I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 14014

The Distinct Impact of Information and Incentives on Cheating

Julien Benistant
Fabio Galeotti
Marie Claire Villeval

JANUARY 2021

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 14014

The Distinct Impact of Information and Incentives on Cheating

Julien Benistant
University of Lyon

Fabio Galeotti
University of Lyon and GATE

Marie Claire Villeval
University of Lyon, GATE and IZA

JANUARY 2021

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

The Distinct Impact of Information and Incentives on Cheating*

We study a dynamic variant of the die-under-the-cup task where players can repeatedly misreport the outcomes of consecutive die rolls to earn more money, either under a non- competitive piece rate scheme or in a two-player competitive tournament. In this dynamic setting we test (i) whether giving continuous feedback (vs. final ex post feedback) on the opponent's reported outcome to both players encourages cheating behavior, and (ii) to what extent this influence depends on the incentive scheme in use (piece rate vs. tournament). We also vary whether the opponent is able to cheat or not. We find that people lie more when placed in a competitive rather than a non-competitive setting, but only if both players can cheat in the tournament. Continuous feedback on the counterpart's reports increases cheating under the piece-rate scheme but not in a competitive setting. Our results provide new insights on the role that feedback plays on cheating behavior in dynamic settings under different payment schemes, and shed light on the origins of the effect of competition on dishonesty.

JEL Classification: C92, M52, D83

Keywords: dishonesty, feedback, peer effects, competitive incentives, experiment

Corresponding author:
Marie Claire Villeval
Université de Lyon
CNRS
GATE UMR 5824
93 Chemin des Mouilles
F-69130 Ecully
France

E-mail: villeval@gate.cnrs.fr

* We are grateful to A. Hopfensitz, A. Kajackaite and M. Lefebvre for useful comments. We also thank Q. Thevenet for research assistance. This research was funded by the LABEX CORTEX (ANR-11-LABX-0042) of Université de Lyon within the program "Investissements d'Avenir" (ANR-11-IDEX-007) operated by the French National Research Agency. It has benefited from the support of IDEXLYON from Université de Lyon (project INDEPTH) within the Programme Investissements d'Avenir (ANR-16-IDEX-0005), and from the support of the project DECISION of the French National Research Agency (ANR-19-CE26-0019). This paper is also based upon research supported by the Thomas Jefferson Fund of the Embassy of France in the United States and the FACE Foundation.

1 Introduction

Occupational fraud and professional misconduct are relatively widespread in market societies (for recent examples in the financial sector, see Cohn et al., 2014; Piskorski et al., 2015; Brown and Minor, 2016; Egan et al., 2019). An aggressive market culture together with a rise of competition between and within organizations are frequently evoked to explain the erosion of moral norms in occupational settings. Employees have to compete hard to get a job, a promotion or a bonus, which may give incentives to inflate CVs, fabricate positive results or adopt fraudulent practices; performance pay schemes may encourage some employees to sabotage competitors, cheat on their own performance, or deceive customers.[1] The responsibility of performance pay schemes in moral erosion is, however, difficult to disentangle from the impact of the feedback on relative performance that is provided to the employees who receive such incentives. Indeed, such feedback can, on its own, encourage cheating through peer effects if it signals, even indirectly and imperfectly, the likely misconduct of other employees. Unfortunately, company data on employees' cheating behavior are rare and they usually do not allow researchers to separate the pure effect of competition from the effect of feedback. Our aim in this study is to identify the separate effect of a competitive incentive scheme from the effect of social information on employees' cheating behavior, by means of a controlled laboratory experiment.

Financial incentives are an important determinant of the level of competition between employees within a firm. A common strategy in companies is to reward employees based on their performance (Prendergast, 1999), either by paying them an individual piece rate or by adopting tournament incentive schemes.[2] But if the development of such strategies has an important impact on productivity levels, a less documented effect is how it affects morality and through which channel. While tournaments generally induce higher effort levels compared to piece rates (e.g., Bull et al., 1987; Van Dijk et al., 2001), they also tend to trigger more cheating behavior (e.g., Cadsby et al., 2010; Schwieren and Weichselbaumer, 2010; Belot and Schröder, 2013; Faravelli et al., 2015; Rigdon and D'Esterre, 2015); for a recent survey, see Gangadharan et al. (2020). But is this due to the materialistic incentive per se, to the fear of a biased competition due to the possible unethical behavior of the competitor, or to the cheating emulation generated by the feedback about the competitor's performance? Indeed, information on others' misbehavior, even imprecise, may lead

[1] Examples of this abound, especially in the domain of credence goods: salesmen lying to customers about the quality of their products in order to sell more and win a sales contest (Poujol, 2009); public officers cheating in national promotion tests (Fitch, 2013); applicants exaggerating or misrepresenting details about their experience or skills in order to increase their chance of being hired in a company (Weiss and Feldman, 2006; Levashina and Campion, 2007).

[2] Piece rates are extensively used to incentivize employees who work in sectors where output is easily observed, such as agricultural workers, salespersons or taxi drivers (*e.g.*, at Uber and Lyft) (Lazear, 2018). Tournaments and relative rankings are commonly employed to determine promotions, hiring, raises and bonuses (see Lazear, 2018). Consider, for example, first-level managers or front-line supervisors who compete for a bonus, assistant professors who compete for a tenure-track position or retailers who participate in a sales contest for a prize.

people to update their beliefs about the existing social norm in the considered context and change their perception of what is socially acceptable. This may, in turn, influence their tendency to cheat. Previous studies found evidence of such peer effects on cheating in non-competitive settings by showing that people tend to cheat more when they are informed, even imperfectly, about others' behavior (see e.g., Gino et al., 2009; Robert and Arnab, 2013; Rauhut, 2013; Lefebvre et al., 2015; Charroin et al., 2021).

Yet, and perhaps surprisingly, the pure impact of social information under competitive pay schemes on cheating behavior has attracted less attention.[3] In such competitive settings, individuals may have a pecuniary interest to mimic their opponents' cheating behavior, and this could increase even more cheating behavior compared to non-competitive settings. But feedback on others' behavior also reduces the uncertainty about the relative performance of their rivals and this may lead to moderate cheating since the success of the competition does not depend on absolute performance levels but on outperforming the competitor. In this study, we shed light on this unexplored area of intersection between incentives and social information by investigating how feedback on a counterpart's performance influences individuals' cheating behavior under competitive *versus* non-competitive pay schemes.

Previous experimental studies on the effects of incentives and feedback on dishonesty have mainly focused on static settings where subjects make one-shot decisions or play over very few rounds. Thus, little is known about the dynamic effects of different pay schemes and feedback on cheating behavior. A recent meta-analysis of Abeler et al. (2019) suggests that dishonest behavior does not change over time. On the other hand, Garrett et al. (2016) found evidence of a gradual escalation of dishonesty and identified signal reduction in the amygdala as the neural mechanism behind such escalation. Besides these conflicting results in non-competitive settings, we do not know how repetition affects cheating under competitive pay schemes and whether the provision of social information affects or not the evolution of cheating behavior over time. Our study addresses also these questions by considering a dynamic setting where subjects can adjust their behavior over time.

To address all the aforementioned research questions, we designed a laboratory experiment where subjects were matched in pairs and played individually and repeatedly over several periods a variant of the die-under-the-cup task (Shalvi et al., 2011; Fischbacher and Föllmi-Heusi, 2013). In each period, subjects had to roll a die in private with no scrutiny and to report the outcome in their computer. Their final performance was computed as

[3] An exception is sabotage. Previous studies showed that individuals informed of their opponents' sabotage behavior increase their own sabotage activities under competitive pay schemes (e.g., Harbring et al., 2007). This is usually motivated by retaliation. While sabotage is about increasing one's own chances of winning a contest by reducing the performance of the opponent, cheating consists in artificially improving one's own output, for example by taking performance-enhancing drugs or forging data (Charness et al., 2014). Both sabotage and cheating directly affect the probability of winning a contest, but their behavioral implications may be different (see e.g. Rigdon and D'Esterre, 2015; Benistant and Villeval, 2019).

the sum of their individual reports in all periods. We varied between-subjects whether participants were paid according to a piece rate or a tournament payment scheme, and whether they received a continuous feedback about their counterpart's reports at the end of each period or only at the end of the task. Feedback provides a noisy signal about the dishonesty of the other player since the reports of several high outcomes may indicate luck or misreporting.

We also designed a variant of the tournament condition where only one pair member could physically roll the die and cheat by misreporting the outcome. The other player was passive; an electronic die was rolled virtually and the outcome was automatically reported by the computer. This allows us to test whether the difference in cheating between the piece-rate and the tournament payment scheme, if any, is driven by the competitive nature of the incentive or by the individuals' beliefs about the dishonest behavior of the rival. If previous differences disappear in the new treatment, this would indicate that dishonest reporting in standard tournaments is driven by the fear of losing an unfair competition and not by the competition itself. If not, this would suggest that the competitive pressure itself generates misbehavior regardless of the counterpart's behavior, for monetary and possibly also non-monetary reasons (e.g., Parco et al., 2005; Cason et al., 2018). To test the effect of feedback on relative performance in this one-sided cheating setting, we also varied whether the feedback on the relative performance was continuous or given only at the end of the competition.

