European Economic Review 170 (2024) 104844

Contents lists available at ScienceDirect

European Economic Review

journal homepage: www.elsevier.com/locate/eer

Lying in competitive environments: Identifying behavioral impacts[☆]

Simon Dato[a,*], Eberhard Feess[b], Petra Nieken[c]

[a] EBS University for Business and Law, EBS Business School, Rheingaustr. 1, 65375 Oestrich-Winkel, Germany
[b] Victoria University of Wellington, School of Economics and Finance, 23 Lambton Quay (Pipitea Campus), Wellington, New Zealand
[c] Karlsruhe Institute of Technology, Institute of Management, Chair of Human Resource Management, Kaiserstr. 89, 76133 Karlsruhe, Germany

ARTICLE INFO

JEL classification:
C90
D82
D91

Keywords:
Private information
Lying
Contest
Competition
Cheating

ABSTRACT

Incentive schemes based on relative performance provide high effort incentives, but may backfire by increasing incentives for misconduct. Previous literature supports this view by demonstrating that, as compared to individual incentive schemes based on absolute performance only, highly competitive environments are associated with higher degrees of lying and cheating. However, it is not clear if this is (mainly) driven by stronger financial incentives or by competition per se and its behavioral effects. We conduct an online experiment with competitive and individual incentive schemes in which the financial incentives to lie are held constant. From a behavioral perspective, a competitive environment may increase the willingness for misconduct via a desire-to-win, but may also decrease it via the negative payoff externality on competitors. Our results provide evidence of a significant lying-enhancing desire-to-win-effect and an insignificant lying-reducing negative externality effect.

1. Introduction

Albeit they are commonplace, incentive schemes based on the relative performance of employees are among the most controversial topics in managerial practice and academia (e.g., Lazear, 1989; Berger et al., 2013; Croson et al., 2015; Kampkötter and Sliwka, 2018). Competitive reward schemes are supposed to increase effort for two reasons. First, they set strong financial incentives due to the discontinuous upward jump in payoffs associated with a higher rank (Grote, 2005). Second, insights from social comparison theory (Festinger, 1954; Garcia et al., 2013) and findings from neuro-physiological research (Fliessbach et al., 2007; Dohmen et al., 2011) suggest that these higher financial incentives are reinforced by various psychological benefits from outperforming others, often summarized as a "desire-to-win" (Charness et al., 2014; Benistant and Villeval, 2019), that are specific to competitive environments.

As a downside, however, competitive schemes may not only lead to higher effort but may also reduce concerns towards morally questionable behavior (Schweitzer et al., 2004; To et al., 2020). Anecdotal evidence documents that employees cut corners to meet their targets and get promotions (Zoltners et al., 2016), allocate their time to window-dressing instead of productive activities (Mitchell et al., 2018; Corgnet et al., 2019), or even commit outright fraud (Brown et al., 2014; Association of Certified Fraud Examiners, 2020). However, as misconduct is also frequently observed with high-powered incentive schemes based on absolute rather than on relative performance measures (Ordóñez et al., 2009; Haßet al., 2015), it is far from obvious that a higher degree of misconduct in competitive compared to non-competitive payment schemes can be attributed to the desire-to-win effect: it might also

☆ Funded by Germany's Excellence Strategy – EXC 2126/1 – 390838866 and by the Federal Ministry of Education and Research (BMBF) and the Baden-Wuerttemberg Ministry of Science as part of the Excellence Strategy of the German Federal and State Governments. The experiments reported in this paper have received the approval from the Ethics Committee of EBS Business School.

* Corresponding author.
E-mail addresses: simon.dato@ebs.edu (S. Dato), eberhard.feess@vuw.ac.nz (E. Feess), petra.nieken@kit.edu (P. Nieken).

https://doi.org/10.1016/j.euroecorev.2024.104844
Received 31 May 2023; Received in revised form 18 August 2024; Accepted 23 August 2024
Available online 26 August 2024

be purely driven by differences in expected financial benefits. Our paper sheds light on this question by designing and conducting an online experiment where all financial impacts, both for the individuals able to commit misconduct and for others, are identical with competitive and non-competitive payment schemes. If we still find a difference in misconduct, then this difference can safely be attributed to a "desire-to-win" effect.

The experimental literature models competitive payment schemes as contests, and finds (almost) consistently that the degree of misconduct is higher than with piece rates (Carpenter et al., 2010; Faravelli et al., 2015; Benistant et al., 2021). However, in all experimental studies we are aware of, the financial incentives for misconduct differ between contests and individual reward schemes (such as piece rates). If contests are not too unbalanced, then the discrete jump sets higher incentives for effort as well as for misconduct, and the utility-maximizing behavior also depends largely on the expectations about the behavior of other contestants (Konrad, 2009). This implies that a simple comparison between the behavior in contests and non-competitive schemes does not allow identifying the psychological impacts of competition.

To the best of our knowledge, our paper is the first to account for this by designing an experiment where the financial benefit from misconduct is identical with and without competition. Specifically, subjects in our online experiment take part in a binary lottery with the outcomes LOW and HIGH, observe the lottery outcome privately, and then report the outcome. In all treatments discussed below, reporting HIGH leads to a higher expected payoff, which provides financial incentives to misreport a privately observed LOW outcome as HIGH. In our contest treatment *C*, two subjects compete in a simultaneous winner-take-it-all contest. The one who reports HIGH (LOW) receives the winner prize (the loser prize). If both subjects submit identical reports, the winner is determined by a random draw. Note that the payoff structure implies that the expected financial benefit from lying is independent of the other contestant's report, as the probability when announcing HIGH instead of LOW increases from 0% to 50% after HIGH and from 50% to 100% after LOW.

We then consider an individual treatment *I* without competition, in which the expected monetary benefit from lying is the same as in treatment *C*. This is achieved by implementing the same prize structure and resembling the opponent's behavior by a random draw. This procedure ensures that the expected increase in the own payoff from lying in *I* is the same as in *C*. However, we then need to go one step further by taking into account that competitive and non-competitive reward schemes may differ not only in their expected financial benefits from misconduct, but also in another important dimension: Due to the zero-sum character of contests, lying in a contest inevitably reduces the payoff of another subject in the experiment. Assuming that most subjects have rather other-regarding than spiteful preferences, and hence put positive weight on other subjects' payoffs, such a negative externality ceteris paribus reduces the willingness to lie. Hence, if we do not find a difference in lying between treatments *C* and *I*, this would not allow the conclusion that there is no desire-to-win effect: it may also be the case that this effect is offset by the negative externality effect.

We account for this issue with our third treatment, which we refer to as the negative externality treatment *N*. This treatment is identical to treatment *I*, except that each active subject is matched with a passive subject ("a bystander") who receives the high payment if and only if the active subject receives the low payment. All financial effects are then identical to treatment *C*, as inflating the outcome of the lottery yields the same expected financial benefit and the same negative payoff externality. The only difference remaining is that there is no competition in treatment *N*, as there is no other subject who might inflate their outcome.

The piecewise comparisons of the treatments allow us to identify the following effects: First and foremost, comparing the lying frequencies in treatments *C* and *N* shows, of course restricted to our experimental framework, whether there is a desire-to-win effect. This comparison is hence crucial for our research question. Second, comparing the two treatments *N* and *I* isolates the negative externality effect. Third, comparing treatments *C* and *I* shows whether the desire-to-win effect from competition is stronger or weaker than the negative externality effect.

Our design ensures that the expected financial benefit from lying is identical across treatments, and independent of the beliefs about the behavior of other subjects. Nevertheless, one might worry that a comparison of the treatments *C* and *N* is affected by differences in beliefs because we use a random draw to simulate the opponent in treatment *N*. We conducted the treatment *N-belief* to address this concern. Instead of using a random draw that resembles the opponent's behavior, we inform subjects that their payoff will be determined by their own report and a report from a randomly drawn participant from a previous treatment *C* session. While this previous participant will not be paid based on the outcome, a passive subject from the current session will receive the high or low payment depending on the game's outcome as in treatment *N*. Thus, subjects in the treatment *N-belief* need to form beliefs just as subjects in treatment *C*.

Note that the desire-to-win effect and the negative externality effect arise from outcome-based preferences, that is, decision-makers evaluate outcomes rather than the underlying actions. Recent experimental evidence indicates that decisions also depend on whether the underlying actions are seen as socially appropriate (Krupka and Weber, 2013; Kimbrough and Vostroknutov, 2016; Barr et al., 2018; Chang et al., 2019). This observation is potentially important for our findings, as our hypotheses rest on the assumption that the social inappropriateness of lying is not (substantially) treatment-dependent. To see why this might matter for our treatment comparisons, suppose lying is seen as less appropriate in treatment *N* compared to *I*, as it imposes a negative externality on other individuals. The higher degree of social inappropriateness then makes reporting HIGH less attractive in *N*. A possible treatment effect between *I* and *N* could then, next to altruism, also be traced back to a preference for norm compliance. Likewise, lying in *C* could be evaluated as less inappropriate than in *N*, because the subject one is matched with can lie as well.

To account for this, we conducted the *Norms* treatment to test whether the social inappropriateness of lying varied across our three main treatments. The data corroborates that lying, compared to reporting a low outcome truthfully, is considered much less appropriate. This documents that a general social norm of not lying exists. More importantly, the difference in the appropriateness ratings between misreporting a low outcome as HIGH and reporting truthfully is very similar and not significantly different across treatments. This provides evidence that the treatment effects cannot be attributed to treatment-dependent differences in the social appropriateness of lying.

Our main results from the treatments *C*, *I*, *N*, and *N-belief* are as follows: First and foremost, 49.6% of all subjects report HIGH in treatment *N*, compared to 57.4% in treatment *C*. This difference is economically meaningful and statistically significant. In addition, we observe a share of HIGH reports of 52.6% in the treatment *N-belief*, which is not significantly different from the share in the treatment *N*. Hence, average lying behavior does not depend on whether one has to form beliefs about the stakes of the game or not. Given that we collected the data for our *N-belief* treatment at a later point in time, we conducted additional sessions of the *C* treatment to account for potential confounding effects when comparing old and new data. In line with the comparison of *N* and *C*, we find that the frequency of high reports in *N-belief* is economically meaningful and statistically significantly lower than in the newly conducted treatment *C-new*. Our experiment hence provides evidence for a desire-to-win effect that reduces moral concerns towards lying: When the financial consequences from lying are identical both for the decision maker and others, then playing against a human who can decide to lie increases the lying frequency compared to a situation where payoffs depend on the own decision and affect another human which has no decision power.

Second, comparing treatments *I* and *N* allows carving out the negative externality effect: Both treatments have identical monetary incentives and differ only in that lying yields a negative externality on the bystander in *N*. Comparing the frequency of high reports in treatments *I* (53.9%) and *N* shows that the negative externality causes a moderate yet statistically insignificant reduction in cheating.

Third, the frequency of HIGH reports is higher in the contest scheme *C* (57.4%) than in the individual scheme *I* (53.9%). but the difference is not significant. The comparisons *C* vs. *N* and *N* vs. *I* are crucial for our main research question, i.e., to identify (i) the desire-to-win and (ii) the negative externality effect. The comparison of treatments *C* and *I* captures the net effect of competition and is therefore interesting from a more applied perspective. Although tentatively suggesting a positive net effect, our experimental results do not allow the conclusion that, for constant financial incentives, competition leads to more lying.

For at least two reasons, however, it seems plausible that the reduction in moral concerns triggered by the desire-to-win effect identified in our experimental framework underestimates the impact in reality. First, all psychological motives underlying the desire-to-win effect discussed in the literature are likely to be more important when competing in real effort tasks compared to competing on a payoff from a lottery (see Piest and Schreck, 2021, for an overview): Outperforming others in a challenging task matters more for the own self-image and reputation effects towards others than succeeding in a lottery contest. Second, our anonymous online experiment leaves hardly any room for what the literature refers to as rivalry "that is characterized by the experience of heightened psychological stakes of competition by the focal actor when competing against the target actor" (Kilduff et al., 2016, p. 1509). Both the experimental literature and field data show that rivalry tends to further reduce moral concerns compared to anonymous competitive settings (Pierce et al., 2013; Kilduff et al., 2016; To et al., 2020). Therefore, our approach to identifying the desire-to-win effect in an online experiment with prizes based on lottery outcomes is conservative. Accordingly, our approach might underestimate the difference in lying between competitive and individual incentive schemes in reality.

Our main contribution is to provide, based on the prediction from a stylized behavioral model, experimental evidence of a desire-to-win effect. Importantly, our theoretical prediction is robust to several extensions: allowing for individual heterogeneity in the strength of the desire-to-win or lying costs to be dependent on the opponent's report in the contest leaves the prediction qualitatively unchanged. Not only is the desire-to-win effect theoretically robust, but the experimental effect is also very unlikely to emerge through different channels. First, we elicit the social norm of lying for each treatment separately and show that it is not treatment-dependent. Second, we discuss several alternative explanations such as loss aversion and image concerns, and show that they cannot explain the significantly higher share of high reports in *C* than in *N*.

The remainder of the paper is organized as follows: Section 2 relates to the literature. We present a simple model in Section 3. Section 4 describes the experimental design, procedures, and our hypotheses. Results are shown in Section 5. In Section 6 we put our findings into perspective by discussing several possible extensions and alternatives to our model and design. We conclude in Section 7.

2. Related literature

Our paper is most closely related to experiments comparing misconduct in competitive and non-competitive treatments. The earlier literature considers real-effort tasks (see the overview by Chowdhury and Gürtler, 2015). Schwieren and Weichselbaumer (2010) use the maze game introduced by Gneezy et al. (2003) and Belot and Schröder (2013) let subjects identify euro coins, and Faravelli et al. (2015) use the matrix task developed by Mazar et al. (2008). Schwieren and Weichselbaumer (2010) compare the individual piece rate treatment to a contest of six subjects, in which only the one who reports the highest number of solved mazes is paid. Overall, they do not find a significant difference between the cheating behavior in the two treatments, but low-performing subjects lie significantly more in the contest than with piece rates. Belot and Schröder (2013) compare piece rates to a four-player contest. The contest winner receives a prize of 50 euro, whereas the other three contestants get nothing. They find that both the productive effort and the lying frequency are significantly higher in the contest. Faravelli et al. (2015) compare piece rates to a two-player contest. Cheating is more frequent in the contest, but this effect disappears when subjects can self-select to the piece rate or the contest treatment.

Most of the experiments just discussed suggest that competitive remuneration systems lead to more misconduct than simple bonus schemes. In contrast to our experiment, however, the financial benefits from misconduct differ between the contest and the piece rate settings. With piece rates, the marginal financial benefit of misconduct is constant and independent of the behavior of all other subjects in the experiment. Conversely, the marginal benefit from misconduct in a contest depends on the number and the behavior of other contestants. These differences in the incentive structures are likely to contribute to the different findings in the literature: In Schwieren and Weichselbaumer (2010), the marginal benefit from cheating in the contest might be perceived as rather low because just one out of six contestants are paid. The fact that Belot and Schröder (2013) find more cheating in the competitive environment might hence be due to the lower number of contestants and the large winner prize of 50 euros. Faravelli et al. (2015) consider only two contestants. The contest also entails a piece rate component, as the winner gets $2 per correctly solved matrix, compared to $1 in the individual piece rate setting. The main difference to our comparison of treatments C and I is hence that the marginal expected financial benefit from cheating differs between treatments.[1]

Most of the recent literature builds on the die-under-the-cup paradigm introduced by Fischbacher and Föllmi-Heusi (2013), which we adopt as well. Subjects roll a die in private, and the payoff structure is designed to induce a strong financial incentive to misreport the outcome.[2] As lying is unobservable, it needs to be studied at an aggregated level. Several recent papers utilize lotteries in the spirit of Fischbacher and Föllmi-Heusi (2013) to compare two-player contests. The advantage of the lottery setting compared to real effort tasks is that the degree of misconduct cannot be influenced by the subjects' abilities and effort costs. Dato et al. (2019) consider a sequential contest with and without lying possibility for the first subject. They find no significant treatment effect on the second subject's lying behavior. The same holds in Dannenberg and Khachatryan (2020), who compare simultaneous contests, in which either both or just one subject can lie. Benistant et al. (2021) find that the lying frequency in a contest is significantly larger than with piece rates if and only if both contestants can lie. The latter two papers derive a rich set of results,[3] but the marginal benefit from lying again differs between contests and piece rates. In Dannenberg and Khachatryan (2020), the results entered by passive subjects are systematically below those of subjects who can lie,[4] which changes the incentive structure of the contestant who can lie. In addition, a subject who rolls a die without the possibility to lie may be seen as a competitor. The latter argument also refers to Dato et al. (2019), where marginal financial benefits from lying are identical across all contest treatments.

