideological epistemology and its significance for modern social science. *Psychological Inquiry, 31,* 1-22.

Conrads, J., Irlenbusch, B., Rilke, R. M., & Walkowitz, G. (2013). Lying and team incentives. *Journal of Economic Psychology, 34,* 1-7.

Cohen, G. L. (2003). Party over policy: The dominating impact of group influence on political beliefs. *Journal of Personality and Social Psychology, 85*(5), 808-822.

Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. In J. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind* (pp. 163-228). New York: Oxford University Press.

Delmas, M. A., & Lessem, N. (2014). Saving power to conserve your reputation? The effectiveness of private versus public information. *Journal of Environmental Economics and Management, 67,* 353-370.

Delton, A. W., Krasnow, M. M., Cosmides, L., & Tooby, J. (2011). Evolution of direct reciprocity under uncertainty can explain human generosity in one-shot encounters. *Proceedings of the National Academy of Sciences, 108,* 13335-13340.

DeMarree, K. G., Clark, C. J., Wheeler, S. C., Briñol, P., & Petty, R. E. (2017). On the pursuit of desired attitudes: Wanting a different attitude affects information processing and behavior. *Journal of Experimental Social Psychology, 70,* 129-142.

Ditto, P. H., Liu, B. S., Clark, C. J., Wojcik, S. P., Chen, E. E., Grady, R. H., ... & Zinger, J. F. (2019a). At least bias is bipartisan: A meta-analytic comparison of partisan bias in liberals and conservatives. *Perspectives on Psychological Science*, *14*(2), 273-291.

Ditto, P. H., Clark, C. J., Liu, B. S., Wojcik, S. P., Chen, E. E., Grady, R. H., ... & Zinger, J. F. (2019b). Partisan Durkee, P. K., Lukaszewski, A. W., & Buss, D. M. (2020). Psychological foundations of human status allocation. *Proceedings of the National Academy of Sciences*, *117*, 21235-21241.

Engelmann, J. M., Herrmann, E., & Tomasello, M. (2012). Five-year olds, but not chimpanzees, attempt to manage their reputations. *PLoS One*, *7*, e48433.

Engelmann, D., & Fischbacher, U. (2009). Indirect reciprocity and strategic reputation building in an experimental helping game. *Games and Economic Behavior*, *67*, 399-407.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149-1160.

Fehr, E. (2004). Don't lose your reputation. *Nature*, *432*, 449-450.

Fehr, E., & Fischbacher, U. (2004). Social norms and human cooperation. *Trends in Cognitive Sciences*, *8*, 185-190.

Fehr, E., Fischbacher, U., & Gächter, S. (2002). Strong reciprocity, human cooperation, and the enforcement of social norms. *Human Nature*, *13*, 1-25.

Frimer, J. A., Skitka, L. J., & Motyl, M. (2017). Liberals and conservatives are similarly motivated to avoid exposure to one another's opinions. *Journal of Experimental Social Psychology*, *72*, 1-12.

Gächter, S., & Fehr, E. (1999). Collective action as a social exchange. *Journal of Economic Behavior & Organization*, *39*, 341-369.

Gampa, A., Wojcik, S. P., Motyl, M., Nosek, B. A., & Ditto, P. H. (2019). (Ideo) logical reasoning: Ideology impairs sound reasoning. *Social Psychological and Personality Science*, *10*(8), 1075-1083.

Goh, J. X., Hall, J. A., & Rosenthal, R. (2016). Mini meta-analysis of your own studies: Some arguments on why and a primer on how. *Social and Personality Psychology Compass*, *10*, 535-549.

Guolo, A., & Varin, C. (2017). Random-effects meta-analysis: the number of studies matters. *Statistical Methods in Medical Research*, *26*(3), 1500-1518.

Hardy, C. L., & Van Vugt, M. (2006). Nice guys finish first: The competitive altruism hypothesis. *Personality and Social Psychology Bulletin*, *32*, 1402-1413.

Hawkins, C. B., & Nosek, B. A. (2012). Motivated independence? Implicit party identity predicts political judgments among self-proclaimed independents. *Personality and Social Psychology Bulletin*, *38*(11), 1437-1452.

Henrich, J., & Gil-White, F. J. (2001). The evolution of prestige: Freely conferred deference as a mechanism for enhancing the benefits of cultural transmission. *Evolution and Human Behavior*, *22*, 165-196.

Hilbig, B. E., & Thielmann, I. (2017). Does everyone have a price? On the role of payoff magnitude for ethical decision making. *Cognition*, *163*, 15-25.

Kahan, D. M., Hoffman, D. A., Braman, D., Evans, D., & Rachlinski, J. J. (2012). They saw a protest: Cognitive illiberalism and the speech-conduct distinction. *Stan. L. Rev.*, *64*, 851.

Kahan, D. M., Peters, E., Dawson, E. C., & Slovic, P. (2017). Motivated numeracy and enlightened self-government. *Behavioural Public Policy, 1*(1), 54–86.

Kajackaite, A., & Gneezy, U. (2017). Incentives and cheating. *Games and Economic Behavior*, *102*, 433-444.