Our results show that participants over-reported the die outcomes in all treatments. Tournament incentives increased over-reporting compared to the individual piece-rate scheme but the difference was almost nullified when the opponent could not cheat. Continuous feedback on the counterpart's reports increased dishonesty compared to a final feedback under the piece-rate pay scheme, suggesting the existence of peer effects, but not in tournaments. Despite the latter result, participants reacted to the perceived level of honesty of their counterpart both in tournaments and in the piece-rate pay scheme.

To better understand the role of beliefs when subjects were continuously informed about their counterpart's performance, we ran a follow-up experiment where we elicited the subjects' beliefs about the prevalence of cheating behavior in each period under tournament and piece-rate incentives, respectively. The results of this second experiment reveal that individuals underestimated more the extent of cheating behavior when they were paid a piece rate compared to when they competed in a tournament. Furthermore, subjects adjusted their perception of the norm and their behavior depending on the social information they received about their counterpart.

The next section reviews the related literature. Section 3 describes the experimental design and procedures of the main experiment. Section 4 reports the results. Section 5 introduces the second experiment and summarizes its main results. Section 6 discusses the findings of both experiments and concludes.

2 Related Literature

Our study is related to three strands of the literature. The first one concerns the impact of relative performance feedback policies on performance under different pay schemes. Previous studies found contrasted effects of feedback on performance (for a recent survey, see Villeval, 2020).[4] Under a flat wage scheme Falk and Ichino (2006) and Mas and Moretti (2009) found strong emulation effects when employees can observe or are observed by other employees working. Bandiera et al. (2010) identified a positive net effect of social incentives on a firm's performance, suggesting that firms can use social feedback as an incentive to motivate workers in place of monetary incentives. In contrast, Barankay (2019) found that removing performance rankings increased employees' productivity in a company. In tournament settings, Gill and Prowse (2012) found that information about the competitor's performance may discourage effort because of disappointment aversion, while Fershtman and Gneezy (2011) found no effect. Eriksson et al. (2009) found no effect of relative feedback on the quantity of effort under a piece rate and a tournament, but identified a negative effect on the quality of effort. Kuhnen and Tymula (2012), Ludwig and Lünser (2012) and Delfgaauw et al. (2014) found heterogeneous effects of feedback depending notably on one's relative performance. In networks where people were paid a piece rate, Beugnot et al. (2019) showed strong positive effects of peers' performance feedback on effort, except for females when information flowed both ways (from the peers to the subject and *vice versa*).

Our study connects to the aforementioned literature by investigating how different payment schemes and feedback policies affect behavior. However, in contrast to that literature, we focus on cheating behavior rather than effort. Since in real settings performance may depend on both effort and potential cheating, past results on relative performance feedback on effort and our novel findings can be seen as complementary.

Our study is also related to the literature on the relationship between competition and unethical behavior. A number of experimental studies report that dishonest behavior occurs more frequently in competitive than non-competitive settings, with some variations in the size, stability and distribution of the effect (Cadsby et al., 2010; Schwieren and Weichselbaumer, 2010; Belot and Schröder, 2013; Faravelli et al., 2015; Rigdon and D'Esterre, 2015; Aydogan et al., 2017).[5] This literature, however, does not explore what is the role

[4] Evidence comes from both observational data in the field and lab experiments. Importantly, in a meta-analysis Herbst and Mas (2015) found that peer effects in the laboratory generalize to the field.

[5] We focus on the literature on tournaments and contests as this is the main target of our study. Other studies investigated the effect of market competition on unethical behavior (e.g., Falk and Szech, 2013; Rabanal and Rud, 2018; Feltovich, 2019), and the spillover effects of competition between employees on unethical behavior (e.g., Buser and Dreber, 2016; Schurr and Ritov, 2016; Banerjee et al., 2018). It is also worth noting the existence of earlier work showing that competition increases sabotage (e.g., Lazear, 1989; Konrad, 2000; Harbring et al., 2007; Carpenter et al., 2010; Harbring and Irlenbusch, 2011). Finally, some studies looked at the link between the strength of the competition and the occurrence of cheating (Conrads et al., 2013; Cartwright and Menezes, 2014).

played by the information about the opponent's performance to determine dishonest behavior. Our contribution is to identify the direct impact of feedback about the counterpart's relative performance on cheating and to investigate whether this effect depends on the incentive scheme in use and for what reason.

Finally, there is a literature on peer effects and dishonesty that is relevant to our study.[6] An influence of peers' behavior on criminal activity has been identified in various studies (Glaeser et al., 1996; Patacchini and Zenou, 2008; Corno, 2017), especially regarding juvenile deliquency (Patacchini and Zenou, 2009; Díaz and Patacchini, 2020). Laboratory experiments have contributed to this field by proposing simpler methods to identify endogenous peer effects. Gino et al. (2009) found that cheating is contagious in a real-effort experiment in which participants could misreport their performance. Using a coin tossing task, Fosgaard et al. (2013) observed that men cheat significantly more when they are told that others have cheated. In a die task, Diekmann et al. (2015) found that people lie more if they are informed about the lying behavior of others. When the die task is repeated, Rauhut (2013) showed that subjects who are told about others' reports lie more (less) when they underestimated (overestimated) others' lying behavior compared to uninformed subjects. Similar results were reported by Soraperra et al. (2017) in a die task, by Robert and Arnab (2013) in the context of a sender-receiver game, and by Bäker and Mechtel (2019) and Lauer and Untertrifaller (2019) in real-effort experiments.

However, such peer effects are highly heterogeneous. In a tax evasion experiment, Fortin et al. (2007) found no evidence that people imitate others' evading behavior but identified strong exogenous peer effects focused on income differences. Lefebvre et al. (2015) observed that tax evasion increased after previous examples of low compliance were disseminated, whereas examples of high compliance did not make people more honest, an asymmetry in contagion that has been replicated by Dimant (2018) in a different setting. Finally, Charroin et al. (2021) found evidence of peer effects but only among individuals who were already behaving dishonestly in isolation. However, none of these studies investigated the potential contagion effect of peers' dishonesty in a tournament setting or evaluated the interaction between incentives and feedback, which is the main aim of our study.

3 Design and Procedures of the Main Experiment

In this section, we introduce the design of the main experiment and detail the procedures.

[6] There is also a very recent literature looking at how the perception of cheating affects fairness views and redistribution (Bortolotti et al., 2017; Klimm, 2019). In our experiment, subjects cannot know for certain whether the opponent is cheating but they may form an opinion. If subjects suspect cheating, they might perceive it as unfair, and decide to cheat as well.

3.1 Design

Our experiment is based on a variant of the die-under-cup task (Fischbacher and Föllmi-Heusi, 2013) repeated for 24 periods. We used a six-faced die with three possible outcomes (represented by colors instead of numbers) to increase the statistical power of our analyzes, like in Dai et al. (2018).[7] The possible outcomes of the die roll were "red", "yellow" or "blue" with equal probability (1/3). The die was inside a sealed cup to increase privacy, as in Shalvi et al. (2011). In each period, participants were asked to roll the die and report the outcome. They earned 0, 1 or 2 points if they reported a blue, a yellow or a red face, respectively. Participants were allowed to roll the die more than once but they were explicitly told to report only the outcome of the first roll. Participants' performance in this task was computed as the sum of points earned across the 24 periods.[8]

We purposely chose a task where performance is only determined by luck and honesty in order to isolate the pure effect of feedback and incentives on cheating behavior.[9] In particular, we get rid of a number of possible confounds that would arise if performance was also determined by effort. First, our results are not influenced by participants' innate abilities or effort, which is known to vary across different payment schemes and feedback policies (see Section 2); second, subjects' beliefs about the true performance of the counterpart are constant within each pair and across conditions; third, we prevent subjective entitlements from arising. Our task also captures real-world settings where an agent over-reports a performance measure (*e.g.*, the quality of a product, company earnings) for personal benefits, and this measure is determined in part or fully by external factors (*e.g.*, the volatility of markets, the quality of work of a previous team). It also reflects what previous studies in organizational economics have used to measure the true and reported performance of "workers" in laboratory "firms" (see d'Adda et al., 2017).

We implemented a 2x2 between-subject design. In all treatments participants were matched in fixed pairs for the whole duration of the experiment. One dimension that we manipulated was the payment scheme: either participants were paid based on a piece-rate scheme or they had to enter a two-player tournament. The other dimension varied whether participants were informed continuously (*i.e.*, at the end of each period) or only at the end of the 24 periods about the reports of their counterpart in each period. We also added a variant of the tournament treatment where only one participant in each pair could misreport. This variant was motivated by the identification of the role of beliefs about the counterpart's dishonesty. We ran two versions of this tournament with asymmetric

[7]One potential disadvantage of having only three outcomes is that the signal is noisier as there is 1/3 chance that the report is honest (compared to 1/6 with a six-faced die). Note, however, that the noise decreases with repetition.