Charness et al. (2014) consider a dynamic real-effort rank-order tournament with flat wages so that all treatments are identical with regards to the financial incentives to cheat. They find that informing subjects about their ranks increases their effort, which reinforces the view that ranking systems may be beneficial in this respect.[5] Furthermore, subjects who are informed about their rank engage in cheating and sabotage. Our identification strategy of the behavioral impacts of competition on misconduct differs in many important respects: First, Charness et al. (2014) do not compare the cheating behavior with information on ranks to a treatment without information so that it cannot be excluded that subjects would have cheated even without information on ranks due to, e.g., self-image concerns or to reduce their anger about a task they disliked. Interpreting ranks as competition, there is hence no comparison of our treatment C to another treatment.[6] Second, while flat wages ensure that differences in treatments are not driven by different financial incentives, we are interested in comparing bonus contracts to competitive remuneration schemes, which would be impossible with flat wages. Third, we compare three treatments to tease out the impact of the negative externality implied by competition.

Benistant and Villeval (2019) analyze a two-player simultaneous real-effort tournament. The lying behavior is neither affected by group identity nor by whether lying increases the own or decreases the opponent's final score. Several papers find that lying is likely to be reinforcing, as subjects who underestimate (overestimate) the lying frequency lie more (less) when they are informed about the actual numbers (Le Maux et al., 2021; Bäker and Mechtel, 2019; Casal et al., 2017; Diekmann et al., 2015).[7] In addition to lying about the own outcome, the literature also considers the possibility of sabotaging the competitors' outcomes. In the seminal paper by Carpenter et al. (2010), sabotage occurs more frequently in contests.[8] Harbring and Irlenbusch (2011) and Conrads et al. (2014) find that sabotage and lying, respectively, increase in the prize spread.[9] These findings reinforce our view that the monetary incentives need to be kept constant to identify the behavioral impacts of competition.

While we introduce a second player to identify the impact of competition, other papers introduce a second player to determine the effects of groups. Conrads et al. (2014) compare an individual piece-rate treatment to a treatment where the two members of a group decide independently on their report and share their payoff equally. Lying is more frequent in the group treatment. A comparable result is found in Danilov et al. (2013) in an experiment with professionals from the financial services sector, provided

[1] In Faravelli et al. (2015), the payment per correctly solved maze is, on average, $1 both in the contest and with piece rates. Marginal financial incentives to cheat, however, are quite different, as those depend in the contest on (i) the own performance, (ii) the own willingness to cheat, and (iii) the expectation on the other contestant's report.

[2] Dai et al. (2018) document that the behavior in the die-under-the cup paradigm provides a good predictor of cheating in the field. For a meta-study on this paradigm with non-strategic set-ups, see Abeler et al. (2019).

[3] Dannenberg and Khachatryan (2020) compare individual to group contests, and Benistant et al. (2021) focus on the impact of feedback and incentives on the lying behavior in dynamic settings. Also, considering a dynamic framework, Necker and Paetzel (2023) find that the lying frequency of strong performers in a real-effort task increases when they learn that they are matched with other strong performers.

[4] The reported outcomes could only be identical if no one lies.

[5] Gill et al. (2019) extend the analysis to a multi-period setting. They find that providing information about the rank has the highest positive effect on effort for subjects at the top and the bottom of the ranking.

[6] However, in individual settings without competition, Charness et al. (2019) find no evidence of cheating in a die-roll task if reports have no impact on payoffs.

[7] Feltovich (2019) frames the decision situation as markets and compares lying in monopolies and different kinds of duopolies. While the marginal financial benefit is highest in the monopoly treatment, the lying frequencies in the duopoly tend to be rather higher than lower. This also suggests a behavioral impact of competition.

[8] As in the papers just discussed, the expected marginal financial benefit from the misconduct differs among treatments.

[9] Dato and Nieken (2014) find that sabotage frequencies of men exceed those of women.

that group identity is prominent. Kocher et al. (2018) find more lying in groups, and Dannenberg and Khachatryan (2020) show that the group effect is more pronounced in competitive settings.

Summing up, while there is a large body of literature that compares cheating and lying in treatments with and without competition, we are not aware of any other paper that keeps both the expected marginal financial benefit from misconduct and the impact on others constant across treatments.

3. The model

To derive the utility-maximizing lying frequencies under competitive and individual incentive schemes, and to disentangle the impact of a *desire-to-win* and the *negative externality* in competition, we analyze the following simple model.

Player i takes part in a lottery, which yields a high outcome $x_i = h$ with probability p_i and a low outcome $x_i = l$ with $1 - p_i$. Player i privately observes x_i and then reports $r_i \in \{l, h\}$. Misreporting the actual outcome by reporting $r_i \neq x_i$ yields (internal) lying costs of c. The report influences player i's monetary payoff, which is either high, w_H, or low, w_L. Player i derives material utility from money according to an increasing function $u(w)$ with $u(w_L) = u_L < u_H = u(w_H)$. We consider three settings. In all settings, player i's probability of receiving w_H instead of w_L increases by 50 percentage points when reporting h instead of l.

Two players $i = 1, 2$ compete with each other in the *Contest* setting C. Both players privately observe the realization of their (independent) lotteries and report the outcome. If only one player reports the high outcome, she receives w_H, while the other player receives w_L. If both reports are identical, a random draw determines who of the two players obtains w_H and w_L, respectively.[10]

Next to the additional material utility from winning denoted by $\Delta u = u_H - u_L$ and lying costs c, player i's objective function is affected by the following two motives. First, winning the contest provides an additional non-monetary utility $\hat{u} > 0$ that can be interpreted as a "desire-to-win" or "competitiveness". Results from experimental economics (Brookins and Ryvkin, 2014; Sheremeta, 2010; Cooper and Fang, 2008) as well as neuroeconomics (Dohmen et al., 2011; Delgado et al., 2008) provide evidence that non-monetary motives shape the evaluation of a competition's outcome.

Second, recall that the competitor receives w_L in case player i wins the contest and receives w_H. Therefore, by reporting h instead of l, player i reduces the utility of the competing player j in two respects: First, she imposes a negative externality on the other player's expected monetary payoff. Second, she reduces the probability that player j may enjoy her non-monetary utility $\hat{u}$ from winning the contest. We assume that player i has social preferences and puts relative weight $\phi \in (0, 1)$ on the other player's utility.

Note that, after observing the high outcome, it is optimal to report h. This directly follows from (i) positive lying costs ($c \geq 0$) and (ii) the weaker regard for the opponent than for herself ($\phi < 1$).[11] We can thus restrict attention to the situation where player i has drawn a low outcome, $x_i = l$. Suppose the other player j submits $r_j = l$ with probability π, then player i's utility of truthfully reporting the low outcome is given by

$$
\begin{aligned}
U_i^C(l) &= \pi \left\{ \frac{1}{2} \left[u_L + \phi \left(u_H + \hat{u} \right) \right] + \frac{1}{2} \left(u_H + \hat{u} + \phi u_L \right) \right\} + (1 - \pi) \left[u_L + \phi \left(u_H + \hat{u} \right) \right] \\
&= u_L + \frac{\pi}{2} (\Delta u + \hat{u}) + \phi \left[u_L + \left(1 - \frac{\pi}{2} \right) (\Delta u + \hat{u}) \right] ,
\end{aligned}
$$

whereas misreporting the low outcome as high yields

$$
\begin{aligned}
U_i^C(h) &= \pi \left(u_H + \hat{u} + \phi u_L \right) + (1 - \pi) \left\{ \frac{1}{2} \left[u_L + \phi \left(u_H + \hat{u} \right) \right] + \frac{1}{2} \left(u_H + \hat{u} + \phi u_L \right) \right\} - c \\
&= u_L + \frac{1}{2} (1 + \pi) (\Delta u + \hat{u}) + \phi \left[u_L + \frac{1 - \pi}{2} (\Delta u + \hat{u}) \right] - c.
\end{aligned}
$$

Comparing the expected utilities shows that player i lies if and only if

$$
c < (1 - \phi) \frac{\Delta u + \hat{u}}{2} \equiv \tilde{c}_C.
$$

Importantly, our framework ensures that the expected financial benefit from reporting HIGH instead of LOW, and hence also the threshold $\tilde{c}_C$, are independent of the probability π that the other player reports HIGH. It follows that each player has a dominant strategy, which is choosing (i) $r_i = h$ if $x_i = h$ and (ii) $r_i = l$ ($r_i = h$) if $x_i = l$ for $c \geq \tilde{c}_C$ ($c < \tilde{c}_C$). These dominant strategies constitute the Nash equilibrium of the game.

In the *Negative Externality* setting N, player i takes part in the same lottery as in C, privately observes the realization, and reports the outcome. Conversely to setting C, however, there is no strategic interaction as her payoff does not depend on the action of another player. Instead, N is a setting of individual decision-making: the probability to obtain w_H is determined by player i's report and two random draws. With probability q, a low report $r_i = l$ leads to a 50/50-lottery between w_L and w_H, whereas the high report $r_i = h$ yields w_H with certainty. With probability $1 - q$, $r_i = l$ yields w_L with certainty, while $r_i = h$ results in the 50/50-lottery between w_L and w_H. Setting N eliminates the competitive nature of setting C so that a desire-to-win does not affect player i's report. The negative externality inherent to competition, however, is maintained: there is a passive individual who receives the low (high)

[10] When interpreting the reports as effort which translates into output directly, this setup resembles the canonical tournament model of Lazear and Rosen (1981) with discrete effort and zero-variance of noise. Having a model without noise is crucial to ensure that financial incentives are independent of the other contestant's report, and identical across settings.

[11] We discuss downward lying and channels through which it might be optimal in Section 6.3.

monetary payoff if player i receives the high (low) monetary payoff. Therefore, social preferences still influence player i's report. The utility of truthfully reporting the low outcome is then

$$U_i^N(l) = u_L + \frac{q}{2}\Delta u + \phi\left[u_L + \left(1 - \frac{q}{2}\right)\Delta u\right],$$

while misreporting the low outcome as high yields

$$U_i^N(h) = u_L + \frac{1}{2}(1+q)\Delta u + \phi\left(u_L + \frac{1-q}{2}\Delta u\right) - c.$$

Comparing the expected utilities shows that player i lies in setting N if and only if

$$c < (1-\phi)\frac{\Delta u}{2} \equiv \tilde{c}_N.$$

Note that the threshold $\tilde{c}_N$ is independent of the probability q, as the impact of q in treatment N resembles the effect of π in setting C: with probability q $(1-q)$, a player in N faces the same financial consequences as a player in C does, given that the other contestant reports LOW (HIGH). Hence, as the expected benefit from lying does not depend on the other player's behavior in C, neither does the expected benefit from lying in N depend on the outcome of the random draw.

The *Individual* setting I is identical to N except that there is no passive individual whose payoff depends on the player's decision. As social preferences are muted, truthfully reporting the low outcome yields expected utility

$$U_i^I(l) = u_L + \frac{q}{2}\Delta u,$$

while misreporting the low outcome as high yields

$$U_i^I(h) = u_L + \frac{1}{2}(1+q)\Delta u - c.$$

Player i hence lies if and only if

$$c < \frac{\Delta u}{2} \equiv \tilde{c}_I.$$

Comparing the thresholds for player i's lying decision in the three settings yields the following proposition.

Proposition 1. *(i) The thresholds $\tilde{c}$ for lying costs c such that player i lies if and only if $c < \tilde{c}$ are larger in settings I and C compared to setting N, $\tilde{c}_C, \tilde{c}_I > \tilde{c}_N$. (ii) The threshold is higher in setting C than in setting I if and only if $\phi < \frac{\hat{u}}{\hat{u}+\Delta u}$.*

Proof. *Part (i).* $\tilde{c}_I - \tilde{c}_N = \phi\frac{\Delta u}{2} > 0$. $\tilde{c}_C - \tilde{c}_N = (1-\phi)\frac{\hat{u}}{2} > 0$. *Part (ii).* $\tilde{c}_C - \tilde{c}_I = (1-\phi)\frac{\hat{u}}{2} - \phi\frac{\Delta u}{2}$ decreases in ϕ. As $\tilde{c}_C - \tilde{c}_I = 0 \Leftrightarrow \phi = \frac{\hat{u}}{\hat{u}+\Delta u}$, the result follows. ■

The intuition for Proposition 1 is as follows: In all settings, misreporting the low outcome as high increases the probability to obtain the high monetary payoff by 50 percentage points. The expected financial (or material) benefits from lying are thus identical across settings. The only difference of setting N to setting I is that player i's decision is also affected by her social preferences towards the passive player j. This reduces her incentive to lie; thus $\tilde{c}_I > \tilde{c}_N$. Next, the only difference between setting C to setting N is that there is also a utility from winning the contest. As player i puts higher weight on her own than on player N's utility, this increases the threshold; hence $\tilde{c}_C > \tilde{c}_N$ (*Part (i)* of the Proposition). *Part (ii)* of the Proposition shows that it depends on the importance of social preferences whether the incentives to lie are larger in settings C or I. If social preferences ϕ are sufficiently large, $\phi > \frac{\hat{u}}{\hat{u}+\Delta u}$, then the threshold $\tilde{c}$ (and hence the lying frequency) is lower in the contest. But if ϕ is low, then the "desire-to-win" dominates the negative externality effect.

4. Experimental design and hypotheses

We designed three treatments that resembled the settings discussed in the previous section.[12] Subjects in treatment C participated in a simultaneous two-player contest and were randomly matched with another subject. Each subject is instructed to roll a six-sided die and report the outcome, where a result of the die roll from one to four (five or six) translates into a LOW (HIGH) outcome. If both subjects reported the same outcome, each of them received the high bonus (winner prize) of 1.20 GBP with a probability of 50%: note that if one subject received the winner prize, the other subject received the loser prize of 0.20 GBP. Otherwise, the one who reported HIGH (LOW) got the winner prize with certainty. The impact of the paired subjects' reports on the contest prizes are summarized in Table 1. All subjects knew that, after the experiment, they would be informed about the report of their competitor and the resulting payment.

Importantly, the financial incentives to lie do not depend on whether the opponent reports HIGH or LOW. While different lying frequencies lead to different likelihoods of competing against a high or low report, the probability of receiving the winner instead of the loser prize increases by 50 percentage points irrespective of the other player's report: If the other player reports LOW, then reporting HIGH instead of LOW yields an increase in receiving the winner prize from 50% to 100%, and if the other player reports

[12] The instructions for all treatments are relegated to the Appendix A.

Table 1
Overview of bonus payments in treatment *C*.

Report of the other participant	Your report	Bonus
Low	Low	Each of you has a 50% chance of getting the 1.20 GBP. This is decided by a random draw.
	High	You: 1.20 GBP Other: 0.20 GBP
High	Low	You: 0.20 GBP Other: 1.20 GBP
	High	Each of you has a 50% chance of getting the 1.20 GBP. This is decided by a random draw.

Table 2
Overview of bonus payments in treatment *N*.

Case	Your report	Bonus
1 45% probability	Low	Each of you has a 50% chance of getting the 1.20 GBP. This is decided by a random draw.
	High	You: 1.20 GBP Other: 0.20 GBP
2 55% probability	Low	You: 0.20 GBP Other: 1.20 GBP
	High	Each of you has a 50% chance of getting the 1.20 GBP. This is decided by a random draw.

HIGH, then reporting HIGH instead of LOW yields an increase in receiving the winner prize from 0% to 50%. Note that this implies that subjective beliefs about the opponent's behavior do not affect financial incentives to lie. Therefore, the incentives are identical across subjects in treatment *C* even if they hold different subjective beliefs on the other contestants' propensity to lie.

In our second treatment *N*, there is only one active player. As in treatment *C*, this player's expected payoff always increases by 50 percentage points when reporting HIGH instead of LOW. Also as in treatment *C*, lying decreases the probability that another participant gets the winner instead of the loser prize by 50 percentage points. However, there is no competitor in treatment *N* who decides upon lying. Instead, the other participant is a passive "bystander" who gets the winner (loser) prize if the active player gets the loser (winner) prize. Thus, active players in treatment *N* did not act in a competitive environment, but the impact of their report on other subjects' bonus payments resembled treatment *C*: The financial consequences of lying are the same, but there is no competition. Comparing the behavior of treatments *C* and *N* hence allows us to isolate the impact of competition.

To resemble treatment *C* as closely as possible, we distinguish between two cases for the active player. In case 1, the active player received 0.20 GBP or 1.20 GBP with 50% probability each when reporting LOW, and 1.20 GBP for sure when reporting HIGH. This case 1 mirrored the situation in treatment *C* when the other contestant reported LOW. In case 2, the active player receives 0.20 GBP with certainty with a LOW report, and 0.20 GBP or 1.20 GBP with 50% probability each with a HIGH report. Case 2 thus mirrored treatment *C* when the other contestant reported HIGH. To assign probabilities for these two cases that are as close as possible to treatment *C*, we executed one session of treatment *C* with 100 subjects upfront. 55 subjects reported HIGH and 45 reported LOW, so that we implemented a probability of 45% that the active player in treatment *N* was in case 1, and a probability of 55% that the active player was in case 2.[13] Subjects received no information about the origin of the probabilities for each case. They were informed about the probabilities for the two cases but did not learn which case they were actually in before submitting their report. After the experiment, they were informed about the case they had been in and the resulting payment. The impact of the report and the random draws (that is, whether the active player is in case 1 or 2) on the subjects' bonus payments are summarized in Table 2.