Kamis, D., Newmark, T., Begel, D., & Glick, I. D. (2016). Cheating and sports: history, diagnosis and treatment. *International Review of Psychiatry, 28*, 551-555.

Karau, S. J., & Williams, K. D. (1993). Social loafing: A meta-analytic review and theoretical integration. *Journal of Personality and Social Psychology, 65*, 681-706.

Kurzban, R., & Leary, M. R. (2001). Evolutionary origins of stigmatization: the functions of social exclusion. *Psychological Bulletin*, *127*, 187-208.

Jackson, D., & Turner, R. (2017). Power analysis for random-effects meta-analysis. *Research Synthesis Methods*, *8*(3), 290-302.

Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology, 37*(11), 2098–2109.

Milinski, M., Semmann, D., & Krambeck, H. J. (2002). Reputation helps solve the 'tragedy of the commons'. *Nature, 415*, 424-426.

Nietzsche, F. (1974). *The gay science: With a prelude in German rhymes and an appendix of songs* (W. Kaufmann, Trans.). New York, NY: Vintage. (Original work published 1887)

Nowak, M. A., & Sigmund, K. (1998). Evolution of indirect reciprocity by image scoring. *Nature, 393*, 573-577.

Palou, P., Ponseti, F. J., Cruz, J., Vidal, J., Cantallops, J., Borràs, P. A., & Garcia-Mas, A. (2013). Acceptance of gamesmanship and cheating in young competitive athletes in

relation to the motivational climate generated by parents and coaches. *Perceptual and Motor Skills*, *117*(1), 290-303.

Pettit, N. C., Doyle, S. P., Lount Jr, R. B., & To, C. (2016). Cheating to get ahead or to avoid falling behind? The effect of potential negative versus positive status change on unethical behavior. *Organizational Behavior and Human Decision Processes*, *137*, 172-183.

Preston, I., & Szymanski, S. (2003). Cheating in contests. *Oxford Review of Economic Policy*, *19*(4), 612-624.

Schippers, M. C. (2014). Social loafing tendencies and team performance: The compensating effect of agreeableness and conscientiousness. *Academy of Management Learning & Education*, *13*, 62-81.

Schwieren, C., & Weichselbaumer, D. (2010). Does competition enhance performance or cheating? A laboratory experiment. *Journal of Economic Psychology*, *31*, 241-253.

Seide, S. E., Röver, C., & Friede, T. (2019). Likelihood-based random-effects meta-analysis with few studies: empirical and simulation studies. *BMC Medical Research Methodology*, *19*(1), 1-14.

Sperber, D., & Baumard, N. (2012). Moral reputation: An evolutionary and cognitive perspective. *Mind & Language*, *27*, 495-518.

Stark, E. M., Shaw, J. D., & Duffy, M. K. (2007). Preference for group work, winning orientation, and social loafing behavior in groups. *Group & Organization Management*, *32*, 699-723.

Stroud, N. J. (2008). Media use and political predispositions: Revisiting the concept of selective exposure. *Political Behavior*, *30*(3), 341-366.

Stroud, N. J. (2010). Polarization and partisan selective exposure. *Journal of Communication, 60*(3), 556-576.

Šukys, S. (2013). Athletes' justification of cheating in sport: Relationship with moral disengagement in sport and personal factors. *Baltic Journal of Sport and Health Sciences, 3*(90), 70-77.

Taber, C. S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science, 50*(3), 755-769.

Tennie, C., Frith, U., & Frith, C. D. (2010). Reputation management in the age of the world-wide web. *Trends in Cognitive Sciences, 14*, 482-488.

Trivers, R. L. (1971). The evolution of reciprocal altruism. *The Quarterly Review of Biology, 46*, 35-57.

Van Vugt, M., & Hardy, C. L. (2010). Cooperation for reputation: Wasteful contributions as costly signals in public goods. *Group Processes & Intergroup Relations, 13*, 101-111.

Vonasch, A. J., Reynolds, T., Winegard, B. M., & Baumeister, R. F. (2018). Death before dishonor: Incurring costs to protect moral reputation. *Social Psychological and Personality Science, 9*, 604-613.

Whitley, B. E. (1998). Factors associated with cheating among college students: A review. *Research in Higher Education, 39*, 235-274.

Winegard, B., Kirsch, A., Vonasch, A., Winegard, B., & Geary, D. C. (2020). Coalitional value theory: An evolutionary approach to understanding culture. *Evolutionary Psychological Science, 6*, 301-318.

Wu, J., Balliet, D., & Van Lange, P. A. (2016a). Reputation management: Why and how gossip enhances generosity. *Evolution and Human Behavior, 37*, 193-201.

Wu, J., Balliet, D., & Van Lange, P. A. (2016b). Reputation, gossip, and human cooperation. *Social and Personality Psychology Compass*, *10*, 350-364.

Zaksaite, S. (2012). The interrelation of micro and macro factors that contribute to cheating in sports. *Sport & EU Review*, *4*(2), 9-23.

Zimniuch, F. (2009). *Crooked: A history of cheating in sports*. Lanham, MD: Taylor Trade Publications.