[8]Subjects performed the same task across treatments. Thus, the potential effects of lucky or unlucky draws in the first few periods equally apply to all conditions and cannot drive differences across treatments.

[9]One can think of our task as a magnifying glass that helps us to better visualize the role that different payment schemes and feedback policies play on cheating behavior.

cheating opportunities, one where the feedback on relative performance was continuous and one where it was given only at the end of the 24 periods. Hence, in total, we ran 6 treatments.

3.1.1 Payment Schemes

One dimension that we manipulated was how participants' payoffs were determined. In the *Piece Rate* treatment (PR, hereafter), reporting a blue outcome paid 0 points, a yellow outcome paid 1 point and a red outcome paid 2 points. Each point was worth €0.18. Participants' total payoff in this task was determined by the sum of the points accumulated throughout the 24 periods, regardless of the counterpart's report. In the *Tournament* treatment (TR, hereafter), the pair member with the highest number of points accumulated at the end of the 24 periods earned €0.36 per point while the counterpart earned €0. As such, the expected payoffs were the same in the two payment schemes, assuming a 50 percent chance of winning the tournament.

We paid subjects based on the sum of their reports in all the 24 periods instead of one or few rounds selected at random for different reasons. First, we thought it was more natural: it reflects more what typically happens in the field. For example, in sales contests, salesmen usually compete over cumulative sales during several weeks (see e.g. Delfgaauw et al., 2013). Second, it prevented frequent ties in the tournament (two competitors having the same total performance). Third, we did not add background risk to the decision-making setting. Finally, we adopted the same payment procedure in all treatments. Hence, any distortions cannot explain treatment differences.

3.1.2 Information

The second dimension that we manipulated was the nature of the feedback received by the participants about their counterpart's reports. In the *Final Feedback* treatment (FF, hereafter), participants were only informed at the end of the experiment about the reports made by their counterpart in each of the 24 periods. Thus, in this treatment there was uncertainty about one's relative performance and about the honesty of the counterpart until the end of the experiment. Providing detailed information about the counterpart's reports at the end of the experiment allowed us to hold constant the observability of the participants' reports made by their counterpart across treatments.

In the *Continuous Feedback* treatment (CF, hereafter), participants were informed about the reports made by their counterpart at the end of each period. A history box with the participant's and the counterpart's reports was displayed on the participant's screen (see Figure 1). As soon as both subjects had entered their reports, the box was updated with the new information, and a new period began. Thus, in this treatment participants were informed about the difference in performance between themselves and their counterpart, and its evolution over time, in each period. This feedback provided also a noisy information about the honesty of the counterpart. Over time, participants could perceive whether the counterpart was dishonest or not by comparing the distribution of reports with the theoretical uniform distribution, but they could never be certain of whether or not a report of the counterpart in a given period was honest.

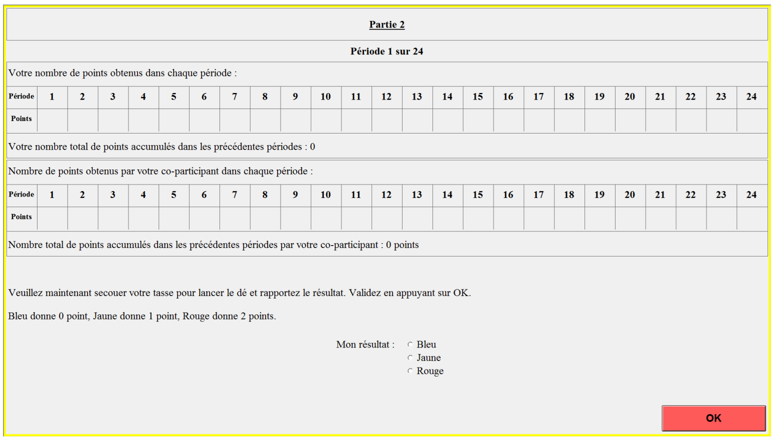


Figure 1: Screenshot of the Decision Screen, Continuous Feedback Treatment

Notes: Translation of the text: Period 1 out of 24 // Your number of points obtained in each period // Your total number of points accumulated in past periods // Number of points obtained by your co-participant in each period // Total number of points accumulated by your co-participant in past periods // Please now shake the cup to roll the die and report the outcome. Validate by pressing 'ok'. Blue pays 0 points, Yellow pays 1 point, Red pays 2 points // My outcome: Blue/Yellow/Red.

3.1.3 Reporting

In the *Tournament* condition, we also varied whether only one or both participants in each pair had to physically roll a die and report the outcome. In the *One-Sided* tournament treatment (TR1, hereafter), only one pair member was requested to roll the die and report the outcome. The other competitor was passive. She or he was shown a video of a die roll and the outcome was automatically reported by the computer. The role of the passive player was randomly assigned to one member in the pair before the first period, and roles were kept fixed throughout the 24 periods. Both subjects within each pair were aware of the profile of their competitor (passive or active). All the other aspects of the task were identical to the other treatments.

The one-sided condition was implemented only for the tournament payment scheme. Indeed, if dishonesty is more widespread in the tournament condition than in the piece-rate condition, it might be because of the competitive nature of the tournament scheme *per se* or because people want to restore the fairness of the competition if they anticipate that their counterpart is cheating. This additional treatment manipulation allowed us to mute the second channel. Table 1 summarizes our treatments.

Table 1: Treatments and Conditions

	Piece-Rate (PR)	Tournament (Two-Sided) (TR)	Tournament (One-Sided) (TR1)
Final Feedback (NF)	PR-FF (42)	TR-FF (44)	TR1-FF (58)
Continuous Feedback (CF)	PR-CF (72)	TR-CF (78)	TR1-CF (62)

Notes: The number of subjects is reported in parenthesis for each condition. Differences in the number of participants reflect (a) different attendance rates across sessions; (b) the fact that we needed more observations in the Continuous Feedback conditions, as the independent observation is the pair and not the participant.

3.1.4 Additional Measures

Both at the beginning of the session (before being informed about the main task) and at the end of the 24 periods, we asked participants to rate their level of happiness and nervousness (emotional arousal), using Self-Assessment Manikin scales (SAM).[10] Before the main task, we also elicited the participants' risk attitude using the procedure of Gneezy and Potters (1997).[11] Finally, at the end of the experiment and after the second round of questions about emotions, we elicited the participants' beliefs about the others' reports. They had to state how many subjects out of 24 (excluding themselves and their counterpart) reported a red, a yellow and a blue outcome in a given period.[12] We elicited the participants' beliefs for the first, the thirteenth and the last period. The objective was to get a rough idea about the participants' perception of the empirical norm in the experiment, and how this evolved over time. One period and one color were randomly selected for payment. Participants received €1.5 for a correct prediction, €1 if their prediction differed within plus or minus 1 from the actual number, €0.5 if it differed within plus or minus 2, and €0 otherwise.

We acknowledge that these beliefs are probably biased since they were elicited at the end of the experiment, after informing the subjects about the distribution of reports made by their counterpart. We chose not to elicit the beliefs during the task to prevent subjects from focusing their attention on those beliefs and on how to best respond to them. Indeed, previous studies found that this may generate unwanted effects and confound the results (see, e.g., Croson, 2000; Gächter and Renner, 2010, in the context of public good games). In the follow-up experiment, we elicited the beliefs between periods, allowing these potential biases to manifest (see details in Section 6).

[10]The Online Appendix A reports the instructions and the illustrations used to report feelings of happiness and nervousness.

[11]Participants were endowed with 150 euro cents and asked to choose how much to invest in a lottery. This lottery had two possible outcomes with equal probability of being selected. In case of a failure, participants lost their investment and in case of a success, the amount invested was multiplied by 2.5. While a risk neutral participant should invest all of this endowment, the amount of the endowment not invested captures the degree of risk aversion. Subjects received a feedback on their earnings in this task only at the end of the session.

[12]Subjects were told that the 24 participants could be from the current session and/or a previous one.

3.2 Procedures

We conducted the experiment at GATE-Lab, Lyon, France. 356 subjects were recruited using HROOT (Bock et al., 2014). 89.9% were students from local engineering, business, and medical schools. The remaining participants were mostly former students while few people came from the general public (*e.g.*, clerical workers from the campus). We conducted 2 sessions for each treatment except for the TR-CF and PR-CF treatments for which we conducted 3 sessions. More sessions were needed for these treatments, as the independent observation is the pair and not the participant. Table B.1 in the Online Appendix B summarizes the participants' characteristics. The experiment was programmed with z-Tree (Fischbacher, 2007). The instructions were directly displayed on the computer screen of the subjects (see Online Appendix A for a translated version of the instructions). Participants received the instructions for the die task only once the mood and the risk elicitation tasks were completed. The instructions used neutral language and clarification questions were answered in private. Each session lasted approximately 45 minutes. Mean earnings were €12.92 (SD=5.83), including a €5 show-up fee.