Note that, from an incentive perspective, it does not make a difference for subjects with lying costs and social preferences whether they are in case 1 or 2, because the expected impact of lying is always the same (both in case 1 and in case 2, lying increases the probability of receiving the money by 50%). This independence also holds when subjects are inequity averse à la Fehr and Schmidt (1999). The reason for this independence is that the inequity is always the same, as, inevitably, one player gets the winner and one the loser prize. The reason why we nevertheless resembled treatment C as closely as possible is that there might be psychological reasons other than social preferences or inequity aversion as to why subjects perceive an increase from 0% to 50% in case of lying differently than an increase from 50% to 100%. To eliminate this potentially confounding factor, we chose probabilities for treatment *N* that mirrored the competitor's behavior in treatment *C*.[14]

Treatment *I* was identical to treatment *N* except that there was no bystander. In treatment *I*, we hence eliminated not only competition (as in treatment *N*), but also the impact of the own behavior on the payoff of someone else, which is still present in

[13] The remaining observations for treatment *C* have been collected together with the two other treatments in a randomized within-session format. The share of high reports is not significantly different between the first and the remaining sessions of the *C* treatment ($p = 0.716$, two-sided Fisher's exact), so the first session predicted average behavior very well.

[14] To even further address this concern, we introduce and discuss the treatment *N-belief* in Section 5.3.2.

Table 3
Overview of bonus payments in treatment *I*.

Case	Your report	Bonus
1 45% probability	Low	You have a 50% chance of getting the 1.20 GBP. This is decided by a random draw.
	High	1.20 GBP
2 55% probability	Low	0.20 GBP
	High	You have a 50% chance of getting the 1.20 GBP. This is decided by a random draw.

treatment *N*. Similar to the original set-up of Fischbacher and Föllmi-Heusi (2013), a subject's report solely determined the own expected bonus. The probabilities for cases 1 and 2 in treatment *I* were identical to those in treatment *N* (see Table 3).

In each treatment, subjects had to answer four control questions before rolling a die and reporting an outcome. Each control question addressed one of the possible cases. For each case, subjects were asked for the probability of receiving the high bonus. If subjects failed to give the correct answer, they could try a second time again before seeing the correct answer.

After submitting their report, subjects in treatments *C*, *I* , and *N* were asked for their belief about the behavior of other subjects in their treatment. Our question reads "What do you think about the behavior of the other participants in this study? Out of all participants (except you) whose actual results of the die roll was LOW (outcome 1 to 4), how many will report HIGH?" Beliefs were stated on a scale from zero to 100%.[15] In addition, we used the Honesty-Humility subscale from the HEXACO to measure fairness, sincerity, and greed avoidance (Ashton and Lee, 2009) ,[16] and measured positive and negative reciprocity following Dohmen et al. (2009). Finally, we asked for sex, age, country of residence, education level, and the number of studies they participated in on the online platform during the last 12 months. We also included an attention check into the Honesty-Humility survey stating "it is important that you pay attention to this study. Please tick disagree".

We ran the sessions with the passive subjects (the bystanders) in treatment *N* after having collected the reports from the active subjects. We refer to the data collected from bystanders as treatment *B*. As the bystanders did not have to make any payoff-relevant decision, we elicited their belief about the misconduct in one of the three treatments *C*, *I*, and *N*. Each bystander received the instructions of either treatment *C*, *I*, or *N*, and was asked to state the belief about the frequency of lying.[17] After stating the belief, all bystanders learned how their bonus was calculated, i.e., they were informed about the procedures from treatment *N* and their role as passive bystanders.

In *Norms*, we examined whether or not lying is assessed differently across treatments. In doing so, we followed the approach of Krupka and Weber (2013). A different set of subjects were given the instructions of one of our three main treatments and asked to rate the social appropriateness of (i) reporting LOW and (ii) reporting HIGH if the actual outcome of the die roll was LOW. Each report had to be rated as *very socially inappropriate*, *socially inappropriate*, *somewhat socially inappropriate*, *somewhat socially appropriate*, *socially appropriate*, or *very socially appropriate*. Each subject was then randomly paired with another subject, who rated the reports from the same main treatment. One of the two possible reports HIGH and LOW was randomly drawn for each pair, and the pair's ratings of that report were compared. If the ratings matched, both received a bonus of 2.50 GBP and zero otherwise. After submitting their ratings, all subjects filled out the same survey as in the other treatments (except for the belief question).

4.1. Sample and procedures

We preregistered our study in the AEA RCT Registry, and the digital object identifier (DOI) is: "10.1259/rct.6824–1.0". We executed our experiments online on the Prolific platform for several reasons. First, we needed a large sample size, as lying is unobservable, and our dependent variable is the share of high reports. Second, subjects needed to be sure that their actual outcome was unobservable. Whereas this is straightforward online, even with clear-cut instructions it might be doubted by some subjects in a classical lab situation. Third, we preferred to collect a sample with subjects differing across age, education, and sex to increase the generalizability of our results.

Prolific is a large online platform where people can participate in research and business studies. We announced a scientific study and a survey on individual decision-making. To ensure high data quality, we required subjects to be fluent in English, to reside in either UK or USA, be at least 18 years old, and have an approval of at least 95%. All subjects were allowed to participate just once. We implemented measures to prevent the restarting of the survey and self-selection into treatments. We informed the subjects that the study took about fifteen minutes and involved filling out a short survey and rolling a die. Subjects also knew the two possible bonus payment levels. If subjects were interested in participating, they followed a link taking them to the first page of our study

[15] For various reasons, we chose against incentivizing the elicitation of beliefs. As we did not observe the actual distribution of results, we would have to use the theoretically predicted distribution to calculate an approximation of actual lying behavior. It is even more critical to keep financial incentives across treatments constant and avoid possible confounds arising from, e.g., treatment differences in the degree of lying estimation complexity.

[16] We used a seven-point scale instead of the original five point-scale.

[17] They were asked precisely the same question as the active subjects in the respective treatment.

(hosted on Qualtrics). This first page was a consent form and only subjects giving their consent entered the study. To avoid that they needed to wait for each other in treatment *C*, we did not play the contest in a live interaction. Instead, subjects in all treatments were informed about the experiment's outcome within two days after participating.

1509 subjects participated in our study in total. We aimed for 300 subjects in each treatment.[18] The randomization was done by Qualtrics which automatically assigned incoming participants to treatments. 292 observations for the treatments *N* and *I*, and 318 for treatment *C*. The treatment *B* has 303 and the treatment *Norms* 303 observations. The data was collected between December 9 and 22, 2020.[19]

We excluded two subjects who did not pass the attention check from the analysis. Recall that subjects had two attempts to answer the control questions in the three main treatments. Overall, 89.22% of all subjects answered all four questions correctly at least after the second attempt and 70% even after the first attempt. Given that we provided a table with the corresponding payments and that subjects had, for each control question, only three options to choose from, it seems reasonable to assume that subjects (97) answering at least one question incorrectly twice did either not understand our set-up or did not pay close attention. This suggests excluding these subjects from the analysis. This view is reinforced by the fact that the percentage of subjects answering incorrectly twice differs among treatments; it is highest with 14.83% in treatment *I* and lowest with 6.29% in treatment *C*. There is no significant difference between treatments *I* and *N* ($p = 0.273$ in a Fisher's exact test), but the percentage of subjects failing to answer correctly even after the second attempt is significantly lower in treatment *C* ($p = 0.001$ compared to treatment *I* and $p = 0.022$ compared to treatment *N*). As these subjects may confound our treatment comparisons, we exclude them and focus our analysis on the sample of 803 observations for our three main treatments (298 for treatment *C*, 247 for treatment *I*, and 258 for treatment *N*).[20] We will refer to this sample as the *main sample* from hereon. To check for the robustness of our results, we will also consider a *restricted sample*, containing only those subjects who correctly answered all questions already in the first attempt. Table A.1 in the Appendix A provides an overview of the number of observations per treatment, demographics, and the share of subjects that answered all four control questions correctly after the first or second try.

4.2. Hypotheses

The comparison of the critical thresholds $\tilde{c}$ between the three settings *C*, *I*, and *N* in Section 3 yield the following hypotheses regarding the behavior of subjects in our experiment:

Hypothesis 1 (*Desire-To-Win Effect*)**.**The frequency of high reports in the negative externality treatment *N* is lower than in the contest treatment *C*.

Both treatments *C* and *N* share identical financial incentives and comprise a negative externality. A subject's payoff, however, depends on the report of some other subject only in *C*: the desire-to-win inherent to such a competition causes a stronger inclination to lie in *C*.

Hypothesis 2 (*Negative Externality Effect*)**.** The frequency of high reports in the negative externality treatment *N* is lower than in the individual treatment *I*.

The treatments *I* and *N* only differ in the negative externality on some other subject that a subject's high report gives rise to in treatment *N*. Social preferences of subjects lead to a lower inclination to lie in *N*. Recall that, due to countervailing effects, we have no hypothesis for the comparison of treatment *C* and *I*.

5. Results

This section provides the experimental test of the hypotheses on the consequences of behavioral impacts on lying in competitive environments. First, we analyze our main variable of interest, the fraction of high reports in the *C*, *I*, and *N* treatments. We then explore the role of subjects' beliefs about lying behavior and demonstrate the robustness of our results with the help of regression analysis and the results of the *N-belief* treatment. Finally, we investigate the social norm of lying.

---

[18] Our main variable of interest is the share of high reports which varies between zero and 100%. From the meta-analysis in Abeler et al. (2019) we expected 28% of subjects that see a low outcome to lie and report high in the *I* treatment. This would result in a baseline effect of 61 percent high reports. A power calculation with a total sample size of 600 (we compare the outcome between two treatments), a power of 0.8, and an alpha of 0.5 leads to a minimum detectable effect size of 0.1128 (two sample Chi-Square test).

[19] We executed a small pilot with 30 subjects in treatment *C* to test our software and set-up. This data is not included in the study. Note that we ran a session with 100 subjects in treatment *C* to collect information for the probabilities of the situations in the other treatments. The rest of the sessions have been executed with a within-session randomization approach for the main treatments. We ran the sessions for the bystanders in treatment *N* separately because these subjects did not have to roll a die.

[20] Recall that we asked four comprehension questions about the design of the treatment and implemented one additional attention check to ensure high data quality. In our preregistration, we stated that we would drop observations that failed all IMC questions and/or spent less than 60 s on the study. No subject took less than 60 s to complete the study. Unfortunately, we did not specify if *failed* means answering correctly in the first attempt. In addition, we did not expect the different levels of comprehension between treatments which might confound the analysis. To avoid stating results that are partly driven by noise due to a lack of understanding of the underlying game, we, therefore, decided to deviate from the preregistered exclusion criteria.

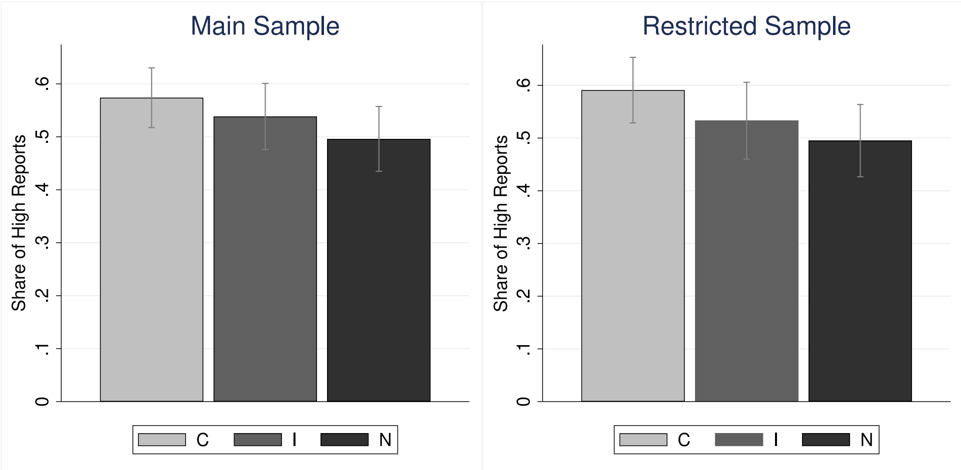


Fig. 1. Share of high reports by treatments for main and restricted sample.

5.1. High reports

Fig. 1 shows the percentage of high reports in treatments C, I, and N for the main sample (left panel) and the restricted sample (right panel). In any case, the share of high reports significantly exceeds 33%, the expected share of high reports under truth-telling ($p < 0.01$, two-sided t-test). In the main sample, the share of high reports is higher in C (57.38%) than in N (49.61%). The difference is statistically significant ($p = 0.073$), supporting Hypothesis 1.[21] This result provides evidence that the desire-to-win fosters lying.[22] Hypothesis 2 states that the share of high reports should be higher in I than in N. In line with Hypothesis 2, we observe high reports more frequently in I (53.85%) than in N, but as the difference is not statistically significant ($p = 0.373$), our results do not support the negative externality effect.[23] Still, the effect goes in the predicted direction, and is strong enough to render the difference in the share of high reports between C and I statistically insignificant ($p = 0.436$). Accordingly, when holding all financial incentives constant, subjects in our experiment do not behave differently in the competitive and the individual treatment.

When considering only subjects who answered the comprehensive questions correctly in the first attempt, we observe a very similar pattern as in the main sample (see the right panel of Fig. 1). In the restricted sample, the shares of high reports in treatments I (53.29%) and N (49.51%) are virtually identical to the main sample, while the corresponding share in C increases to 59.09%. Notwithstanding the lower number of observations, the level of significance of the desire-to-win effect on lying increases when comparing C and N ($p = 0.046$). The differences in the share of high reports between I and N ($p = 0.477$) and C and I ($p = 0.237$) remain statistically insignificant. Overall, all the results from the main sample also prevail with the restricted sample, which documents the robustness of our results. Furthermore, it appears that more mindful subjects care more about winning a contest, as the desire-to-win effect is stronger in the restricted sample.

5.2. Beliefs

We asked subjects for their belief about the percentage of other participants reporting HIGH when the actual outcome is LOW. There are two main insights: first, the ranking of beliefs among the different treatments coincides with the actual behavior. Subjects in treatment N expect a lower share of liars (53.36%, main sample; 52.39%, restricted sample) than subjects in treatment I (main sample 54.42%, $p = 0.2951$; restricted sample 54.71%, $p = 0.1780$) or treatment C (main sample 57.14%, $p = 0.0297$; restricted sample 57.25%, $p = 0.0175$).[24] The second insight is that subjects overestimate the actual degree of lying in all treatments. To see this, recall that the share of high reports contains lies *and* truthful high reports. In all three treatments, the average belief about the lying propensity is higher than 50%. A belief of 50% translates into a share of high reports of roughly 67%, while the actual shares of high reports are below 60% in all treatments.

[21] Our results are a lower bound for the desire-to-win effect. First, the desire-to-win might be more pronounced if the players exert effort in the contest. Second, in the N treatment subjects might perceive they are playing against the computer instead of a human participant. Albeit to a lower extent, the desire-to-win could then also affect lying behavior in treatment N.

[22] Note that the employed test statistics on the difference between the number of high reports in C and N are conservative for two reasons. First, a considerable fraction of subjects (approx. 33%) in both treatments observe HIGH. This reduces the ratio of observed high reports between treatments and, hence, leads to an underestimation of differences in lying across treatments. Second, although we have a specific prediction about the direction of the difference, we employ a two-sided Fisher's exact test in the paper unless stated otherwise.

[23] Fischbacher and Föllmi-Heusi (2013) also conducted a treatment with a negative externality. In line with our results, they find a statistically insignificant reduction in the lying frequency.

[24] All p-values for the comparison of beliefs are for Wilcoxon rank-sum tests.

Table 4
Determinants of lying frequencies. Probit regressions with the report as the dependent variable.

	Main sample			Restricted sample		
	(1)	(2)	(3)	(4)	(5)	(6)
C treatment	0.196*	0.235**	0.208*	0.242**	0.284**	0.255**
	(0.107)	(0.109)	(0.109)	(0.119)	(0.122)	(0.123)
I treatment	0.106	0.138	0.132	0.095	0.091	0.082
	(0.112)	(0.114)	(0.115)	(0.128)	(0.130)	(0.131)
Belief share liars			0.006***			0.006***
			(0.002)			(0.002)
Positive reciprocity		−0.020	−0.031		−0.015	−0.023
		(0.046)	(0.047)		(0.052)	(0.053)
Negative reciprocity		0.021	0.015		0.008	−0.000
		(0.051)	(0.051)		(0.057)	(0.057)
Fairness		−0.114***	−0.102***		−0.165***	−0.153***
		(0.036)	(0.036)		(0.041)	(0.041)
Sincerity		0.019	0.011		0.015	0.009
		(0.041)	(0.041)		(0.046)	(0.047)
Greed avoidance		−0.013	−0.010		−0.002	0.002
		(0.038)	(0.038)		(0.043)	(0.043)
# prev. studies		0.001*	0.001		0.001	0.001
		(0.001)	(0.001)		(0.001)	(0.001)
Female		−0.022	−0.020		−0.031	−0.040
		(0.095)	(0.095)		(0.108)	(0.108)
Undergrad or higher		0.134	0.154*		0.120	0.130
		(0.092)	(0.092)		(0.106)	(0.106)
Age		−0.005	−0.004		−0.004	−0.003
		(0.004)	(0.004)		(0.004)	(0.004)
Constant	−0.010	0.539**	0.120	−0.012	0.763**	0.362
	(0.078)	(0.274)	(0.306)	(0.087)	(0.307)	(0.345)
Observations	803	803	803	630	630	630
Pseudo R2	0.003	0.024	0.034	0.005	0.038	0.046
Log likelihood	−552.598	−540.675	−535.498	−432.255	−417.948	−414.586

Robust standard errors in parentheses.
* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

We elicited the same belief from the passive subjects in treatment *B*. Here, the average expected share of false high reports is not treatment-specific (56.19% in treatment *C*, 58.32% in *I*, and 56.75% in *N*; $p \geq 0.446$ for all pairwise comparisons). We did not implement comprehensive questions in treatment *B*. Given our insights from the *C*, *I*, and *N* treatments, the missing treatment differences might be caused by subjects who did not reflect seriously enough about the situation because they could not influence their payoff.

5.3. Robustness

5.3.1. Additional control variables

Next, we conduct a regression analysis to test whether (i) our main result is robust to adding control variables and (ii) personal characteristics contribute to the lying behavior. Table 4 depicts the results from probit regressions with the report as the dependent variable and treatment *N* as baseline. We use the main sample in the first three specifications and the restricted sample in specifications (4) to (6). In specifications (1) and (4), we replicate the central finding: As the coefficient of the dummy for treatment *C* is positive and significant, the desire-to-win effect leads to a larger share of high reports. Furthermore, the coefficient of the dummy for treatment *I* is positive as predicted by Hypothesis 2, but – as in the non-parametric analysis – not significant.

In specifications (2) and (5), we add controls for personal preferences, characteristics, and demographics. In the last step, we add a control for the subject's belief about the share of liars in their treatment in specifications (3) and (6). We observe that higher preferences for fairness correlate negatively and higher beliefs positively with high reports.[25] Other characteristics and demographics have no significant impact. In all specifications, the coefficient for treatment *C* remains economically and statistically significant, providing evidence for the desire-to-win effect.

5.3.2. Belief formation in the *N* treatment

Recall that the own *financial* benefit from lying is constant and depends neither on the treatment nor on the assigned case in treatments *I* and *N* nor the opponent's report in *C*. Thus, the payoff structure ensures that for subjects with outcome-based

[25] The belief needs to be interpreted with caution, as it might (at least partially) rationalize the own lying behavior.

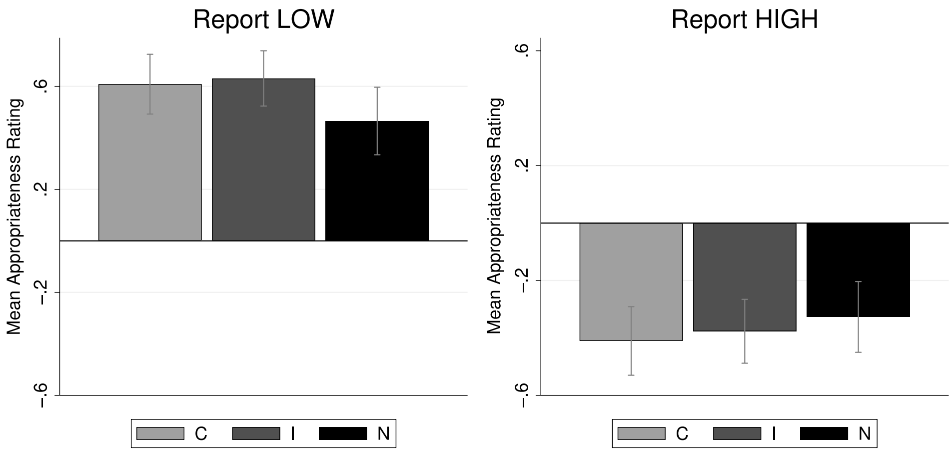


Fig. 2. The left (right) panel shows the mean appropriateness rating of report LOW (HIGH) given the actual result was LOW by treatments.

preferences, the belief about the opponent's report is irrelevant. Nevertheless, subjects might perceive the situation differently in the *C* and the *N* treatment: In treatment *C*, they need to form beliefs about the share of high reports, while this share is exogenously given in the *N* treatment. To address this concern, we designed the *N-belief* treatment.[26] This treatment is identical to the *N* treatment with the only difference being that the payment is not determined by matching the subject's report with a random draw but with the report of another person that previously participated in the *C* treatment. This change in the experimental design ensures that subjects in the *C and* the *N-belief* treatments have to form beliefs, and that the beliefs are formed in similar ways as they receive identical information.

We executed the *N-belief* treatment in 2024 on Prolific using the same requirements as in the previous treatments (see Section 4.1) and excluded participants from previous sessions of this study. We decided to collect new data for the *C* treatment as well to rule out that differences between the treatments are due to different times of data collection. We will refer to this data as the *C-new* treatment from hereon. We also needed to run the *B* treatment again to collect the data for the passive subjects (bystanders) of the *N-belief* treatment. Again, we aimed for 300 observations in each treatment and collected 300 observations for the *N-belief* and the *B* treatment and 298 for the *C-new* treatment. Two subjects failed the attention check in the *C-new* treatment and were dropped from the analysis. In line with our previous analysis, we again restrict our sample to subjects who answered all four control questions correctly after the second attempt in the *N-belief* and the *C* treatment. This leaves us with 249 observations for the *N-belief* treatment and 262 for the *C-new* treatment. If we further restrict our sample to subjects that answered all four control questions correctly in the first attempt, we have 249 subjects in the *N-belief* and 207 in the *C-new* treatment. The share of high reports is 61.83% in the *C-new* treatment and 52.61% in the *N-belief* treatment. Note first that we do not find significant differences between the share of high reports in the *C* treatment and the *C-new* treatment, indicating that the behavior is stable over time ($p = 0.301$). More importantly, we also do not observe significant differences when comparing the share of high reports between the *N* and the *N-belief* treatment ($p = 0.534$). In line with our previous findings, the difference in the share of high reports is economically and statistically significant when comparing the *N-belief* and the *C-new* treatment ($p = 0.040$, and $p = 0.008$ for the restricted sample). Given that the share of high reports in the *C* and the *C-new*, as well as the *N* and the *N-belief* treatment, did not differ significantly, we also ran an analysis with the pooled data from both *C* and both *N* treatments. The share of high reports is 59.46% in the *C-pooled* treatments and significantly different from the share of high reports 51.08% in the *N-pooled* treatments ($p = 0.007$ and $p = 0.001$ for the restricted sample). These results further strengthen the support for Hypotheses 1.

5.3.3. Social norm of lying

To assess whether our treatment comparisons in Section 5.1 can be traced back to outcome-based impacts of a (i) desire-to-win and (ii) negative externality, we elicited the social appropriateness of truthfully reporting the low outcome as well as misreporting it as high for the three treatments *C*, *I*, and *N* separately. One possible concern could be that the social appropriateness of lying is negatively affected by the negative payoff externality in *N* and *C*, thereby contaminating the comparisons to *I*: subjects could then decide to lie less in *N* (or *C*) than *I* not only because they care about the other subject's payoff, but also because they prefer to comply with social norms. Fig. 2 depicts the mean appropriateness rating of the two possible reports for each treatment. Recall that subjects had to choose one of six possible ratings between *very socially inappropriate* (coded as -1) and *very socially appropriate* (coded as 1).

[26] We thank the associate editor and an anonymous reviewer for suggesting this additional robustness check.

For all three main treatments, we observe a strong social norm of honesty. The difference between the two norm ratings of reporting a low outcome truthfully or not is a measure of how much less appropriate it is to lie instead of reporting truthfully. This difference is significantly different from zero for the *C*, *I*, and *N* treatments (two-sided t-test $p < 0.01$).[27] Importantly, the social norm of reporting honestly is not treatment-dependent, as we do not find a significant treatment effect ($p > 0.32$). There is even a somewhat weaker norm of behaving honestly in *N*, as the difference in norm ratings is lowest for this treatment (0.79 in *N* compared to 1.01 in *I* and 1.02 in *C*). Overall, we conclude that our treatment comparisons in Section 5.1 do not mask an underlying effect of social norms.

6. Discussion

In the following, we analyze extensions of our model and consider other behavioral motives that might potentially impact the degree of misconduct. First, we extend the model by allowing for (i) heterogeneity in the desire-to-win and (ii) conditional lying costs. We then discuss alternative explanations for the desire-to-win effect. We derive if and how these extensions and alternative motives impact our theoretical analysis and contrast them with our empirical findings.[28] We also discuss the generalizability and limitations of our study.

6.1. Extensions

First, we extend the model by allowing for heterogeneity in the individual desire-to-win parameters. In Section 3, we assume that $\hat{u}_i = \hat{u}$ for all players i. In reality, individuals may care differently about winning. With heterogeneous levels of $\hat{u}$, it no longer holds that the desire-to-win effect is positive for *every* player i. To see this, recall that the own desire-to-win boosts incentives to lie, while the opponent's desire-to-win mitigates the willingness to lie via social preferences. Consider a player i with a strictly positive, yet below average desire-to-win, so $\hat{u}_i < \mathbb{E}[\hat{u}]$ holds. If this player has sufficiently high other-regarding preferences, then the positive effect is offset by the negative effect. Accordingly, it is possible that this player lies in setting *N* but not in *C*: lying would increase their chances to gain a small extra utility, but at the same time lower the opponent's chances to obtain a much larger extra utility. However, this can only happen for players with a strictly below-average desire-to-win. For players with an average desire-to-win, the effect is positive as players care more about their own payoff.

As the lying cost threshold in *C* depends on the individual desire-to-win, the distribution of lying costs c becomes relevant. If the support of the distribution of lying costs is large enough to also include honest players and the probability mass of the distribution is evenly distributed, then the desire-to-win effect remains positive. We prove in the Appendix A that a sufficient though not necessary condition for a higher lying frequency in treatment *C* than in *N* is that lying costs c are uniformly distributed. For extreme distributions of c, the desire-to-win effect could become negative. As an example, assume that lying costs are so low that everyone lies in *N*, while the player with the lowest lying costs would not lie in *C* for a sufficiently large degree of other-regarding preferences. However, while theoretically interesting, such extreme distributions of lying costs seem unlikely.

Next, we extend the model by allowing for conditional lying costs. In our model, we have assumed that lying yields a constant cost of c. In treatment *C*, however, it may be that lying costs vary with the opponent's report: misreporting a low outcome as high might yield less guilt if one expects or learns that the opponent lies as well (Abeler et al., 2019). In the Appendix A, we, therefore, assume that a player's lying cost in *C* is higher if the opponent submits a low report. We show that the positive desire-to-win result is robust: only if a player in *C* expects that the opponent is more likely to submit a low than a high report and the difference between high and low lying costs is sufficiently large, the expected lying cost in *C* can be larger than in *N* so that the desire-to-win effect can become negative. Our results as well as those from the literature indicate that both conditions are typically not fulfilled. First, and foremost, our belief elicitation reveals that subjects in *C* on average expect that two out of three participants submit high reports. Second, the results of Dato et al. (2019) suggest that lying costs in a contest do not substantially vary with the opponent's report. Considering a sequential contest where players can lie about the binary outcome of a lottery, they find that the frequency of high reports of second movers does not depend on whether first movers report high or low.[29]

6.2. Alternative explanations

We discuss three additional motives that could influence the lying decisions differently across treatments: a desire-not-to-lose, loss aversion, and image concerns.

[27] For the *C*, *I*, and *N* treatments, it holds that the modal response is to rate a truthful low report as *very socially appropriate* and misreporting the outcome as high as *socially inappropriate*.

[28] In the main text, we restrict attention to intuitive explanations. A formal analysis can be found in Appendix A.2.

[29] A similar result is found in Feess et al. (2022) for the situation where the financial interests of the two players are aligned rather than opposed. Their data suggest that lying costs are deontological, and therefore independent of the behavior of the other player.

Desire-not-to-lose

Instead of the desire-to-win, individuals may have a desire-not-to-lose, driven e.g. by a status loss from losing. A first way of modeling this motive is to substitute the utility gain from winning with a utility loss from losing the contest. We find that such a model is the mirror image to a desire-to-win: As the incentive to lie depends solely on the utility spread between winning and losing, it does not matter whether this spread is enhanced by an extra utility from winning or a disutility from losing. Accordingly, the higher frequency of high reports in treatment C compared to N might as well be driven by a desire-not-to-lose. Notwithstanding that these motives may be quite different from a psychological point of view, our model suggests that their impact on lying is the same.

A second way of thinking about the desire-to-win and desire-not-to-lose, however, allows discriminating between the two motives. Suppose the extra utility from winning or the disutility from losing arises if and only if both players submit identical reports, that is when both players can still win or lose after having submitted their reports. In this case, a desire-to-win strengthens the incentive to submit the same report as the opponent, while the opposite holds for a desire-not-to-lose. Hence, the lying incentive depends on the belief about the opponent's report: The higher the probability that the other player reports HIGH, the higher is the incentive to report HIGH with a desire-to-win, and the lower is it with a desire-not-to-lose. The regression analysis in Table 4 reveals that the likelihood of submitting a high report increases with the belief about the share of liars, which contrasts the prediction from a desire-not-to-lose. Moreover, players in C on average expect more than two out of three opponents to report high. When the belief about the share of high reports is larger than 50%, a desire-to-win predicts more lying in C than N, while the opposite holds for a desire-not-to-lose. Our result of significantly more high reports in C than N thus clearly supports a desire-to-win.

Loss aversion

Several studies suggest that loss aversion increases lying (Shalvi, 2012; Grolleau et al., 2016; Schindler and Pfattheicher, 2017; Garbarino et al., 2019). With a fixed reference point, lying decreases the likelihood of incurring a loss, so loss aversion should also foster lying in our setup. Importantly, as loss aversion is a *general* concept for decision-making under risk and uncertainty, it concerns all of our three settings. Our formal analysis in the Appendix A provides three insights: First, the incentive to lie increases with the degree of loss aversion. Second, this increase is independent of the other player's report, as lying increases the probability of avoiding the loss always by 50 percentage points. Third, the lying incentive increase with the reference point. The same reasoning applies to settings I and N, as the probability of ending up in case 1, q, cancels out.[30] Hence, as long as the reference point is identical across settings, the impact of loss aversion on lying is identical in C, N, and I. Our hypotheses remain unchanged.

Garbarino et al. (2019) study the impact of loss aversion on lying when the reference point equals the expected material utility with truth-telling. Applied to our setup, the reference point would be a weighted sum of u_L and u_H, with the weights given by the probability q in I and N. In C, however, the weights are determined by the belief about the share of liars and are, hence, endogenous. The lower the belief about the share of liars, the higher the expected payoff with truth-telling, and thus the reference point. Due to loss aversion, the incentive to lie increases with the reference point (due to an increase in the potential perceived loss). As subjects in the experiment overestimated the actual degree of lying in C, the reference point would be lower than in I and N, which ceteris paribus reduces their incentive to lie. Thus, if loss aversion was a driver of lying behavior in our experiment, it would either not affect our treatment comparisons or predict less lying in C.

So far, we have assumed that the reference point is fixed at the very moment where the lying decision is taken. With expectation-based loss aversion à la Kőszegi and Rabin (2006, 2007), the reference point is stochastic and shaped by rational expectations about the own equilibrium choice. Note that, in equilibrium, the probability of a gain (when a player expects to win but ends up losing) is identical to the probability of a loss (when a player wins while having expected to lose). Due to loss aversion, however, the player always incurs a net loss. This net loss is weighted with the aforementioned probability, which increases with the uncertainty about the prize. It then holds that the attractiveness of lying increases with π, the subjective probability that the opponent reports LOW in C (and the corresponding probability q in N): the more one expects the opponent to report LOW, the more (less) uncertain is the outcome after reporting truthfully LOW as well (lying and reporting HIGH). As our data show that, on average, $\pi < q$ holds, expectation-based loss aversion does not qualify as a potential alternative to the desire-to-win effect.