4 Behavioral Conjectures

A significant amount of evidence shows that individuals suffer from a finite moral cost of cheating (*e.g.,* Mazar et al., 2008; Fischbacher and Föllmi-Heusi, 2013; Kajackaite and Gneezy, 2017; Abeler et al., 2019). Part of this cost can be explained by a potential loss of reputation from being perceived as a cheater by others or even in one's own eyes (Dufwenberg and Dufwenberg, 2018; Gneezy et al., 2018; Abeler et al., 2019; Khalmetski and Sliwka, 2019). When subjects' actions cannot be scrutinized by the experimenter or by other participants, those who lie tend to do it partially instead of fully.[13] In our experiment, scrutiny by the experimenter and by the counterpart (either continuously or at the end) is limited. Indeed, nobody knows what the truth is in each period and, over time, the distribution of reports only gives a probability on whether the individual cheated or not. Hence, we expect a significant fraction of the participants to cheat but mostly partially to lower their chance of being detected. Note that in our repeated environment a subject cheats partially if (s)he misreport the outcome of the die role in some periods but does not cheat maximally in all.[14] We state our first conjecture as follows:

[13]For instance, Abeler et al. (2019) found that 59% of those who cheat report the highest possible number when the experimenter can observe the true outcome. In contrast, when the true outcome is not observable, this number drops to 27%. This is an estimate and it is computed as $[(32.9\%-10\%)\times(10/9)]/(100\%-6.4\%)$ where the numerator is the fraction of subjects who behave as income maximizers (see Fischbacher and Föllmi-Heusi, 2013), while the denominator is the proportion of all cheaters.

[14] As in previous studies on cheating, we assume that no participant misreports a die roll to his or her disadvantage. Our definition of partial cheating is a natural extension of the one used in one-shot cheating games if one considers the 24 periods all together. A partial cheater is someone who neither tells the truth nor maximizes his or her payoff (Fischbacher and Föllmi-Heusi, 2013) throughout the 24 rounds. This definition applies, for example, to subjects who *sometimes* or *all the time* report yellow instead of blue, or

Conjecture 1 *(Cheating Behavior):* Most participants cheat partially by over-reporting their actual outcomes but not in every periods.

Based on the literature on cheating behavior under competition and because payoffs depend on the ability to outperform the competitor, we expect that individuals who face competitive incentives are more likely to behave dishonestly than subjects who are paid an individual piece rate. Therefore, we anticipate more dishonesty in the tournament than in the piece-rate payment scheme. Furthermore, we expect that when the opponent cannot cheat in the tournament, participants are less likely to over-report their own outcome than when their opponent can cheat. If participants suffer from a cost of cheating, they have no interest to misreport more than necessary when they know that their opponent cannot cheat. We thus expect them to cheat just enough to win the tournament. We also expect cheating to be similar under the one-sided tournament and the piece-rate pay scheme. If we assume that the marginal benefit of cheating does not influence the willingness to cheat—as previous studies have demonstrated (see, e.g., Fischbacher and Föllmi-Heusi, 2013; Kajackaite and Gneezy, 2017; Abeler et al., 2019)[15]—subjects should, in expectation, cheat to the same extent in the one-sided tournament and under a piece-rate pay scheme, even if each supplementary misreported point in the former increases the payoff twice as much as in the latter. Our second conjecture is as follows:

Conjecture 2 *(Competition):* Regardless of the feedback, (a) participants cheat more under a tournament payment scheme than an individual piece-rate scheme; (b) they cheat less when their opponent is not allowed to cheat compared to when (s)he has such possibility; and (c) they are equally likely to cheat in the one-sided tournament and under the piece-rate payment scheme.

Previous studies reported a behavioral contagion effect of dishonesty in settings that are comparable to our individual piece-rate environment with continuous feedback (*e.g.,* Gino et al., 2009; Diekmann et al., 2015; Charroin et al., 2021). One mechanism behind these peer effects is a preference for conformity and it may be channeled by changes in beliefs (Rauhut, 2013): individuals who are informed about the prevalence of cheating behavior in their group, update their beliefs about the appropriateness of cheating, and adjust their behavior accordingly to imitate others.[16] In particular, underestimators adjust their cheating behavior upward, while overestimators downward. This contagion effect may be

to subjects who *sometimes* report red instead of yellow.

[15] In cheating games, lying does not increase even if one multiplies the marginal benefit by a factor of 50 as long as subjects fear to be exposed as liars (Kajackaite and Gneezy, 2017).

[16] In our dynamic setting, the information flows bidirectionally: a player does not only receive social information from the counterpart but also transmits information regarding his or her own behavior to the counterpart (like in previous studies on peer effects where social information is provided simultaneously—see, e.g., Falk and Ichino 2006). Sharing information may have an effect on behavior, too. Disentangling the effect of sharing vs. receiving social information is an interesting line for future research but it is beyond the scope of the current study. In deriving our conjectures, we assume that the effect of receiving social information dominates the one of sharing.

asymmetric, as bad examples may have more influence than good ones (e.g. Lefebvre et al., 2015; Dimant, 2018). A possible explanation is that most people tend to underestimate the extent of cheating in non-competitive settings (Diekmann et al., 2015). Assuming that this also applies to our setting (and we provide evidence of this in the follow-up experiment), we expect participants who are paid an individual piece rate to behave more dishonestly when they are continuously informed about their counterpart's reports.

Subjects in the tournament treatments may expect the opponent to cheat more compared to subjects who are paid a piece rate to avoid a null payoff. The effect of a continuous feedback in the tournament depends on whether the subjects' beliefs regarding the cheating behavior of their opponent are more or less precise compared to the beliefs of the subjects who are paid a piece rate. If subjects in the tournament tend to underestimate more the extent of cheating behavior, they will cheat more when they receive a continuous feedback compared to when they receive a final feedback. On the opposite, they will not cheat more if their beliefs are more accurate. They might even cheat less if they overestimate the extent of cheating behavior. If the opponent cannot cheat, the signal they receive about their counterpart's reports is uninformative regarding the prevalence of cheating in the population. So we expect no difference in cheating behavior between the one-sided tournaments with continuous and final feedback.[17]

In addition to the aforementioned effects, feedback bears a strategic value in the tournament. Unlike in the individual piece-rate setting, if one does not mimic the dishonest behavior of the opponent (s)he will lose the tournament and earn nothing. Hence, in the tournament we expect people to stick more to their opponent's performance when they are continuously informed about this performance. Our third conjecture is as follows:

Conjecture 3 *(Feedback):* (a) Under a piece-rate pay scheme, subjects cheat more when they receive continuous rather than final feedback; in the tournament, the difference depends on the players' beliefs about the prevalence of cheating behavior in the population; (b) In one-sided tournaments, both continuous and final feedback have the same effect on misreporting; (c) Under continuous feedback, participants in the tournament are more likely to adjust their level of cheating to the reports of their counterpart compared to subjects in the individual piece-rate scheme in order not to lose the tournament.

Regarding the effect of feedback, our main experiment allows us to examine whether subjects cheat more (or less) when they receive a continuous rather than a final feedback. However, we cannot identify whether this is because subjects underestimate more (or less) the extent of cheating behavior in the population. We explore this in the follow-up experiment where we elicit the subjects' beliefs during the task.

[17] Note that this holds for risk neutral individuals. A risk averse subject might cheat more when feedback is final rather than continuous to increase his or her chance of winning against a possibly very lucky opponent.

5 Results of the Main Experiment

We start by presenting our results on reporting behavior in the different treatments and then, we analyze the dynamics of reporting over time.

5.1 Cheating Behavior Across Treatments

Our first result is in line with the previous literature:

Result 1 (*Cheating*): In all treatments, a significant fraction of the participants over-reported the highest outcome (red) and under-reported the lowest outcome (blue) but not in every period, indicating partial cheating. They also under-reported the intermediate outcome (yellow) in the TR-CF, TR-FF and TR1-FF treatments.

Table 2: Reports by Treatment

Treatment	Blue outcome reported	Yellow outcome reported	Red outcome reported	Average value of reports		Full extent cheating
PR-FF	23.7%	30.9%	45.4%	1.22	(0.05)	4.8%
PR-CF	17.4%	31.2%	51.4%	1.34	(0.05)	1.4%
TR-FF	11.5%	21.8%	66.7%	1.55	(0.06)	27.3%
TR-CF	12.8%	23.7%	63.5%	1.51	(0.05)	10.2%
TR1-FF	20.1%	27.5%	52.4%	1.32	(0.06)	6.9%
TR1-CF	18.5%	30.5%	51%	1.32	(0.05)	9.7%
Total	16.6%	27.3%	56.1%	1.38	(0.02)	9.4%

Notes: PR stands for Piece Rate; TR for Tournament, and TR1 for One-sided Tournament. CF stands for Continuous Feedback and FF for Final Feedback. Standard errors are in parentheses. The first three columns report the percentages of blue (low outcome), yellow (medium outcome) and red (high outcome) colors reported over the 24 periods. The fourth column reports the average reported value over the 24 periods. A reported red outcome pays 2 points, a yellow one pays 1 point and a blue one pays 0 points. In the FF treatments, the average is computed at the participant level. In the CF treatments, it is computed at the pair level, except in the TR1-CF treatment where we only consider the active players. The last column reports the percentage of participants who reported the maximum number of points (48 points) in the task (full extent cheating). Passive participants who could not cheat in the TR1 treatments are excluded from the statistics reported in this table.