Image concerns

The literature has shown that image concerns are an important driver of lying (Abeler et al., 2019). In line with their interpretation, we have captured (internal) self-image concerns via lying costs. In addition to self-image concerns, social image concerns (often also referred to as reputation concerns) towards the experimenter or other participants might reduce the willingness to lie. To see the consequences of social image concerns in our framework, it is instructive to distinguish between those towards the experimenter and those towards other subjects. Recall first that liars in our treatments can never be singled out, as the outcome of the lottery is neither observable to us nor to other subjects. As a consequence, image concerns towards the experimenter are identical across treatments. In treatments C and N, the report is observable to the other contestant (C) or the bystander (N), who are both affected by the subject's report. Hence, image concerns should be similar in C and N, and should not contaminate the identification of the desire-to-win effect (though we acknowledge that image concerns in C might be affected by the expectation about the other contestant's behavior). By contrast, in treatment I, there are only image concerns towards the experimenter, as there is no other subject. Since image concerns would hence lead to less lying in N than in I, they go in the same direction as the negative

[30] Note that this result also holds for a more general value function that exhibits diminishing sensitivity as proposed by Kahneman and Tversky (1979).

externality effect itself. The fact that we do not observe a significant treatment difference in the share of high reports between *I* and *N* suggests that image concerns do not play a strong role in our experiment. At least partially, this may be driven by the fact that the bystander in treatment *N* is informed only two days after the experiment about the active subject's report and that the perceived social distance is large in online experiments.

6.3. Limitations and generalizability

In this paper, we restrict attention to unethical behavior in the form of lying about the outcome of a lottery (a die roll), thereby following a widespread experimental approach. We opted for the die-rolling task, as (i) behavior in this paradigm has been shown to correlate with unethical behavior outside the lab (Cohn et al., 2015; Hanna and Wang, 2017; Cohn and Maréchal, 2018; Dai et al., 2018), and (ii) it can be easily applied to non-strategic and strategic decision settings.[31] Other prominently studied variants of unethical behavior are deception in sender-receiver games (Gneezy, 2005), taking money designated for donation (Ariely et al., 2009; Kirchler et al., 2016; Feess et al., 2023), and lying about the performance in real-effort tasks (Mazar et al., 2008; Ruedy and Schweitzer, 2010). We decided against sender-receiver games, as these are typically strategic by design: accordingly, it seemed unnatural to design several non-strategic treatments in the spirit of *N* and *I*. By contrast, using money designated for donation as unethical behavior fits nicely for treatment *I*, but the emerging two-sided negative externality (on the donation agency and the bystander/opponent) in treatments *N* and *C* would impose undesirable cognitive challenges on the subjects. It is worthwhile noting that deception, taking money designated for donation, and lying are different types of unethical behavior.[32] Accordingly, it is not directly obvious that our results generalize to all types of unethical behavior, and we leave it to future research to study the potential differences in behavior.

In our model and our estimation of the lying behavior in the experiment, we have assumed that subjects report a high outcome of the lottery truthfully. While misreporting a high outcome as low reduces the expected payoff and yields lying costs, there might, in principle, nevertheless be psychological motives that could lead to downward lying. We see three potential channels for that but argue that the necessary conditions for these channels are not likely to be satisfied for a large number of subjects in our experiment (for all three channels, see the formal analysis in the Appendix A).

First, with heterogeneous desire-to-win parameters, a player might be inclined to engage in downward lying if the opponent's (individual or expected) desire-to-win is so far above the own that other-regarding preferences dominate the lower payoff, the own desire-to-win, and lying costs. These conditions, however, seem to be very restrictive.

Second, if the extra utility from winning accrues only with identical reports, then downward lying increases the likelihood of getting the extra utility from winning if one expects that the opponent is more likely to submit a low report. However, as subjects in treatment *C* expect that the opponent reports high far more often than low, this cannot rationalize downward lying.[33]

Third, downward lying might potentially be rationalized by expectation-based loss aversion. If one expects that the opponent reports HIGH, then truthfully reporting a high outcome yields a 50/50 gamble between winning and losing. By contrast, downward lying yields the loser prize with certainty and hence avoids the gamble. If a player is sufficiently loss-averse, this could, in principle, trigger downward lying. This, however, requires that the degree of loss aversion is so strong that the player (if she has not overly strong social preferences) would be willing to choose first-order stochastically dominated options. Since the requirements for downward lying are very restrictive for all three channels, we do not believe that downward lying is a severe issue in our experiment.

As an alternative to our lottery setting, a promising possibility would have been a real-effort task, where subjects can increase the probability of receiving a higher payoff by misreporting their performance. We believe, however, that using a real-effort task comes at a cost: first, lying behavior is likely to depend on the kind of task, as subjects who are less capable in the task chosen may feel more entitled to inflate their actual performance. Accordingly, the results would be sensitive to the type of task, implying a low degree of generalizability. Second, using a lottery is conservative, as image concerns are likely to be more important for (challenging) real-effort tasks, thereby boosting the desire-to-win effect. Thus, if the difference in lying between treatments *C* and *N* is significant in the lottery setting, it should be even more pronounced in real-effort settings.

A real-effort setting might be desirable from an applied perspective because it increases the external validity. For instance, it would be interesting to understand whether the potentially larger desire-to-win effect leads to a significant difference in lying between the competitive incentive scheme *C* and the individual incentive scheme *I*. Furthermore, one could combine the lottery- and the real-effort-setting by experimentally implementing a production function as in Lazear and Rosen (1981): a subject works on a real-effort task, and the impact of noise on individual output is captured by a lottery draw that is private information and needs to be reported. We believe that this is an interesting avenue for future research.

Our experimental approach allows us to disentangle different channels through which (here, lying) behavior might be affected, which is the main purpose of our contribution. We acknowledge that this implies abstracting from a variety of possibly important contextual and personal influence factors. For instance, the degree of lying and, as a consequence, also the negative externality-

[31] In mind games (Jiang, 2013; Potters and Stoop, 2016; Kajackaite and Gneezy, 2017), subjects are asked to think about a number and are paid if and only if the observable outcome of a die roll matches their number. This is closely related to the report about the (unobservable) outcome of a lottery. We hence suspect that the results of mind games should be similar to our lottery.

[32] For conceptual definitions of different types of unethical behavior in strategic settings, see Sobel (2020).

[33] Theoretically, one might rationalize downward lying when subjects suffer from losing only with identical reports. Our analysis in Section 6.2, however, suggests that the subjects' lying behavior is not driven by this interpretation of the desire-not-to lose motive.

and the desire-to-win-effect might depend on the context: the negative externality might influence the behavior more strongly in a promotion tournament among colleagues than in a professional sports contest, where the competitive zero-sum nature and the corresponding incentives for unethical behavior might have been internalized and somewhat accepted by all contestants.[34] Furthermore, the literature has shown that the willingness to behave unethically is shaped by individual attributes (Kish-Gephart et al., 2010), (competitive) pressure (Welsh and Ordóñez, 2014; Mitchell et al., 2023), and the perception of being treated fairly (Schweitzer et al., 2004; Cadsby et al., 2010). Accordingly, our results should not be interpreted as evidence for a universal desire-to-win effect, as its existence and magnitude are likely to be shaped by the aforementioned factors.

7. Conclusion

In the last decades, incentive schemes based on the relative performance of employees have been criticized, and many companies abolished or at least mitigated them. Arguably, while competitive pressure may set high incentives to perform well, it may also incentivize employees to game the system, cheat, and even commit outright fraud. Laboratory experiments on cheating and lying in contests also support the view that competition leads to a higher degree of misconduct, but the channels through which competition affects cheating and lying have so far not been explored systematically. First, as the expected marginal financial benefit from misconduct tends to be higher in competitive payment schemes, it cannot be excluded that differences in the observed behavior are not driven by competition-specific characteristics, but by differences in financial incentives. We account for this issue by designing our non-competitive payment scheme (treatment I) such that the expected financial benefit from lying about the outcome of a lottery is the same as in our competitive payment scheme (treatment C).

We identify and disentangle two channels through which lying might be affected differently in competition than under individual incentive schemes: the lying-reducing negative externality effect and the lying-increasing desire-to-win effect. First, our experimental results reveal that there is a significant positive desire-to-win effect. Subjects in a competition (treatment C) submit more high reports than subjects in a situation with (i) identical financial incentives and (ii) the same negative externality (treatment N). To prove the robustness of this result, we design an additional individual incentive scheme (treatment *N-belief*) to account for the possibility that differences in the belief formation process contaminate our identification of the desire-to-win effect. After collecting new data for this and the competitive incentive scheme, the desire-to-win effect remains positive and significant. This is our main contribution: By keeping all financial incentives constant with and without competition, we show that the desire-to-win effect leads to a significantly higher frequency of lying. Second, our results document a negative, yet insignificant negative externality effect, which arises from a comparison of behavior in N to an individual incentive scheme without negative externality (treatment I). Apparently, the subjects did not care strongly about the opponent's payoff when making the lying decision.

Furthermore, we elicit the social inappropriateness of lying and, by showing that it does not differ across incentive schemes, demonstrate that our results do not mask an underlying social norms effect. Moreover, we discuss and analyze several possible extensions of the model and show that our hypotheses are robust to heterogeneity in the desire-to-win and conditional lying costs. Furthermore, we demonstrate that loss aversion, a desire-not-to lose, and image concerns do not qualify as alternative explanations for the desire-to-win effect. Finally, we discuss the limitations of our contribution, put our results into perspective, and provide possible avenues for future research: although clearly beyond the scope of this paper, it would be interesting to see if the desire-to-win effect is specific to lying or carries over to other types of unethical behavior. Likewise, from an applied perspective, it would be interesting to know whether the desire-to-win effect survives or even aggravates when introducing a real-effort task or varying other contextual factors.

Appendix A

A.1. Descriptive statistics

See Table A.1.

Table A.1
Descriptive statistics for all treatments. Results for the main treatments are based on the main sample, the results for the restricted sample are reported in parentheses. We report average results for all variables except for the number of subjects.

Treatment	# subjects	Female	Age	Undergrad or higher	# prev. Studies	HIGH report
C	298(242)	0.51(0.50)	34.66(34.42)	0.60(0.62)	57.32(58.70)	0.574(0.591)
I	247(182)	0.45(0.40)	34.25(32.92)	0.49(0.54)	66.07(65.07)	0.539(0.533)
N	258(206)	0.48(0.46)	34.29(34.17)	0.54(0.56)	63.40(61.17)	0.496(0.495)
C-new	262(207)	0.49(0.49)	41.97(42.28)	0.61(0.65)	138.41(139.78)	0.618(0.652)
N-belief	249(249)	0.49(0.49)	41.63(41.63)	0.61(0.61)	132.31(132.31)	0.526(0.526)
B	303	0.50	33.45	0.60	63.39	–
Norm	303	0.50	33.68	0.61	61.57	–

[34] Along these lines, Lance Armstrong argued in court that, given the behavior of others in professional cycling, his sophisticated doping scheme mainly restored fair competition, allowing the most capable athlete to win.

A.2. Proofs

Heterogeneous desire-to-win

We assume that players may differ in their desire-to-win preferences. Suppose that the $\hat{u}_i$'s are realizations of independent and identically distributed random variables, drawn from a distribution $f(\hat{u}_i)$. Furthermore, assume that lying costs are distributed uniformly on an interval $[0, \bar{c}]$ with $\bar{c}$ sufficiently high to ensure that some types refrain from lying in any case. First, suppose that the individual desire-to-win is private information, while the distribution, and thus also the expected desire-to-win $\mathbb{E}[\hat{u}]$, is common knowledge. With a truthful report $r_i = l$, player i's utility is

$$\begin{aligned} U_i^C(l) &= \pi \left\{ \frac{1}{2}\left[u_L + \phi\left(u_H + \mathbb{E}[\hat{u}]\right)\right] + \frac{1}{2}\left(u_H + \hat{u}_i + \phi u_L\right) \right\} + (1-\pi)\left[u_L + \phi\left(u_H + \mathbb{E}[\hat{u}]\right)\right] \\ &= u_L + \frac{\pi}{2}\left(\Delta u + \hat{u}_i\right) + \phi\left[u_L + \left(1 - \frac{\pi}{2}\right)\left(\Delta u + \mathbb{E}[\hat{u}]\right)\right] . \end{aligned}$$

When reporting $r_i = h$, her utility is

$$\begin{aligned} U_i^C(h) &= \pi\left(u_H + \hat{u}_i + \phi u_L\right) + (1-\pi)\left\{ \frac{1}{2}\left[u_L + \phi\left(u_H + \mathbb{E}[\hat{u}]\right)\right] + \frac{1}{2}\left(u_H + \hat{u}_i + \phi u_L\right) \right\} \\ &= u_L + \frac{1}{2}(1+\pi)\left(\Delta u + \hat{u}_i\right) + \phi\left[u_L + \frac{1-\pi}{2}\left(\Delta u + \mathbb{E}[\hat{u}]\right)\right] - c. \end{aligned}$$

Comparing the expected utilities shows that player i lies if and only if

$$c < (1-\phi)\frac{\Delta u}{2} + \frac{1}{2}\left(\hat{u}_i - \phi\mathbb{E}[\hat{u}]\right) \equiv \tilde{c}_C^i.$$

Comparing the threshold to the threshold in treatment N, $\tilde{c}_N = (1-\phi)\frac{\Delta u}{2}$, yields

$$\tilde{c}_C^i < \tilde{c}_N \quad \Leftrightarrow \quad \hat{u}_i < \phi\mathbb{E}[\hat{u}] .$$

Players with a sufficiently low (high) desire-to-win would be less (more) inclined to lie in C than in N. To assess whether the expected share of lies is higher or lower in C than in N, it suffices to compare the expected lying cost threshold $\mathbb{E}\left[\tilde{c}_C^i\right]$ in C to the corresponding threshold $\tilde{c}_N$ in N:

$$\mathbb{E}\left[\tilde{c}_C^i\right] = \int_{\hat{u}_i}\left[(1-\phi)\frac{\Delta u}{2} + \frac{1}{2}\left(\hat{u}_i - \phi\mathbb{E}[\hat{u}]\right)\right] f\left(\hat{u}_i\right) d\hat{u}_i = \tilde{c}_N + (1-\phi)\frac{\mathbb{E}[\hat{u}]}{2} > \tilde{c}_N.$$

On average, the threshold is higher in C than in N so that lying is on average more attractive in C than in N. Hence, the desire-to-win effect remains positive.

Second, suppose that the individual $\hat{u}_i$'s are common knowledge. Then,

$$\tilde{c}_C^{i,j} = \tilde{c}_N + \frac{1}{2}\left(\hat{u}_i - \phi\hat{u}_j\right)$$

is the threshold of player i with $\hat{u}_i$ when competing against player j with $\hat{u}_j$. Averaging over the possible opponent's types, the expected threshold for player i is

$$\mathbb{E}\left[\tilde{c}_C^i|\hat{u}_i\right] = \int_{\hat{u}_j}\left[\tilde{c}_N + \frac{1}{2}\left(\hat{u}_i - \phi\hat{u}_j\right)\right] f\left(\hat{u}_j\right) d\hat{u}_j = \tilde{c}_N + \frac{1}{2}\left(\hat{u}_i - \phi\mathbb{E}[\hat{u}]\right),$$

and, hence, equal to the previous case. Accordingly, we still get

$$\mathbb{E}\left[\tilde{c}_C^i\right] = \tilde{c}_N + (1-\phi)\frac{\mathbb{E}[\hat{u}]}{2} > \tilde{c}_N.$$

Hence, we have shown that the effect remains positive with a heterogeneous desire-to-win, no matter whether the opponent's desire-to-win is observable or not. Note that the uniform distribution of lying costs that includes honest types is a sufficient but not a necessary condition for a positive desire-to-win-effect. Still, there exist specific distributions for which the desire-to-win effect is negative. Consider a distribution of lying costs with all probability mass on the interval below $\tilde{c}_N$ but above the threshold in C of the type with the lowest desire-to-win: then all players would lie in scenario N, whereas the player with the lowest desire-to-win would not lie in C.

Conditional lying costs

Consider treatment C and assume that player i incurs lying cost c_H (c_L) if $r_j = l$ ($r_j = h$) with $c_H = c + \omega$ and $c_L = c - \omega$, where $\omega > 0$. Misreporting the low outcome as high, $r_i = h$, then yields

$$U_i^C(h) = \pi\left(u_H + \hat{u} + \phi u_L - c_H\right) + (1-\pi)\left\{ \frac{1}{2}\left[u_L + \phi\left(u_H + \hat{u}\right)\right] + \frac{1}{2}\left(u_H + \hat{u} + \phi u_L\right) - c_L \right\}.$$

With a truthful report $r_i = l$, player i's utility is

$$U_i^C(l) = u_L + \frac{\pi}{2}(\Delta u + \hat{u}) + \phi\left[u_L + \left(1 - \frac{\pi}{2}\right)(\Delta u + \hat{u})\right].$$

Comparing the two expected utilities shows that player i lies if and only if

$$c < (1-\phi)\frac{\Delta u}{2} + (1-\phi)\frac{\hat{u}}{2} - (2\pi - 1)\omega \equiv \tilde{c}_C^c.$$

Recall that $\tilde{c}_N = (1-\phi)\frac{\Delta u}{2}$. It then holds that the willingness to lie is higher in C than in N if and only if

$$(1-\phi)\frac{\hat{u}}{2} > (2\pi - 1)\,\omega$$

For $\pi \leq \frac{1}{2}$, the desire-to-win effect is always positive: if player i expects that the opponent is more likely to report H than L, the expected lying cost in C is lower than the certain lying cost c in scenario N. Only if the player expects that other players predominantly submit low reports, it is possible that the desire-to-win effect is not positive. But even with $\pi = 1$, it holds that the desire-to-win effect remains positive as long as the difference between the lying costs in C is not too large, i.e., if $\omega < \frac{(1-\phi)\hat{u}}{2}$.

Desire-not-to-lose

Instead of a desire-to-win effect, suppose the players' utility can be reduced by $\hat{u}$ in case they lose. We consider two versions: First, we assume that the disutility accrues whenever a player gets the loser prize. Second, we assume that the disutility accrues only when both players have submitted identical reports.

In the first version, player i's utility with a truthful low report is

$$\begin{aligned} U_i^C(l) &= \pi\left\{\frac{1}{2}\left(u_L - \hat{u} + \phi u_H\right) + \frac{1}{2}\left[u_H + \phi\left(u_L - \hat{u}\right)\right]\right\} + (1-\pi)\left(u_L - \hat{u} + \phi u_H\right) \\ &= u_L - \hat{u} + \frac{\pi}{2}(\Delta u + \hat{u}) + \phi\left[u_L - \hat{u} + \left(1 - \frac{\pi}{2}\right)(\Delta u + \hat{u})\right]\ , \end{aligned}$$

whereas misreporting the low outcome as high yields

$$\begin{aligned} U_i^C(h) &= \pi\left[u_H + \phi\left(u_L - \hat{u}\right)\right] + (1-\pi)\left\{\frac{1}{2}\left(u_L - \hat{u} + \phi u_H\right) + \frac{1}{2}\left[u_H + \phi\left(u_L - \hat{u}\right)\right]\right\} - c \\ &= u_L - \hat{u} + \frac{1}{2}(1+\pi)(\Delta u + \hat{u}) + \phi\left[u_L - \hat{u} + \frac{1-\pi}{2}(\Delta u + \hat{u})\right] - c. \end{aligned}$$

Comparing the expected utilities shows that player i lies if and only if

$$c < (1-\phi)\frac{\Delta u + \hat{u}}{2} = \tilde{c}_C\ ,$$

which shows that the lying cost threshold with a fear-of-losing is identical to the one with a desire-to-win in C.

Now consider the second version, in which a desire-not-to-lose can be effective only if both players have submitted identical reports. Suppose the outcome of i's lottery was low. Then truthfully submitting $r_i = l$ yields

$$\begin{aligned} U_i^C(l) &= \pi\left\{\frac{1}{2}\left(u_L - \hat{u} + \phi u_H\right) + \frac{1}{2}\left[u_H + \phi\left(u_L - \hat{u}\right)\right]\right\} + (1-\pi)\left(u_L + \phi u_H\right) \\ &= u_L + \frac{\pi}{2}(\Delta u - \hat{u}) + \phi\left[u_L - \frac{\pi}{2}\hat{u} + \left(1 - \frac{\pi}{2}\right)\Delta u\right]\ , \end{aligned}$$

whereas misreporting the low outcome as high and submitting $r_i = h$ yields

$$\begin{aligned} U_i^C(h) &= \pi\left(u_H + \phi u_L\right) + (1-\pi)\left\{\frac{1}{2}\left(u_L - \hat{u} + \phi u_H\right) + \frac{1}{2}\left[u_H + \phi\left(u_L - \hat{u}\right)\right]\right\} - c \\ &= u_L - \frac{1-\pi}{2}\hat{u} + \frac{1+\pi}{2}\Delta u + \phi\left[u_L + \frac{1-\pi}{2}(\Delta u - \hat{u})\right] - c. \end{aligned}$$

Comparing the expected utilities shows that player i lies if and only if

$$c < (1-\phi)\frac{\Delta u}{2} + (1+\phi)\frac{2\pi - 1}{2}\hat{u} \equiv \tilde{c}_C^l$$

The threshold $\tilde{c}_C^l$ increases with π: the higher the expected probability that the opponent reports low, the less likely is it to suffer from the desire-not-to-lose after reporting high. Note that the threshold is lower than $\tilde{c}_N$ if and only if $\pi < \frac{1}{2}$.

Under an analogous interpretation of the desire-to-win, which accrues only when the players have submitted identical reports, it would be optimal to lie if and only if

$$c < (1-\phi)\frac{\Delta u}{2} - (1+\phi)\frac{2\pi - 1}{2}\hat{u} \equiv \tilde{c}_C^w$$

Contrary to $\tilde{c}_C^l$, the threshold $\tilde{c}_C^w$ decreases with π: the higher the expected probability that the opponent reports low, the less likely it is to benefit from the desire-to-win after reporting high. Note that the threshold is lower than $\tilde{c}_N$ if and only if $\pi > \frac{1}{2}$.

Loss aversion

We consider two versions, which differ with respect to reference point formation. First, we consider a fixed reference point as in Kahneman and Tversky (1979). Second, we analyze a variant with an expectation-based reference point à la (Kőszegi and Rabin, 2006, 2007).

Fixed Reference Point:

Suppose that players are loss averse and compare the actual material utility u_k, $k \in \{L, H\}$, to a reference point $r \in (u_L, u_H)$, so winning (losing) yields a gain (loss). Let $\Delta = u_k - r$ denote the difference between the actual outcome and the reference point. The comparison is evaluated according to a piece-wise linear value function

$$\mu(\Delta) = \begin{cases} \Delta & \text{if } \Delta \geq 0 \\ \lambda\Delta & \text{if } \Delta < 0 \end{cases},$$

where $\lambda > 1$ expresses loss aversion. The overall utilities in C then read:

$$U_i^C(l) = \pi\left\{\frac{1}{2}\left[u_L + \phi u_H - \lambda\left(r - u_L\right)\right] + \frac{1}{2}\left(u_H + \phi u_L + u_H - r\right)\right\} + (1-\pi)\left[u_L + \phi u_H - \lambda\left(r - u_L\right)\right] \ ,$$

so that

$$U_i^C(l) = u_L + \frac{\pi}{2}\Delta u + \phi\left[u_L + \left(1 - \frac{\pi}{2}\right)\Delta u\right] + \frac{\pi}{2}\left(u_H - r\right) - \left(1 - \frac{\pi}{2}\right)\lambda\left(r - u_L\right),$$

whereas misreporting the low outcome as high yields

$$U_i^C(h) = \pi\left(u_H + \phi u_L + u_H - r\right) + (1-\pi)\left\{\frac{1}{2}\left[u_L + \phi u_H - \lambda\left(r - u_L\right)\right] + \frac{1}{2}\left(u_H + \phi u_L + u_H - r\right)\right\} - c$$

so that

$$U_i^C(h) = u_L + \frac{1+\pi}{2}\Delta u + \phi\left(u_L + \frac{1-\pi}{2}\Delta u\right) + \frac{1+\pi}{2}\left(u_H - r\right) - \frac{1-\pi}{2}\lambda\left(r - u_L\right) - c.$$

Comparing the expected utilities shows that player i lies if and only if

$$c < (1-\phi)\frac{\Delta u}{2} + \frac{u_H - r + \lambda\left(r - u_L\right)}{2} \equiv \tilde{c}_C^{LA}.$$

The threshold $\tilde{c}_C^{LA}$ is increasing in λ and r. The more loss averse the player is, the more attractive lying is. Likewise, a higher reference point renders lying more attractive.

For setting N, the utilities $U_i^N(h)$ and $U_i^N(l)$ are identical to their counterparts in C, except for the fact that π needs to be replaced by q. Just as π in C, q cancels out in the comparison so that

$$\tilde{c}_N^{LA} = \tilde{c}_C^{LA}.$$

Accordingly, with a fixed reference point, loss aversion predicts no difference in lying between C and N.

Stochastic Expectation-based Reference Point:

Now suppose that the reference point is stochastic and shaped by rational expectations. Consider a player i who expects to submit $r_i = l$ and anticipates that the opponent will submit $r_j = l$ with probability π. The reference point is shaped by the resulting expected lottery over material outcomes: the player expects to receive u_H (u_L) with probability $\frac{\pi}{2}$ ($\frac{2-\pi}{2}$). If player i expects to submit report $r_i = h$, the reference point is given by the lottery that yields u_H (u_L) with probability $\frac{1+\pi}{2}$ ($\frac{1-\pi}{2}$). The psychological gain–loss utility is then determined by a comparison of possibly realized and expected outcomes, where every comparison is weighted with its expected occurrence probability. Loss aversion is again incorporated by the fact that a loss is multiplied with $\lambda > 1$. The player's overall expected utility is given by the sum of expected material and psychological utility.

We consider two equilibrium concepts, which differ in the assumption of whether the reference point adapts to the actual choice or not. First, with choice-unacclimating expectations (UPE), actual and expected choice may differ from each other. Suppose a player has formed the expectation to report low, but changes their mind and submits a high report: if the uncertainty about the final outcome is resolved quickly after the report has been submitted, it seems reasonable that the expected outcome is still shaped by the expectation to report low when evaluating the outcome. The equilibrium concept *Personal Equilibrium (PE)* then requires that expected and actual choices are internally consistent: reporting low (high) is only a PE when there is no incentive to deviate from the expectation to report low (high). Second, with *choice-acclimating expectations*, the expected and the actual report are identical by assumption. This applies to situations where the uncertainty about the realized material outcome is resolved long after the report has been submitted. During this time, the expectation must have been adapted to the actual choice, so that the reference point needs to be shaped by the actual choice when evaluating outcomes. The report which, when expecting to submit and when actually submitting it, yields the highest expected utility, is the *Choice-acclimating Personal Equilibrium (CPE)*.

First, suppose expectations are choice-unacclimating and denote the expected utility from reporting r_i while having expected to report r_i^e by $U(r_i|r_i^e)$. Consider the situation in which the actual outcome of the lottery for i was LOW. With $r_i = h$ and $r_i^e = l$, player i expects to obtain u_H (u_L) with probability $\frac{\pi}{2}$ ($\frac{2-\pi}{2}$), but receives u_H (u_L) with probability $\frac{1+\pi}{2}$ ($\frac{1-\pi}{2}$). Hence, with the probability $\frac{\pi}{2} \cdot \frac{1-\pi}{2}$, the player incurs a loss of size Δu, because of expecting to win but ending up losing. Likewise, i expects to obtain u_L with probability $\frac{2-\pi}{2}$ but ends up receiving u_H with probability $\frac{1+\pi}{2}$, so that the player incurs a gain of size Δu with probability $\frac{2-\pi}{2} \cdot \frac{1+\pi}{2}$. Accordingly, the utility $U_i^C(h|l)$ is given by

$$U_i^C(h|l) = u_L + \frac{1+\pi}{2}\Delta u + \phi\left[u_L + \frac{1-\pi}{2}\Delta u\right] + \frac{2-\pi}{2}\frac{1+\pi}{2}\Delta u - \frac{\pi}{2}\frac{1-\pi}{2}\lambda\Delta u - c,$$

while

$$U_i^C(l|l) = u_L + \frac{\pi}{2}\Delta u + \phi\left[u_L + \left(1 - \frac{\pi}{2}\right)\Delta u\right] - (\lambda - 1)\frac{2-\pi}{2}\frac{\pi}{2}\Delta u.$$

Reporting $r_i = l$ truthfully is a PE if and only if $U_i^C(l|l) \geq U_i^C(h|l)$

$$U_i^C(l|l) \geq U_i^C(h|l) \;\Leftrightarrow\; c \geq (1-\phi)\frac{\Delta u}{2} + \frac{2 + (\lambda-1)\pi}{4}\Delta u \equiv \tilde{c}_C^{PE,l}$$

In the same way, lying and submitting $r_i = h$ is a PE if and only if $U_i^C(h|h) \geq U_i^C(l|h)$. With

$$U_i^C(h|h) = u_L + \frac{1+\pi}{2}\Delta u + \phi\left(u_L + \frac{1-\pi}{2}\Delta u\right) - (\lambda - 1)\frac{1-\pi}{2}\frac{1+\pi}{2}\Delta u - c$$

and

$$U_i^C(l|h) = u_L + \frac{\pi}{2}\Delta u + \phi\left[u_L + \left(1 - \frac{\pi}{2}\right)\Delta u\right] + \left(\frac{1-\pi}{2}\frac{\pi}{2} - \frac{1+\pi}{2}\frac{2-\pi}{2}\right)\Delta u,$$

it holds that lying is a PE if and only if

$$U_i^C(h|h) \geq U_i^C(l|h) \;\Leftrightarrow\; c \leq (1-\phi)\frac{\Delta u}{2} + \frac{1 + \lambda + (\lambda-1)\pi}{4}\Delta u \equiv \tilde{c}_C^{PE,h}.$$

Note that $\tilde{c}_C^{PE,h} > \tilde{c}_C^{PE,l}$, so that lying (truth-telling) is the unique PE if and only if $c < \tilde{c}_C^{PE,l}$ ($c > \tilde{c}_C^{PE,h}$), whereas both are a PE if $c \in \left(\tilde{c}_C^{PE,l}, \tilde{c}_C^{PE,h}\right)$.

Second, suppose expectations are choice-acclimating. If the actual outcome of the lottery for i was LOW,

$$U_i^C(h|h) \geq U_i^C(l|l) \;\Leftrightarrow\; c \leq (1-\phi)\frac{\Delta u}{2} + (\lambda-1)\frac{2\pi - 1}{4}\Delta u \equiv \tilde{c}_C^{CPE}.$$

For setting N, the utilities are $U_i^N(h|h)$, $U_i^N(l|h)$, $U_i^N(h|l)$, $U_i^N(l|l)$, and the equilibrium thresholds are again identical to their counterparts in C, except for the fact that π needs to be replaced by q. It holds that all thresholds increase with π (or q, respectively). Accordingly, loss aversion with a stochastic reference point (shaped by rational expectations) predicts more lying in C than in N if and only if $\pi \geq q$.

Downward lying

The analysis of downward lying is identical to the above analysis of upward lying, except that lying costs arise with a low report. Therefore, the threshold for downward lying is the additive inverse of the upward lying threshold. Hence, downward lying can be optimal if and only if the lying cost threshold for upward lying is negative.

With heterogeneous and observable desire-to-win parameters, we get

$$\begin{aligned}\tilde{c}_C^{i,j} < 0 \quad &\Leftrightarrow \quad (1-\phi)\frac{\Delta u}{2} + \frac{1}{2}\left(\hat{u}_i - \phi\hat{u}_j\right) < 0\\ &\Leftrightarrow \quad \hat{u}_j > \frac{\hat{u}_i}{\phi} + \frac{1-\phi}{\phi}\Delta u\end{aligned}$$

With the alternative desire-to-win, it holds that $\tilde{c}_C^{w} \geq 0$ whenever $\pi \leq \frac{1}{2}$. If $\pi > \frac{1}{2}$,

$$\begin{aligned}\tilde{c}_C^{w} < 0 \quad &\Leftrightarrow \quad (1-\phi)\frac{\Delta u}{2} - (1+\phi)\frac{2\pi - 1}{2}\hat{u} < 0\\ &\Leftrightarrow \quad \phi > \frac{\Delta u - (2\pi-1)\hat{u}}{\Delta u + (2\pi - 1)\hat{u}}\end{aligned}$$

With expectation-based loss aversion and choice-acclimating expectations, it holds that $\tilde{c}_C^{CPE} \geq 0$ whenever $\pi \geq \frac{1}{2}$. If $\pi < \frac{1}{2}$,

$$\begin{aligned}\tilde{c}_C^{CPE} < 0 \quad &\Leftrightarrow \quad (1-\phi)\frac{\Delta u}{2} + (\lambda - 1)\frac{2\pi - 1}{4}\Delta u < 0\\ &\Leftrightarrow \quad \lambda > 1 + \frac{2(1-\phi)}{1-2\pi} \equiv \tilde{\lambda}.\end{aligned}$$

Note that if $\phi < \frac{1}{2}$, then $\tilde{\lambda} \geq 2\ \forall \pi \in \left[0, \frac{1}{2}\right)$.

A.3. Instructions

Consent Form

Welcome to our study!

First, we give some general information about our team, the aim of the study, and data protection.

Aim and data collection:

We are interested in individual decision making and personal characteristics. *[I, C, N, N-belief, C-new, and bystanders in N treatment: We ask you to answer a survey about your attitudes towards others and give some predictions about the behavior of individuals.* [*only for [I, C, N, N-belief, C-new treatments: In addition, you will be asked to roll a die and report the outcome.] [only for Norm treatment: We ask you to answer a survey about your attitudes towards others and evaluate the choices of other participants.]* We will ask you to give us your Prolific ID to ensure that we can pay you. In our study, we will also use the demographic information such as age or education you provided on Prolific.

Important: All information we provide in this study is true. You will never get inaccurate information.

Risks and benefits:

There are no physical or emotional risks associated with this study that would go beyond the risks of daily life. Your participation in this study will help us to better understand individual decision making.

Payment:

You will receive a fixed payment of 1.40 GBP for taking part in our study. In addition, you can earn a bonus. The payment will be sent to you within two days after completion of this study.