Support to Result 1. Table 2 reports, for each treatment, the frequency of each reported outcome, the average value of the reported outcomes and the frequency of subjects who cheated to the full extent (reporting the highest outcome in the 24 periods). If participants were reporting the outcomes of their rolls honestly, we should observe each possible outcome to be reported 33% of the time. Table 2 shows that participants reported a red outcome more than 33% of the time in all treatments (between 45.4% and 66.7%), while they reported a blue or a yellow outcome less than 33% of the time in all treatments (blue: between 11.6% and 23.7%; yellow: between 21.8% and 30.9%). Two-sided Wilcoxon signed rank tests[18] comparing the actual and expected reports for each outcome show

[18] All the non-parametric tests are two-sided, except if specified otherwise. One could argue that binomial tests would be more suited but in our CF treatments the independent observation is at the pair level, which

that participants significantly over-reported the red outcome and under-reported the blue outcome ($p < 0.001$ for each treatment).[19] They also significantly under-reported the yellow outcome but only in the TR-CF, TR-FF and TR1-FF treatments ($p < 0.001$ in both TR treatments and $p = 0.022$ in the TR1-FF treatment).

In all treatments, the average reported value (between 1.22 and 1.55 points) is higher than the expected value of 1 under the assumption of honest reporting. Wilcoxon signed rank tests comparing the average reported performance and the expected performance show that participants over-reported significantly in all treatments ($p < 0.001$ in all treatments). Most participants did not, however, cheat to the full extent. The percentage of participants who reported 48 points in total (*i.e.*, the maximum number of points that one could report over the 24 periods) is lower than 10%, except in the TR-CF and TR-FF treatments where it reaches 10.2 % and 27.3%, respectively.

To assess the occurrence of cheating throughout the experiment, we split the 24 periods in blocks of 4 periods.[20] We find that cheating occurred from the first block up to the last one in all treatments (Wilcoxon signed rank tests; $p < 0.05$), with only one case marginally significant at the 10% level (in the third block of the PR-FF treatment: $p = 0.086$). These findings support our Conjecture 1.

We now state our second result relative to treatment comparisons.

Result 2 (*Competition*): **a)** Participants over-reported more in the TR than in the PR treatments. **b)** They over-reported more in the TR than in the TR1 treatments where competition was less fierce since it was common knowledge that the opponent could not cheat. **c)** They did not over-report systematically more in the TR1 than in the PR treatments.

Support to Result 2. Table 3 reports the p-values of a Dunn's test.[21] It shows that participants over-reported significantly more in the two TR treatments than in any other treatment, as predicted by Conjecture 2a. This holds regardless of whether feedback was continuous or final (TR-CF *vs.* PR-FF/PR-CF: $p < 0.001$ and $p = 0.019$, respectively; TR-FF *vs.* PR-FF/PR-CF: $p < 0.001$ and $p = 0.009$, respectively). A statistical analysis pooling the CF and FF conditions together gives similar results: participants report significantly more points in the TR treatments compared to the PR ones (Wilcoxon rank sum test, $p < 0.001$). This is in line with previous research showing that competitive incentives increase dishonesty. The results of Table 3 also indicate that the values reported are

calls for an analysis at the pair level and prevents us to use binomial tests. In the FF treatments, binomial tests give similar results than Wilcoxon signed-rank tests.

[19] In this section, all statistical analyses are done at the pair level for the CF treatments and at the participant level for the FF treatment. Participants who were passive in the TR1 treatments are always excluded from the analysis.

[20] Figure C.1 of the Online Appendix C displays the average reported value across blocks of 4 periods, by treatment. Splitting the data into blocks of four periods avoids the fluctuations of a period-by-period analysis but it is arbitrary. As a robustness test, we conducted the same analysis with blocks of 2 and 6 periods, and the results are qualitatively the same.

[21] This test is based on a two-sided Kruskal-Wallis test and computes multiple pairwise comparisons on the values of the reports between all treatments.

significantly higher in the TR than in the TR1 treatments (TR-CF *vs.* TR1-FF/TR1-CF: $p = 0.015$ and $p = 0.014$, respectively; TR-FF *vs.* TR1-FF/TR1-CF: $p = 0.006$ for both comparisons). This is the case also when we compare all the TR treatments pooled together *vs.* all the TR1 treatments together (Two-sided Wilcoxon rank sum test, $p < 0.001$). This result indicates that individuals are more honest in a tournament when they face an opponent who cannot cheat, as predicted by Conjecture 2b. In competitive settings, honesty is conditional on one's opponent's honesty since it determines the fierceness of the competition. Believing (sincerely or strategically) that the competition may be unfair motivates one's own dishonesty; muting the possibility for the opponent to cheat strengthens moral conduct.

Table 3: Dunn's Test: Pairwise Comparisons of the Value of the Reports Between Treatments

Treatments	PR-FF	PR-CF	TR-FF	TR-CF	TR1-FF
PR-CF	**0.035**	-	-	-	-
TR-FF	**<0.001**	**0.008**	-	-	-
TR-CF	**<0.001**	**0.019**	0.383	-	-
TR1-FF	**0.068**	0.419	**0.006**	**0.015**	-
TR1-CF	**0.062**	0.424	**0.006**	**0.014**	0.494

Notes: This table reports p-values of a Dunn's test comparing the average reported value between all treatments. For the FF treatments the independent observations are at the participant level, while for the CF treatments the independent observations are at the pair level. Participants that could not misreport in the TR1 treatments are excluded from the analysis.

The tests reported in Table 3 also show that compared to piece-rate incentives, tournaments *per se* do not systematically induce more cheating. The values reported in the TR1 treatments are only marginally significantly different from those in PR-FF (TR1-CF *vs.* PR-FF: $p = 0.062$ and TR1-FF *vs.* PR-FF: $p = 0.068$), and not higher than those in PR-CF (TR1-CF *vs.* PR-FF: $p = 0.424$ and TR1-FF *vs.* PR-FF: $p = 0.420$). Furthermore, if we pool the treatments together, we find no significant difference between the reports in all the PR treatments and all the TR1 treatments (Wilcoxon rank sum test, $p = 0.282$). Dishonesty in a piece-rate scheme is similar to dishonesty in a tournament when the opponent cannot cheat. This is in line with Conjecture 2c.

We now state our third result which is related to the impact of the frequency of feedback on misreporting.

Result 3 (*Feedback*): Participants cheated significantly more when they received continuous feedback compared to final feedback, but only in the PR treatments.

Support to Result 3. In the PR treatments participants were more dishonest when they received continuous rather than final feedback about their counterpart's reports (PR-

FF *vs.* PR-CF: $p = 0.035$), which reveals the presence of peer effects. This result supports Conjecture 3a. By contrast, being informed continuously of the opponent's reports affected misreporting neither in the TR nor in the TR1 treatments (TR-FF *vs.* TR-CF: $p = 0.383$; TR1-FF *vs.* TR1-CF: $p = 0.494$).

For the TR1 treatments, this result, which is consistent with Conjecture 3b, is not surprising since continuous feedback adds little information value compared to final feedback.[22] This finding also suggests that observing repeatedly a counterpart that does not cheat does not make people more honest. This is consistent with the asymmetric peer effects found in Lefebvre et al. (2015) but it could also be due to the fact that truth-reporting is in our case compulsory and does not reveal anything of the opponent's intrinsic honesty.

As we explain in Section 4, we expect Result 3 to be driven by the extent to which subjects underestimate others' cheating behavior at the beginning of the task. In particular, we expect subjects in the PR treatments to underestimate the extent of cheating behavior, while we expect subjects in TR to display more accurate initial beliefs (which would explain that we find a significant effect of continuous *vs.* final feedback in the PR treatments and not in the TR treatments). However, since the beliefs elicited in the main experiment are influenced by the social information received throughout the 24 periods (recall that we elicited these beliefs at the end of the experiment), we cannot use them to test whether subjects underestimate more the extent of cheating behavior in PR-CF than TR-CF. We investigate this point with our follow-up experiment that we introduce in section 6.

To check the robustness of our results we ran a multinomial logit regression in which the dependent variable is the outcome reported by participant i in period t. The independent variables include treatment dummies with the PR-FF treatment taken as the reference, a time trend, a dummy variable for male participants, risk attitude and age, a dummy variable coding whether the participant is a student, and a fixed-effect for sessions ran during the same months.[23] The marginal effects are reported in Table B.3 of the Online Appendix B. Table 4 reports the results of multiple pairwise comparisons of the marginal effects between all treatments computed from the aforementioned multinomial logit model.

Table 4 shows that participants were more likely to report a red outcome and less likely to report a yellow or a blue outcomes in the TR treatments than in the PR treatments. Furthermore, reports differed between the TR-FF and the TR1 treatment but they were not sensitive to feedback in the TR and TR1 treatments. These results are consistent with the ones obtained from non-parametric tests. However, reports did not differ significantly between the TR-CF and the TR1 treatments, although the sign of the coefficients are in line with the non-parametric tests. Also, only reports of the blue outcome differed between the PR-FF and the PR-CF treatments. Overall, these results support Conjectures 2a, 2b,

[22] This is true for risk neutral individuals. But we obtain the same findings when controlling for risk attitudes in the regressions presented below.

[23] This variable controls for the fact that the experimental sessions have been run during three distinct months.

3a and 3b; they only partially support Conjectures 2c.

Table 4: Pairwise Comparisons of Treatments for All Outcomes

	PR-FF	PR-CF	TR-FF	TR-CF	TR1-FF
Blue outcome (0 points)					
PR-CF	**-0.048***	-	-	-	-
	(0.025)				
TR-FF	**-0.113*****	**-0.065*****	-	-	-
	(0.026)	(0.024)			
TR-CF	**-0.094*****	**-0.046***	0.019	-	-
	(0.026)	(0.024)	(0.025)		
TR1-FF	-0.044	0.005	**0.070****	0.051	-
	(0.032)	(0.031)	(0.031)	(0.031)	
TR1-CF	**-0.061****	-0.013	**0.052****	0.033	-0.018
	(0.027)	(0.026)	(0.026)	(0.026)	(0.029)
Yellow outcome (1 point)					
PR-CF	0.020	-	-	-	-
	(0.027)				
TR-FF	**-0.084*****	**-0.104*****	-	-	-
	(0.030)	(0.031)			
TR-CF	**-0.059****	**-0.079*****	0.025	-	-
	(0.029)	(0.029)	(0.032)		
TR1-FF	-0.044	**-0.064****	0.040	0.015	-
	(0.029)	(0.031)	(0.032)	(0.032)	
TR1-CF	-0.018	-0.038	**0.066***	0.041	0.026
	(0.032)	(0.034)	(0.034)	(0.035)	(0.032)
Red outcome (2 points)					
PR-CF	0.028	-	-	-	-
	(0.043)				
TR-FF	**0.198*****	**0.169*****	-	-	-
	(0.048)	(0.050)			
TR-CF	**0.153*****	**0.125*****	-0.044	-	-
	(0.046)	(0.048)	(0.052)		
TR1-FF	**0.088***	0.059	**-0.110****	-0.066	-
	(0.050)	(0.055)	(0.056)	(0.055)	
TR1-CF	**0.079***	0.051	**-0.118****	-0.074	-0.008
	(0.047)	(0.052)	(0.053)	(0.052)	(0.053)

Notes: The table reports pairwise comparisons of the marginal effects between all treatments for each outcome. Comparisons are computed from the model reported in Table B.3 of the Online Appendix B. Robust standard errors clustered at the pair (CF treatments) or individual level (FF treatments) are in parentheses. *** indicates significance at the 1% level, ** at the 5% level, and * at the 10% level.

Finally, in the Online Appendix we report further analyses exploring the heterogeneity of the effects of incentives and social information on behavior.[24]

[24] In particular, we investigated gender effects and found that males lied significantly more than females, which drove the high level of dishonesty in the TR-FF treatment. We also explored the impact of our treatment manipulations on happiness and nervousness. We found that participants were less happy and more nervous in the TR treatments, possibly because competition induces more stress.

5.2 Dynamics of Cheating Behavior

We now consider how cheating behavior evolved over time and introduce our fourth result.

Result 4 (*Escalation effects*): The likelihood to over-report the highest outcome increased over time in the TR-CF and the PR-CF treatments, while cheating was more stable in the other treatments.

Support to Result 4. Figure 2 reports the marginal effects of the period variable on the likelihood to report a blue, a yellow or a red outcome in each treatment. Estimates are computed from a multinomial logit regression model similar to the one reported in Table B.3 of the Online Appendix B, but with the inclusion of an interaction between each treatment dummy and the time trend variable. In both the PR-CF and the TR-CF treatments we observe a significant increase of reported red outcomes over time ($p = 0.012$ and $p = 0.001$, respectively). Reports of the yellow outcome decrease significantly over time only in the TR-CF treatment ($p = 0.001$). This reveals an escalation effect in the presence of social information.[25] Not surprisingly, in the TR1-FF treatment we do not observe any effect of time. The marginal effect of the period is significant for none of the outcomes ($p = 0.321$, 0.560 and 0.682 for the blue, yellow and red outcome, respectively).

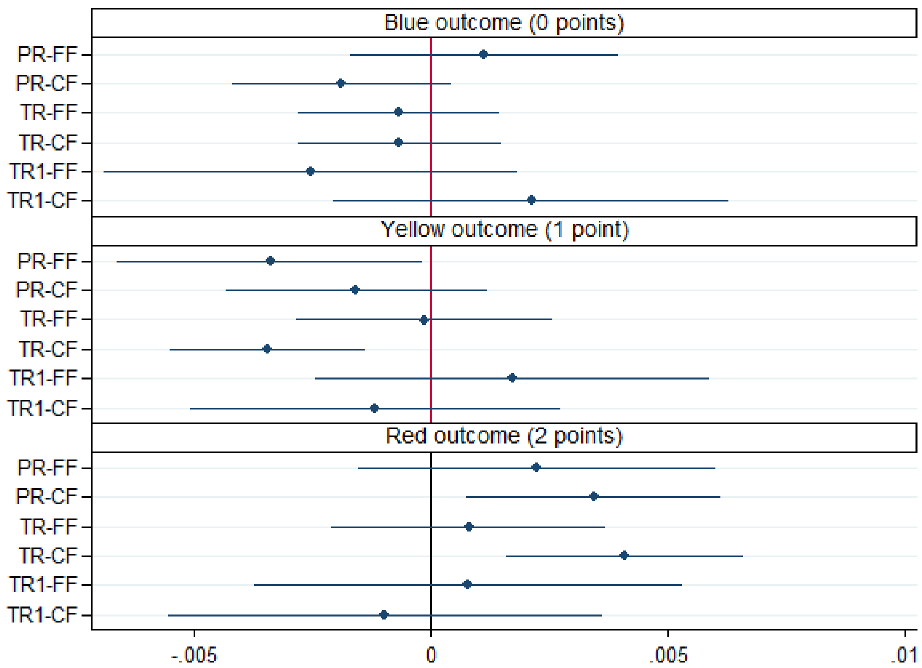


Figure 2: Marginal Effects of Time on the Probability to Report Each Outcome, by Treatment.
Notes: The figure displays the marginal effects of the period variable on the likelihood to report each possible outcome in each treatment. Estimates are computed from a multinomial logit regression model. The horizontal lines represent 95% confidence intervals.

[25] In the PR-FF treatment, participants were less likely to report a yellow outcome over time ($p = 0.040$), but not more likely to report a red outcome over time ($p = 0.239$). Thus, we cannot conclude that they lied more over time in this treatment.

To support the idea that the escalation of cheating in the PR-CF and TR-CF treatments results from social interactions,[26] we follow the strategy used in Falk and Ichino (2006) and replicated in Rosaz et al. (2016). We calculate the within-pair and between-pair standard deviations (S.D., hereafter) of the participants' total value of reports at the end of the 24 periods. A comparison of the within-pair and between-pair S.D. between the CF and the FF treatments allows us to determine whether social interactions have a significant impact in the PR-CF and TR-CF treatments. If social interactions generate imitation and/or rivalry, the within-pair S.D. should be lower and the between-pair S.D. should be higher in the CF treatments than in the corresponding FF treatments. Finally, the difference between the between-pair S.D. and the within-pair S.D. should be larger in the CF treatments than in the corresponding FF treatments.

The average within-pair S.D. is equal to 3.692 in the PR-CF treatment and 5.758 in the PR-FF treatment, suggesting a lower within-pair heterogeneity in the PR-CF than in the PR-FF treatment. A similar difference can be observed in the TR treatments where the within-pair S.D. is equal to 2.865 and 8.035 in the TR-CF and TR-FF treatments, respectively. Wilcoxon rank sum tests show that the difference in the within-pair S.D. between the PR-CF and PR-FF treatments is significant at the 10% level ($p = 0.053$), and the difference between the TR-CF and TR-FF treatments is significant at the 1% level ($p < 0.001$).