Confidentiality:

The information collected in this study may be published in a report or a journal article and presented to interested parties, including possibly, but not exclusively, members of editorial boards and scientific committees. In no circumstances will your Prolific ID be disclosed to people outside the research group. No personal data (e.g. your IP address) will be collected. Other information (e.g., survey responses, time of the study) will be kept by the researchers and may be used for further studies.

Your rights as a participant:

Participation is entirely voluntary. You may leave the survey at any time without any penalty or prejudice.

Do you wish to participate?

- Yes, continue
- No, leave survey

Please enter you Prolific ID

——————-

[Instructions for Treatment C]

Before the survey starts, you will play a simple game where you can earn an additional bonus that will be added to the 1.40 GBP you receive for the survey.

You will be matched with another participant who is also taking part in this study. You will not learn this participant's ID nor will they learn yours.

In this game, you will have to roll a six-sided die. You are free to choose how to obtain the outcome of a die roll, by using either a physical or a virtual die. Hence, you will roll the die in private, so that the outcome cannot be seen by anyone else. After you will have privately observed the outcome of the die roll, we will ask you to report the result of your die roll.

An outcome of

means the result is "Low".

An outcome of

means the result is "High".

The participant you are matched with will also roll a die in private and report either "High" or "Low". You will not be informed about their report before you enter your report. Also, the other participant will not be informed about your report before entering their report.

The table below shows how your and the other participant's bonus payment depend on the reports of both of you. In all cases, one of you will receive a bonus of 1.20 GBP and the other one a bonus of 0.20 GBP.

Report of the other participant	Your report	Bonus
Low	Low	Each of you has a 50% chance of getting the 1.20 GBP. This is decided by a random draw.
	High	You: 1.20 GBP Other: 0.20 GBP
High	Low	You: 0.20 GBP Other: 1.20 GBP
	High	Each of you has a 50% chance of getting the 1.20 GBP. This is decided by a random draw.

The bonus will be sent to you within two days after completion of this study. You will also receive a message via the Prolific system informing you about the reports of both participants and the resulting payment. The other participant will also receive such a message.

[Control questions]

Before we start the game, we want to make sure you have understood the set-up.
Please answer the questions below to the best of your knowledge.

[Table from above was shown here.]
Imagine the following situation:

[Each of the following questions was shown on a separate page. Participants could try twice and were informed about the correct answer afterwards.]
Suppose you have reported “Low”. If the other one reports “Low”, how likely is it that you get the bonus of 1.20 GBP?
Suppose you have reported “Low”. If the other one reports “High”, how likely is it that you get the bonus of 1.20 GBP?
Suppose you have reported “High”. If the other one reports “Low”, how likely is it that you get the bonus of 1.20 GBP?
Suppose you have reported “High”. If the other one reports “High”, how likely is it that you get the bonus of 1.20 GBP?
[Instructions for Treatment N]

Before the survey starts, you will play a simple game where you can earn an additional bonus that will be added to the 1.40 GBP you receive for the survey.

You will be matched with another participant who is also taking part in this study. You will not learn this participant’s ID nor will they learn yours.

In this game, you will have to roll a six-sided die. You are free to choose how to obtain the outcome of a die roll, by using either a physical or a virtual die. Hence, you will roll the die in private, so that the outcome cannot be seen by anyone else. After you will have privately observed the outcome of the die roll, we will ask you to report the result of your die roll.

An outcome of

means the result is “Low”.

An outcome of

means the result is “High”.

The participant you are matched with will also fill out a survey and receive 1.40 GBP but will not roll a die and cannot submit a report. As explained in detail below, this participant’s bonus payment depends on your report.

With 45% you will randomly be assigned to Case 1. With the remaining probability of 55% you will be randomly assigned to Case 2. Before you submit your report, you do not know if you will be assigned to Case 1 or 2.

The table below shows how your and the other participant’s bonus payment depend on the assigned case and your report. In all cases, one of you will receive a bonus of 1.20 GBP and the other one a bonus of 0.20 GBP.

Case	Your report	Bonus
1 45% probability	Low	Each of you has a 50% chance of getting the 1.20 GBP. This is decided by a random draw.
	High	You: 1.20 GBP Other: 0.20 GBP
2 55% probability	Low	You: 0.20 GBP Other: 1.20 GBP
	High	Each of you has a 50% chance of getting the 1.20 GBP. This is decided by a random draw.

The bonus will be sent to you within two days after completion of this study. You will also receive a message via the prolific system informing you about the case and the resulting payment. The other participant will also receive a message with information about your report, the case, and the resulting payment.

[Control questions]
Before we start the game, we want to make sure you have understood the set-up.
Please answer the questions below to the best of your knowledge.

[Table from above was shown here.]
Imagine the following situation:

[Each of the following questions was shown on a separate page. Participants could try twice and were informed about the correct answer afterwards.]
Suppose you have reported “Low.” How likely is it that you get the bonus of 1.20 GBP if you are in Case 1?
Suppose you have reported “Low”. How likely is it that you get the bonus of 1.20 GBP if you are in Case 2?
Suppose you have reported “High.” How likely is it that you get the bonus of 1.20 GBP if you are in Case 1?
Suppose you have reported “High.” How likely is it that you get the bonus of 1.20 GBP if you are in Case 2?

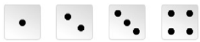
- 0% (Never)
- 50%
- 100% (Always)

[Instructions for Treatment I]

Before the survey starts, you will play a simple game where you can earn an additional bonus that will be added to the 1.40 GBP you receive for the survey.

In this game, you will have to roll a six-sided die. You are free to choose how to obtain the outcome of a die roll, by using either a physical or a virtual die. Hence, you will roll the die in private, so that the outcome cannot be seen by anyone else. After you will have privately observed the outcome of the die roll, we will ask you to report the result of your die roll.

An outcome of



means the result is “Low”.

An outcome of

means the result is “High”.

With 45% you will randomly be assigned to Case 1. With the remaining probability of 55% you will be randomly assigned to Case 2. Before you submit your report, you do not know if you will be assigned to Case 1 or 2.

The table below shows how your bonus payment depends on the assigned case and your report. In all cases, you will receive a bonus of 1.20 GBP or 0.20 GBP.

Case	Your report	Bonus
1 45% probability	Low	You have a 50% chance of getting the 1.20 GBP. This is decided by a random draw.
	High	1.20 GBP
2 55% probability	Low	0.20 GBP
	High	You have a 50% chance of getting the 1.20 GBP. This is decided by a random draw.

The bonus will be sent to you within two days after completion of this study. You will also receive a message via the Prolific system informing you about the case and the resulting payment.

[Control questions]

Before we start the game, we want to make sure you have understood the set-up.

Please answer the questions below to the best of your knowledge.

[Table from above was shown here.]

Imagine the following situation:

[Each of the following questions was shown on a separate page. Participants could try twice and were informed about the correct answer afterwards.]

Suppose you have reported “Low.” How likely is it that you get the bonus of 1.20 GBP if you are in Case 1?

Suppose you have reported ‘‘Low’’. How likely is it that you get the bonus of 1.20 GBP if you are in Case 2?

Suppose you have reported “High.” How likely is it that you get the bonus of 1.20 GBP if you are in Case 1?

Suppose you have reported “High.” How likely is it that you get the bonus of 1.20 GBP if you are in Case 2?

- 0% (Never)
- 50%
- 100% (Always)

[Instructions for Treatment N-belief]

Before the survey starts, you will play a simple game where you can earn an additional bonus that will be added to the 1.40 GBP you receive for the survey.

You will be matched with another participant who is also taking part in this study. You will not learn this participant’s ID nor will they learn yours.

In this game, you will have to roll a six-sided die. You are free to choose how to obtain the outcome of a die roll, by using either a physical or a virtual die. Hence, you will roll the die in private, so that the outcome cannot be seen by anyone else. After you will have privately observed the outcome of the die roll, we will ask you to report the result of your die roll.

An outcome of

means the result is “Low”.

An outcome of

means the result is “High”.

The participant you are matched with will also fill out a survey and receive 1.40 GBP but will not roll a die and cannot submit a report. As explained in detail below, this participant’s bonus payment depends on your report.

You will be assigned to one of two possible cases. Before you submit your report, you do not know if you will be assigned to Case 1 or 2.

The table below shows how your and the other participant’s bonus payment depends on the assigned case and your report. In all cases, one of you will receive a bonus of 1.20 GBP and the other one a bonus of 0.20 GBP.

Case	Your report	Bonus
1	Low	Each of you has a 50% chance of getting the 1.20 GBP. This is decided by a random draw.
	High	You: 1.20 GBP Other: 0.20 GBP
2	Low	You: 0.20 GBP Other: 1.20 GBP
	High	Each of you has a 50% chance of getting the 1.20 GBP. This is decided by a random draw.

The case you are assigned to is determined as follows: In a previous study we did on Prolific, two players were matched. BOTH players rolled a die in private and simultaneously reported their privately observed results. An outcome of one to four meant that the result was “Low”, whereas an outcome of five or six meant that the result was “High”. The bonus payments of these previous players were determined based on the following table.

Report of previous player 1	Report of previous player 2	Bonus
1	Low	Each player has a 50% chance of getting the 1.20 GBP. This is decided by a random draw.
	High	Player 2: 1.20 GBP Player 1: 0.20 GBP
2	Low	Player 2: 0.20 GBP Player 1: 1.20 GBP
	High	Each player has a 50% chance of getting the 1.20 GBP. This is decided by a random draw.

We will randomly select one player from the previous study. If this player’s report was “Low”, you will be assigned to Case 1. If this player’s report was “High”, you will be assigned to Case 2. Note that the participant you are matched with has not taken part in the previous study and cannot influence the case to which you will be assigned.

The bonus will be sent to you within two days after completion of this study. You will also receive a message via the prolific system informing you about the case and the resulting payment. The other participant will also receive a message with information about your report, the case, and the resulting payment.

[Control questions]

Before we start the game, we want to make sure you have understood the set-up.

Please answer the questions below to the best of your knowledge.

[Table from above was shown here.]

Imagine the following situation:

[Each of the following questions was shown on a separate page. Participants could try twice and were informed about the correct answer afterwards.]

Suppose you have reported “Low.” How likely is it that you get the bonus of 1.20 GBP if you are in Case 1?

Suppose you have reported “Low”. How likely is it that you get the bonus of 1.20 GBP if you are in Case 2?

Suppose you have reported “High.” How likely is it that you get the bonus of 1.20 GBP if you are in Case 1?

Suppose you have reported “High.” How likely is it that you get the bonus of 1.20 GBP if you are in Case 2?

- 0% (Never)
- 50%
- 100% (Always)

Can the participant you are matched with influence the case you will be assigned to?

- Yes
- No

Your assignment to Case 1 or Case 2 depends on

- a coin flip
- The report of a randomly selected player from a previous study

[I, C, N, N-belief, C-new treatments]

Getting started:

If you have not done so yet, please get a die or use a virtual one.

Now please roll the die and report either "High" or "Low."

Remember

means the result is "Low", and

means the result is "High".

Please report either "High" or "Low" by checking one of the two boxes below.

- High
- Low

[Belief]

What do you think about the behavior of the other participants in this study? Out of all participants (except you) whose actual result of the die roll was "Low" (outcome 1 to 4), how many will report "High"?

[Answer was recorded via a slider ranging from zero to 100%]

[Belief 2 only for Treatment N-belief]

What do you think about the behavior of the other participants in previous? Out of all participants whose actual result of the die roll was "Low" (outcome 1 to 4), how many have reported "High"?

[Answer was recorded via a slider ranging from zero to 100%]

[Instructions for passive subjects (bystanders) in the N Treatment]

You will receive 1.40 GBP for answering this survey. In addition, you will receive a bonus.

In the following, we will show you the set-up for a study we recently ran on the Prolific platform.

We will ask you for your belief about the behavior of the participants in the study we just ran. Your answer neither influences your fixed payment nor your chance of getting the bonus. For the scientific value of our study, it is important that you state your belief truthfully.

We first show you the exact instructions these participants saw. Then, we will ask you for your belief.

On the next screen, we will show you the exact instructions. All participants also received 1.40 GBP for answering a survey:

[Subjects then saw the instructions of the C, N, or I Treatment. We also conducted the B treatment for the N-belief treatment. The subjects then saw either the C, N-belief, or I Treatment instructions.]

[Belief bystanders]

After reading the instructions of the study we recently ran on the Prolific platform, we now ask you to state your belief. What do you think about the behavior of the participants in this study?

Out of all participants whose actual result of the die roll was "Low" (outcome 1 to 4), how many will have reported "High"?

[Answer was recorded via a slider ranging from zero to 100%]

[If subjects saw the instructions from the C Treatment]

We now explain to you how your bonus is calculated. In a study similar to the one just shown to you, participants also decided on whether to report "Low" or "High". In contrast to the study just shown to you, there was no interaction with other participants. However, one participant was randomly matched with you. If this participant gets the high bonus of 1.20 GBP, you get the low bonus of 0.20 GBP. Also, if this participant gets the low bonus of 0.20 GBP, you get the high bonus of 1.20 GBP. This participant knew that you get 0.20 GBP if they get 1.20 GBP and the other way round.

[If subjects saw the instructions from the N Treatment or N-belief Treatment]

We now explain to you how your bonus is calculated. In the study just shown to you, you played the passive role, i.e. you were randomly matched with one of the participants. If this participant gets the high bonus of 1.20 GBP, you get the low bonus of 0.20 GBP. Also, if this participant gets the low bonus of 0.20 GBP, you get the high bonus of 1.20 GBP.

[If subjects saw instructions from the I Treatment]

We now explain to you how your bonus is calculated. In a study similar to the one just shown to you, participants also decided on whether to report "Low" or "High." One participant was randomly matched with you. If this participant gets the high bonus of 1.20 GBP, you get the low bonus of 0.20 GBP. Also, if this participant gets the low bonus of 0.20 GBP, you get the high bonus of 1.20 GBP. This participant knew that you get 0.20 GBP if they get 1.20 GBP and the other way round.

[Instructions Norm treatment]

We will describe the design of a study on decision making which we ran on the Prolific platform. Participants in this study decided between different options. We will ask you to evaluate the degree at which these possible choices are socially appropriate or not. Specifically, for each possible choice, we will ask you to rate this choice as ‘‘socially appropriate’’ and thus ‘‘consistent with moral or proper social behavior’’ or ‘‘socially inappropriate’’ and thus ‘‘inconsistent with moral or proper social behavior’’.

By socially appropriate, we mean choices that most people agree to be the ‘‘correct’’ or ‘‘ethical’’ choice. Another way to think about this is that, if an individual selects a socially inappropriate choice, then many other people might be angry at the individual for doing so. For each option, please answer as truthfully as possible, based on your own view of what constitutes socially appropriate or socially inappropriate behavior.

To give you an idea of how this task will proceed, we will go through an example and show you how you will report your responses. Note that the example only serves to familiarize yourself with rating choices as socially appropriate or inappropriate. After the example, we will describe the actual situation for which you will rate choices.

Example:

At a local coffee shop a person observes that someone has left their wallet on a table. The person then has four possible choices: (1) take the wallet, (2) ask others nearby if they own the wallet (3) do nothing (4) or hand the wallet to the shop manager.

The person needs to pick one out of these four possible choices.

The table below presents a list of all of the person's possible choices. If this was the actual situation and not the example, we would ask you to rate each of those four choices as "very socially inappropriate", "socially inappropriate", "somewhat socially inappropriate", "somewhat socially appropriate", "socially appropriate" or "very socially appropriate" by ticking the respective box.

Possible choices	Very socially inappropriate	Socially inappropriate	Somewhat socially inappropriate	Somewhat socially appropriate	Socially appropriate	Very socially appropriate
Take the wallet						
Ask others nearby if the wallet belongs to them						
Do nothing						
Hand the wallet to the manager						

Recall that by "socially appropriate" we mean choices that most people agree is the ‘‘correct’’ or ‘‘ethical’’ thing to do. To see how to fill the table suppose hypothetically and arbitrarily that your opinions are as follows: (1) taking the wallet is "very socially inappropriate, " (2) asking others nearby if the wallet belongs to them is "socially appropriate", (3) leaving the wallet where it is "somewhat socially inappropriate", and (4) handing the wallet to the shop manager is "very socially appropriate". Then, you would need to indicate your responses as follows:

Possible choices	Very socially inappropriate	Socially inappropriate	Somewhat socially inappropriate	Somewhat socially appropriate	Socially appropriate	Very socially appropriate
Take the wallet	x					
Ask others nearby if the wallet belongs to them					x	
Do nothing			x			
Hand the wallet to the manager						x

After these explanations we now proceed to our actual study which we ran on Prolific:

Person A, a participant in that study, had to make a choice by picking one of two options. We will ask you to rate each possible choice just as in the example above.

Your bonus payment will be calculated as follows: First, the software will randomly select one of Person A's possible choices. Secondly, the software will randomly match you with another participant that also evaluates Person A's possible choices. If your report for the selected choice matches the report of this participant, you will receive a bonus of 2.50 GBP. Otherwise your bonus will be zero.