To calculate the standard deviations in the PR-FF and TR-FF treatments we used the actual pairs of participants. As a robustness test, we also ran a simulation-based approach.[27] The results of this simulation are displayed in Figure C.2 in Online Appendix C; they confirm the conclusions of the non-parametric tests.[28] The between-pair S.D. is significantly higher in the PR-CF and TR-CF treatments (6.696 and 8.122, respectively) than in the PR-FF and TR-FF treatments (5.103 and 6.269, respectively) (see Figure C.3 in Online Appendix C). We also compare the difference between the between-pair S.D. and the within-pair S.D. in both CF treatments with the difference computed for the real and simulated pairs in the corresponding FF treatments. Wilcoxon rank sum tests show that this difference is significantly larger for the pairs in the CF treatments than for the pairs in the FF treatments (PR treatments: $p < 0.001$; TR treatments: $p < 0.001$). Figure C.4 in Online Appendix C shows that this difference is larger in the CF treatments than in the simulated pairs of the FF treatments.

Overall, these analyses of within- and between-pair standard deviations support the

[26]One might argue that these results simply reflect a decrease of the moral cost of cheating over repetitions regardless of the social interactions (e.g., Garrett et al., 2016). However, such an effect should be independent of the presence of continuous feedback. But we do not find a significant effect of time in the FF treatments, at the exception of PR-FF discussed in footnote 25.

[27]We generated 30,000 configurations of all the hypothetical pairs formed with the participants of the PR-FF and TR-FF treatments that we compared with the average within-pair S.D. in the PR-CF and TR-CF treatments, respectively.

[28]Moreover, not surprisingly, we found no difference between the within- and between-pair S.D. in the TR1-FF and TR1-CF treatments.

existence of peer effects due to the dissemination of social information within pairs, which could explain the observed escalation of reports over time in the PR-CF and TR-CF treatments.

We can next test whether the effect of feedback is larger in the TR-CF than in the PR-CF treatments (Conjecture 3c). This brings us to the following result.

Result 5 (*Treatment Differences in Escalation*): The contagiousness of dishonesty was not significantly larger in tournaments than in the piece-rate pay scheme.

Support to Result 5. If the effect of feedback was larger in TR-CF than in PR-CF, we should have seen a lower within-pair standard deviation of the participants' total value of reports in TR-CF compared to PR-CF. However, a Wilcoxon rank sum test rejects the hypothesis that the within-pair average standard deviation was different between the two treatments ($p = 0.217$). We thus reject Conjecture 3c.

6 A Follow-up Experiment to Identify the Role of Beliefs

6.1 Design and Procedures

In the main experiment we found that participants cheated more when they received a continuous rather than a final feedback under piece-rate incentives, but not in tournaments. We argue that this is because uninformed participants underestimated more the extent of cheating behavior under a piece-rate pay scheme than in a tournament where the anticipation of the opponent's behavior has strategic value. To test this interpretation, we ran a second experiment with 102 new participants from the same pool of subjects. The experiment comprised of two conditions: the belief-PR-CF and belief-TR-CF treatments (bPR-CF and bTR-CF, hereafter) with 46 and 56 subjects, respectively.

These two treatments were equivalent to the TR-CF and PR-CF treatments of the main experiment, except that we elicited the participants' beliefs about the prevalence of cheating behavior at the beginning of each period. We asked the participants to indicate how many participants out of 24 (selected at random from past sessions) reported a red, a yellow and a blue outcome in a given period. At the end of the experiment, the computer randomly drew one period and one color for payment. Participants received €1.5 for a correct prediction, €1 if their prediction was incorrect by $+/-$ 1, €0.5 if it was incorrect by $+/-$ 2, and €0 otherwise. All the other procedures were equivalent to the main experiment.

6.2 Results

As in the main experiment, we find that participants over-reported the die rolls.[29] On average, participants over-reported more in bTR-CF than bPR-CF. However, the difference is not large enough to achieve statistical significance (Wilcoxon rank sum test on the average report, $p = 0.348$). A possible explanation is that the belief elicitation induced participants to excessively focus on the behavior of the others and on interpersonal comparisons. This could have accrued the desire to outperform the counterpart for purely intrinsic reasons—such as raising one's own status or dominance (Charness et al., 2014)—and amplified cheating. Indeed, comparing behavior in the PR-CF and TR-CF with behavior in the bPR-CF and bTR-CF shows that participants reported significantly less reds (64% *vs.* 68%, $p = 0.091$) and more yellows (27% *vs.* 22%, $p = 0.027$) in the main experiment compared to the follow-up experiment.[30] We also replicate the dynamics that we observe in the main treatment under continuous feedback.[31]

Turning to the beliefs, we can measure the extent to which participants underestimated the amount of cheating at the beginning of the task when feedback was not provided yet. In particular, we can compute to which extent the beliefs differed from the actual distribution of reports for period 1. We find that participants underestimated the average report by 13.03% (2.62%) in bPR-CF (bTR-CF). The difference is statistically significant (Wilcoxon rank sum test, $p = 0.055$). More precisely, participants in bPR-CF underestimated the proportion of reds in period 1 by 19.39%, while participants in bTR-CF overestimated it by 3.69% ($p = 0.011$). Furthermore, participants in bPR-CF (bTR-CF) overestimated (underestimated) the proportion of yellows by 17.04% (20.83%) (bPR-CF *vs.* bTR-CF, $p < 0.001$). They also tended to overestimate more the proportion of blues (by 35.19%) compared to participants in bTR-CF (by 23.26%), but the difference is not significant ($p = 0.571$). These findings indicate that before getting any feedback participants who were paid a piece rate tended to underestimate more the extent of cheating compared to participants who competed in a tournament. This supports our explanation regarding Result 3: compared to a final feedback, the effect of a continuous feedback on dishonesty is positive and significant only if subjects initially underestimated the extent of cheating.

In the Online Appendix D.4, we also look at how participants updated their beliefs over

[29]Pooling all periods together, the average report (1.46 in bPR-CF, 1.55 in bTR-CF) is significantly higher than the expected value of 1 (Two-sided Wilcoxon signed rank tests, $p < 0.001$ in both treatments). Also, participants significantly reported a red outcome more than 33% of the time (60.7% in bPR-CF, $p < 0.001$; 67.3% in bTR-CF, $p < 0.001$), and a blue or a yellow outcome less than 33% of the time (blue: 14.5% in bPR-CF, $p < 0.001$, and 12.5% in bTR-CF, $p < 0.001$; yellow: 24.8% in bPR-CF, $p = 0.006$, and 20.2% in bTR-CF, $p < 0.001$). The percentage of participants who cheated to the full extent (by reporting 48 points in total) is 15.22% in bPR-CF, and 19.64% in bTR-CF.

[30]The difference is larger under piece-rate incentives (red: 51% *vs.* 67%, $p = 0.067$; yellow: 31% *vs.* 25%, $p = 0.026$), possibly because of ceiling effects in tournaments.

[31]The likelihood to over-report the highest outcome increased over time in both the bTR-CF and the bPR-CF treatments (in line with Result 4). In addition, the within-pair average standard deviation was similar between the two treatments, suggesting that the contagiousness of dishonesty was analogous in the two treatments (in line with Result 5) (see the details of this analysis in Online Appendix D.3).

time once they began to receive information about the behavior of their counterpart, and whether these beliefs had an impact on behavior. We find that that participants expected a higher (lower) average report in the group of 24 if they had previously under-estimated (over-estimated) the norm. In addition, participants tended to over-report more, the more they thought the others engaged in this behavior. These results reveal interesting patterns but they should be interpreted with caution due to potential endogeneity issues.

7 Discussion and Conclusion

In this study, we investigated how feedback about others' reported outcomes influences individuals' cheating behavior in both competitive and non-competitive dynamic settings. We also tested whether the level of cheating in competitive settings results mainly from the existence of competition or from the anticipation of an unfair competition due to the opponent's dishonest reporting.

A first important result is that competitive incentives make individuals more likely to cheat compared to an individual piece-rate pay scheme. This replicates in a new dynamic setting previous findings obtained in studies using real-effort tasks (*e.g.,* Schwieren and Weichselbaumer, 2010; Rigdon and D'Esterre, 2015; Faravelli et al., 2015; Belot and Schröder, 2013). We also found that individuals are less likely to cheat in tournaments when it is common knowledge that their opponent cannot misreport. Cheating differs only weakly in this context compared to a piece-rate scheme. This suggests that the higher level of dishonesty when both competitors can lie is mostly driven by a fear of losing a potentially unfair and fiercer competition.

We also found that individuals, when receiving a continuous feedback on their counterpart's reports, became more dishonest over time. This is in line with the idea of a behavioral contagiousness of dishonesty (e.g., Gino et al., 2009; Robert and Arnab, 2013; Lauer and Untertrifaller, 2019; Charroin et al., 2021). The contagiousness of dishonesty was not significantly larger in tournaments than in the piece-rate pay scheme. Rivalry over the prize in tournaments did not generate larger peer effects than under the piece rate scheme. Under the piece-rate pay scheme the peer effects may be driven by pure conformity, inequality aversion and rivalry in terms of status attached to relative ranking.

Finally, we found that providing continuous feedback, instead of a final feedback, on the counterpart's reports increased cheating at the aggregate level in piece-rate settings but not in tournament settings. We argued that this is because subjects had different initial expectations about the cheating behavior of the others, and we confirmed this intuition in a follow-up experiment.