For example, if the example above would be the actual task and the possible choice ‘‘Leave the wallet where it is’’, was selected by the software, we would compare your report with the report of the other participant for this choice. If your report had been ‘‘somewhat socially inappropriate’’, then your bonus would be 2.50 GBP if the participant you are matched with also evaluated the choice as "somewhat socially inappropriate", and zero otherwise.

We now present the situation for which we will ask you to rate the participants‘ possible choices. The participants have also been recruited on the Prolific platform. On this screen, you will read the exact instructions that participants in the original study have seen.

[Subjects then saw the instructions of the C, N, or I Treatment.]

[For subjects that saw instructions from the C Treatment]

You have now read the exact instructions that participants in the original study have seen. In short, the situation can be summarized as follows:

Person A was matched with another participant. Both participants had to roll a die in private and report either "High" or "Low."

Both participants would get a bonus, but only one could get the high bonus. After both participants submitted their report, both reports were compared. If only one participant reported "High", this participant got the high bonus whereas the other participant got the low bonus. If both participants submitted the same report (both "High" or both "Low"), a random draw decided who got the high and who the low bonus.

For both participants reporting "High" instead of "Low" increased the probability to get the high bonus by 50%.

[For subjects that saw instructions from N Treatment]

You have now read the exact instructions that participants in the original study have seen. In short, the situation can be summarized as follows:

Person A had to roll a die in private and then report either "High" or "Low." Person A was matched with another passive participant.

Both participants would get a bonus, but only one could get the high bonus. If Person A got the high bonus, the other passive participant got the low bonus. Likewise, if Person A got the low bonus, the other passive participant got the high bonus.

It depends on Person A's report who got the high and who the low bonus. In any case, reporting "High" instead of "Low" increased the probability for Person A to receive the high bonus and, in turn, decreased the probability for the passive participant to receive the high bonus, by 50%.

[For subjects that saw instructions from I Treatment]

You have now read the exact instructions that participants in the original study have seen. In short, the situation can be summarized as follows:

Person A had to roll a die in private and then report either "High" or "Low."

Person A could earn a high or a low bonus, and reporting "High" instead of "Low" increased the probability to receive the high bonus by 50% in any case.

[All subjects in Norm Treatment]

Suppose Person A has rolled the die and the actual result is "Low" (die roll of 1,2, 3, or 4 leads to "Low").

Please rate each of the two possible choices of Person A as "very socially inappropriate", "socially inappropriate", "somewhat socially inappropriate", "somewhat socially appropriate", "socially appropriate," or "very socially appropriate". Please tick the respective box.

Possible choices	Very socially inappropriate	Socially inappropriate	Somewhat socially inappropriate	Somewhat socially appropriate	Socially appropriate	Very socially appropriate
Report low						
Report high						

Appendix B. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.euroecorev.2024.104844.

References

Abeler, J., Nosenzo, D., Raymond, C., 2019. Preferences for truth-telling. Econometrica 87 (4), 1115–1153.

Ariely, D., Bracha, A., Meier, S., 2009. Doing good or doing well? Image motivation and monetary incentives in behaving prosocially. Amer. Econ. Rev. 99 (1), 544–555.

Ashton, M.C., Lee, K., 2009. The HEXACO-60: A short measure of the major dimensions of personality. J. Pers. Assess. 91 (4), 340–345.

Association of Certified Fraud Examiners, 2020. Report to the Nations. Technical Report, Association of Certified Fraud Examiners, Inc.

Bäker, A., Mechtel, M., 2019. The impact of peer presence on cheating. Econ. Inq. 57 (2), 792–812.

Barr, A., Lane, T., Nosenzo, D., 2018. On the social inappropriateness of discrimination. J. Public Econ. 164, 153–164.

Belot, M., Schröder, M., 2013. Sloppy work, lies and theft: A novel experimental design to study counterproductive behaviour. J. Econ. Behav. Organ. 93, 233–238.

Benistant, J., Galeotti, F., Villeval, M.C., 2021. The distinct impact of information and incentives on cheating. In: GATE WP.
Benistant, J., Villeval, M.C., 2019. Unethical behavior and group identity in contests. J. Econ. Psychol. 72, 128–155.
Berger, J., Harbring, C., Sliwka, D., 2013. Performance appraisals and the impact of forced distribution–an experimental investigation. Manage. Sci. 59 (1), 54–68.
Brookins, P., Ryvkin, D., 2014. An experimental study of bidding in contests of incomplete information. Exp. Econ. 17 (2), 245–261.
Brown, J.L., Fisher, J.G., Sooy, M., Sprinkle, G.B., 2014. The effect of rankings on honesty in budget reporting. Account. Organ. Soc. 39 (4), 237–246.
Cadsby, C.B., Song, F., Tapon, F., 2010. Are you paying your employees to cheat? An experimental investigation. BE J. Econ. Anal. Policy 10 (1).
Carpenter, J., Matthews, P.H., Schirm, J., 2010. Tournaments and office politics: Evidence from a real effort experiment. Amer. Econ. Rev. 100 (1), 504–517.
Casal, S., DellaValle, N., Mittone, L., Soraperra, I., 2017. Feedback and efficient behavior. PLOS ONE 12 (4), 1–21.
Chang, D., Chen, R., Krupka, E., 2019. Rhetoric matters: A social norms explanation for the anomaly of framing. Games Econom. Behav. 116, 158–178.
Charness, G., Blanco-Jimenez, C., Ezquerra, L., Rodriguez-Lara, I., 2019. Cheating, incentives, and money manipulation. Exp. Econ. 22 (1), 155–177.
Charness, G., Masclet, D., Villeval, M.C., 2014. The dark side of competition for status. Manage. Sci. 60 (1), 38–55.
Chowdhury, S.M., Gürtler, O., 2015. Sabotage in contests: a survey. Public Choice 164 (1), 135–155.
Cohn, A., Maréchal, M.A., 2018. Laboratory measure of cheating predicts school misconduct. Econ. J. 128 (615), 2743–2754.
Cohn, A., Maréchal, M.A., Noll, T., 2015. Bad boys: How criminal identity salience affects rule violation. Rev. Econ. Stud. 82 (4), 1289–1308.
Conrads, J., Irlenbusch, B., Rilke, R.M., Schielke, A., Walkowitz, G., 2014. Honesty in tournaments. Econom. Lett. 123 (1), 90–93.
Cooper, D.J., Fang, H., 2008. Understanding overbidding in second price auctions: An experimental study. Econ. J. 118 (532), 1572–1595.
Corgnet, B., Martin, L., Ndodjang, P., Sutan, A., 2019. On the merit of equal pay: Performance manipulation and incentive setting. Eur. Econ. Rev. 113, 23–45.
Croson, R., Fatas, E., Neugebauer, T., Morales, A.J., 2015. Excludability: A laboratory study on forced ranking in team production. J. Econ. Behav. Organ. 114, 13–26.
Dai, Z., Galeotti, F., Villeval, M.C., 2018. Cheating in the lab predicts fraud in the field: An experiment in public transportation. Manage. Sci. 64 (3), 1081–1100.
Danilov, A., Biemann, T., Kring, T., Sliwka, D., 2013. The dark side of team incentives: Experimental evidence on advice quality from financial service professionals. J. Econ. Behav. Organ. 93, 266–272.
Dannenberg, A., Khachatryan, E., 2020. A comparison of individual and group behavior in a competition with cheating opportunities. J. Econ. Behav. Organ. 177, 533–547.
Dato, S., Feess, E., Nieken, P., 2019. Lying and reciprocity. Games Econom. Behav. 118, 193–218.
Dato, S., Nieken, P., 2014. Gender differences in competition and sabotage. J. Econ. Behav. Organ. 100, 64–80.
Delgado, M.R., Schotter, A., Ozbay, E.Y., Phelps, E.A., 2008. Understanding overbidding: using the neural circuitry of reward to design economic auctions. Science 321 (5897), 1849–1852.
Diekmann, A., Przepiorka, W., Rauhut, H., 2015. Lifting the veil of ignorance: An experiment on the contagiousness of norm violations. Ration. Soc. 27 (3), 309–333.
Dohmen, T., Falk, A., Fliessbach, K., Sunde, U., Weber, B., 2011. Relative versus absolute income, joy of winning, and gender: Brain imaging evidence. J. Public Econ. 95 (3–4), 279–285.
Dohmen, T., Falk, A., Huffman, D., Sunde, U., 2009. Homo reciprocans: Survey evidence on behavioural outcomes. Econ. J. 119 (536), 592–612.
Faravelli, M., Friesen, L., Gangadharan, L., 2015. Selection, tournaments, and dishonesty. J. Econ. Behav. Organ. 110, 160–175.
Feess, E., Kerzenmacher, F., Muehlheusser, G., 2023. Morally questionable decisions by groups: Guilt sharing and its underlying motives. Games Econom. Behav. 140, 380–400.
Feess, E., Kerzenmacher, F., Timofeyev, Y., 2022. Utilitarian or deontological models of moral behavior - What predicts morally questionable decisions? Eur. Econ. Rev. 149, 104264.
Fehr, E., Schmidt, K.M., 1999. A theory of fairness, competition, and cooperation. Q. J. Econ. 114 (3), 817–868.
Feltovich, N., 2019. The interaction between competition and unethical behaviour. Exp. Econ. 22 (1), 101–130.
Festinger, L., 1954. A theory of social comparison processes. Hum. Relat. 7 (2), 117–140.
Fischbacher, U., Föllmi-Heusi, F., 2013. Lies in disguise–an experimental study on cheating. J. Eur. Econom. Assoc. 11 (3), 525–547.
Fliessbach, K., Weber, B., Trautner, P., Dohmen, T., Sunde, U., Elger, C.E., Falk, A., 2007. Social comparison affects reward-related brain activity in the human ventral striatum. Science 318 (5854), 1305–1308.
Garbarino, E., Slonim, R., Villeval, M.C., 2019. Loss aversion and lying behavior. J. Econ. Behav. Organ. 158, 379–393.
Garcia, S.M., Tor, A., Schiff, T.M., 2013. The psychology of competition: A social comparison perspective. Perspect. Psychol. Sci. 8 (6), 634–650.
Gill, D., Kissová, Z., Lee, J., Prowse, V., 2019. First-place loving and last-place loathing: How rank in the distribution of performance affects effort provision. Manage. Sci. 65 (2), 494–507.
Gneezy, U., 2005. Deception: The role of consequences. Amer. Econ. Rev. 95 (1), 384–394.
Gneezy, U., Niederle, M., Rustichini, A., 2003. Performance in competitive environments: Gender differences. Q. J. Econ. 118 (3), 1049–1074.
Grolleau, G., Kocher, M.G., Sutan, A., 2016. Cheating and loss aversion: Do people cheat more to avoid a loss? Manage. Sci. 62 (12), 3428–3438.
Grote, R.C., 2005. Forced Ranking: Making Performance Management Work. Harvard Business School Press Boston, MA.
Hanna, R., Wang, S.-Y., 2017. Dishonesty and selection into public service: Evidence from India. Am. Econ. J.: Econ. Policy 9 (3), 262–290.
Harbring, C., Irlenbusch, B., 2011. Sabotage in tournaments: Evidence from a laboratory experiment. Manage. Sci. 57 (4), 611–627.
Haß, L.H., Müller, M.A., Vergauwe, S., 2015. Tournament incentives and corporate fraud. J. Corp. Finance 34, 251–267.
Jiang, T., 2013. Cheating in mind games: The subtlety of rules matters. J. Econ. Behav. Organ. 93, 328–336.
Kahneman, D., Tversky, A., 1979. Prospect theory: An analysis of decision under risk. Econometrica 47 (2), 263–292.
Kajackaite, A., Gneezy, U., 2017. Incentives and cheating. Games Econom. Behav. 102, 433–444.
Kampkötter, P., Sliwka, D., 2018. More dispersion, higher bonuses? On differentiation in subjective performance evaluations. J. Labor Econ. 36 (2), 511–549.
Kilduff, G.J., Galinsky, A.D., Gallo, E., Reade, J.J., 2016. Whatever it takes to win: Rivalry increases unethical behavior. Acad. Manag. J. 59 (5), 1508–1534.
Kimbrough, E.O., Vostroknutov, A., 2016. Norms make preferences social. J. Eur. Econom. Assoc. 14, 608–638.
Kirchler, M., Huber, J., Stefan, M., Sutter, M., 2016. Market design and moral behavior. Manage. Sci. 62 (9), 2615–2625.
Kish-Gephart, J.J., Harrison, D.A., Treviño, L.K., 2010. Bad apples, bad cases, and bad barrels: meta-analytic evidence about sources of unethical decisions at work. J. Appl. Psychol. 95 (1), 1.
Kocher, M.G., Schudy, S., Spantig, L., 2018. I lie? We lie! Why? Experimental evidence on a dishonesty shift in groups. Manage. Sci. 64 (9), 3995–4008.
Konrad, K.A., 2009. Strategy and Dynamics in Contests. Oxford University Press.
Kőszegi, B., Rabin, M., 2006. A model of reference-dependent preferences. Q. J. Econ. 121 (4), 1133–1165.
Kőszegi, B., Rabin, M., 2007. Reference-dependent risk attitudes. Amer. Econ. Rev. 97 (4), 1047–1073.
Krupka, E.L., Weber, R.A., 2013. Identifying social norms using coordination games: Why does dictator game sharing vary? J. Eur. Econom. Assoc. 11 (3), 495–524.
Lazear, E.P., 1989. Pay equality and industrial politics. J. Polit. Econ. 97 (3), 561–580.
Lazear, E.P., Rosen, S., 1981. Rank-order tournaments as optimum labor contracts. J. Polit. Econ. 89 (5), 841–864.
Le Maux, B., Masclet, D., Necker, S., 2021. Monetary Incentives and the Contagion of Unethical Behavior. ZEW-Centre for European Economic Research Discussion Paper.

Mazar, N., Amir, O., Ariely, D., 2008. The dishonesty of honest people: A theory of self-concept maintenance. J. Mar. Res. 45 (6), 633–644.
Mitchell, M.S., Baer, M.D., Ambrose, M.L., Folger, R., Palmer, N.F., 2018. Cheating under pressure: A self-protection model of workplace cheating behavior. J. Appl. Psychol. 103 (1), 54.
Mitchell, M.S., Rivera, G., Treviño, L.K., 2023. Unethical leadership: A review, analysis, and research agenda. Pers. Psychol..
Necker, S., Paetzel, F., 2023. The effect of losing and winning on cheating and effort in repeated competitions. J. Econ. Psychol. 98, 102655.
Ordóñez, L.D., Schweitzer, M.E., Galinsky, A.D., Bazerman, M.H., 2009. Goals gone wild: The systematic side effects of overprescribing goal setting. Acad. Manag. Perspect. 23 (1), 6–16.
Pierce, J.R., Kilduff, G.J., Galinsky, A.D., Sivanathan, N., 2013. From glue to gasoline: How competition turns perspective takers unethical. Psychol. Sci. 24 (10), 1986–1994.
Piest, S., Schreck, P., 2021. Contests and unethical behavior in organizations: a review and synthesis of the empirical literature. Manag. Rev. Q. 71 (4), 679–721.
Potters, J., Stoop, J., 2016. Do cheaters in the lab also cheat in the field? Eur. Econ. Rev. 87, 26–33.
Ruedy, N.E., Schweitzer, M.E., 2010. In the moment: The effect of mindfulness on ethical decision making. J. Bus. Ethics 95, 73–87.
Schindler, S., Pfattheicher, S., 2017. The frame of the game: Loss-framing increases dishonest behavior. J. Exp. Soc. Psychol. 69, 172–177.
Schweitzer, M.E., Ordóñez, L., Douma, B., 2004. Goal setting as a motivator of unethical behavior. Acad. Manag. J. 47 (3), 422–432.
Schwieren, C., Weichselbaumer, D., 2010. Does competition enhance performance or cheating? A laboratory experiment. J. Econ. Psychol. 31 (3), 241–253.
Shalvi, S., 2012. Dishonestly increasing the likelihood of winning. Judgm. Decis. Mak. 7 (3), 292–303.
Sheremeta, R.M., 2010. Experimental comparison of multi-stage and one-stage contests. Games Econom. Behav. 68 (2), 731–747.
Sobel, J., 2020. Lying and deception in games. J. Polit. Econ. 128 (3), 907–947.
To, C., Kilduff, G.J., Rosikiewicz, B.L., 2020. When interpersonal competition helps and when it harms: An integration via challenge and threat. Acad. Manag. Ann. 14 (2), 908–934.
Welsh, D.T., Ordóñez, L.D., 2014. The dark side of consecutive high performance goals: Linking goal setting, depletion, and unethical behavior. Organ. Behav. Hum. Decis. Process. 123 (2), 79–89.
Zoltners, A.A., Sinha, P., Lorimer, S.E., 2016. Wells Fargo and the Slippery Slope of Sales Incentives. Harvard Business School Press.