Our study provides new evidence on the role that feedback on relative performance plays on cheating behavior under competitive and non-competitive payment schemes. It shows that competitive incentives, compared to non-competitive incentives, encourage dishonesty even when there is uncertainty about the opponents' degree of dishonesty and thus, about the actual fierceness of the competition. It also stresses out the importance of people's expectations in driving cheating behavior. In non-competitive settings, providing relative feedback generate peer effects with deleterious effects in terms of cheating. Future research could investigate the exact nature of the effect of social information in non-competitive settings, in order to disentangle the role of normative imitation, inequality aversion and rivalry. A related research question is whether the fierceness of the competition is something that is only relevant when people compete for money or whether it also drives dishonesty when people compete for status or rank. One could, for example, compare two PR conditions in which either both agents or only one can cheat, and where symbolic rewards are assigned to the best performer. Another useful extension of our work would be to explore how precisely social information changes the moral cost of lying, in particular through the perception of what is socially acceptable.

Overall, although one needs to remain cautious before extrapolating evidence from the lab, our findings suggest that organizations should introduce competitive incentive schemes and continuous feedback policies with care when agents have some discretion in reporting, as both tend to increase the contagiousness of dishonesty. Adopting a too strict market culture that emphasizes competitiveness and continuous peer monitoring may have deleterious effects for the ethical conduct of employees.

References

Abeler, J., D. Nosenzo, and C. Raymond (2019). Preferences for Truth-Telling. *Econometrica 87*(4), 1115–1153.

Aydogan, G., A. Jobst, K. D'Ardenne, N. Müller, and M. G. Kocher (2017). The Detrimental Effects of Oxytocin-Induced Conformity on Dishonesty in Competition. *Psychological Science 28*(6), 751–759.

Bäker, A. and M. Mechtel (2019). the Impact of Peer Presence on Cheating. *Economic Inquiry 57*(2), 792–812.

Bandiera, O., I. Barankay, and I. Rasul (2010). Social incentives in the workplace. *Review of Economic Studies 77*(2), 417–458.

Banerjee, R., N. D. Gupta, and M. C. Villeval (2018). The spillover effects of affirmative action on competitiveness and unethical behavior. *European Economic Review 101*, 567–604.

Barankay, I. (2019). Rank incentives: Evidence from a randomized workplace experiment. *Mimeo*.

Belot, M. and M. Schröder (2013). Sloppy work, lies and theft: A novel experimental design to study counterproductive behaviour. *Journal of Economic Behavior and Organization 93*(C), 233–238.

Benistant, J. and M. C. Villeval (2019). Unethical behavior and group identity in contests. *Journal of Economic Psychology 72*, 128–155.

Beugnot, J., B. Fortin, G. Lacroix, and M. C. Villeval (2019). Gender and Peer Effects on Performance in Social Networks. *European Economic Review 113*, 207–224.

Bock, O., I. Baetge, and A. Nicklisch (2014). Hroot: Hamburg Registration and Organization Online Tool. *European Economic Review 71*, 117–120.

Bortolotti, S., I. Soraperra, M. Sutter, and C. Zoller (2017). Too Lucky to be True Fairness views under the shadow of cheating. *CESifo Working Paper* (6563), 1–55.

Brown, J. and D. Minor (2016). Misconduct in financial services: Differences across organizations. *Harvard Business School Working Paper 16-022*.

Bull, C., A. Schotter, and K. Weigelt (1987). Tournaments and piece rates: An experimental study. *Journal of Political Economy 95*(1), 1–33.

Buser, T. and A. Dreber (2016). The Flipside of Comparative Payment Schemes. *Management Science 62*(9), 2626–2638.

Buser, T., A. Dreber, and J. Mollerstrom (2017). The impact of stress on tournament entry. *Experimental Economics 20*(2), 506–530.

Cadsby, C. B., F. Song, and F. Tapon (2010). Are you paying your employees to cheat? an experimental investigation. *The BE Journal of Economic Analysis & Policy 10*(1).

Carpenter, J., P. H. Matthews, and J. Schirm (2010). Tournaments and Office Politics : Evidence from a Real Effort Experiment. *American Economic Review 100*(1), 504–517.

Cartwright, E. and M. L. C. Menezes (2014). Cheating to win: Dishonesty and the intensity of competition. *Economics Letters 122*(1), 55–58.

Cason, T. N., W. A. Masters, and R. M. Sheremeta (2018). Winner-take-all and proportional-prize contests: Theory and experimental results. *Journal of Economic Behavior and Organization 0*, 1–14.

Charness, G., D. Masclet, and M. C. Villeval (2014). The Dark Side of Competition for Status. *Management Science 60*(1), 38–55.

Charroin, L., B. Fortin, and M. C. Villeval (2021). Homophily, peer effects and dishonesty. In *GATE Working Paper*. GATE.

Cohn, A., E. Fehr, and M. A. Maréchal (2014). Business culture and dishonesty in the banking industry. *Nature 516*(7529), 86–89.

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

Corno, L. (2017). Homelessness and crime: do your friends matter? *The Economic Journal 127*(602), 959–995.

Croson, R. T. (2000). Thinking like a game theorist: factors affecting the frequency of equilibrium play. *Journal of Economic Behavior & Organization 41*(3), 299–314.

Dai, Z., F. Galeotti, and M. C. Villeval (2018). Cheating in the Lab Predicts Fraud in the Field: An Experiment in Public Transportation. *Management Science 64*(3), 983–1476.

Delfgaauw, J., R. Dur, A. Non, and W. Verbeke (2014). Dynamic incentive effects of relative performance pay: A field experiment. *Labour Economics 28*, 1–13.

Delfgaauw, J., R. Dur, J. Sol, and W. Verbeke (2013). Tournament incentives in the field: Gender differences in the workplace. *Journal of Labor Economics 31*(2), 305–326.

Diekmann, A., W. Przepiorka, and H. Rauhut (2015). Lifting the veil of ignorance: An experiment on the contagiousness of norm violations. *Rationality and Society 27*(3), 309–333.

Dimant, E. (2018). Contagion of Pro- and Anti-Social Behavior Among Peers and the Role of Social Proximity. *Mimeo*.

Dufwenberg, M. and M. A. Dufwenberg (2018). Lies in disguise – A theoretical analysis of cheating. *Journal of Economic Theory 175*(175), 248–264.

Díaz, C. and E. Patacchini (2020). Parents, neighbors and youth crime. *IZA Discussion Paper* (13906).

d'Adda, G., D. Darai, N. Pavanini, and R. A. Weber (2017, 02). Do Leaders Affect Ethical Conduct? *Journal of the European Economic Association 15*(6), 1177–1213.

Egan, M., G. Matvos, and A. Seru (2019). The market for financial adviser misconduct. *Journal of Political Economy 127*(1), 233–295.

Eriksson, T., A. Poulsen, and M. C. Villeval (2009). Feedback and incentives: Experimental evidence. *Labour Economics 16*(6), 679–688.

Falk, A. and A. Ichino (2006). Clean Evidence on Peer Effects. *Journal of Labor Economics 24*(1), 39–57.

Falk, A. and N. Szech (2013). Morals and Markets. *Science 340*(6133), 707–712.

Faravelli, M., L. Friesen, and L. Gangadharan (2015). Selection, tournaments, and dishonesty. *Journal of Economic Behavior and Organization 110*, 160–175.

Feltovich, N. (2019). The interaction between competition and unethical behaviour: experimental evidence. *Experimental Economics 22*, 101–130.

Fershtman, C. and U. Gneezy (2011). The tradeoff between performance and quitting in high power tournaments. *Journal of the European Economic Association 9*(2), 318–336.

Fischbacher, U. (2007). Z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics 10*(2), 171–178.

Fischbacher, U. and F. Föllmi-Heusi (2013). Lies in Disguise – An Experimental Study on Cheating. *Journal of the European Economic Association 11*(40), 525–547.

Fitch, B. D. (2013). *Law enforcement ethics*. Sage.

Fortin, B., G. Lacroix, and M. C. Villeval (2007). Tax evasion and social interactions. *Journal of Public Economics 91*(11-12), 2089–2112.

Fosgaard, T. R., L. G. Hansen, and M. Piovesan (2013). Separating Will from Grace: An experiment on conformity and awareness in cheating. *Journal of Economic Behavior & Organization 93*, 279–284.

Gächter, S. and E. Renner (2010). The effects of (incentivized) belief elicitation in public goods experiments. *Experimental Economics 13*(3), 364–377.

Gangadharan, L., P. J. Grossman, and J. Vecci (2020). Antisocial behavior in the workplace. In K. Zimmermann (Ed.), *Handbook of Labor, Human Resources and Population Economics*. Springer.

Garrett, N., S. C. Lazzaro, D. Ariely, and T. Sharot (2016). The brain adapts to dishonesty. *Nature Neuroscience 19*(October), 1–6.

Gill, D. and V. Prowse (2012). A structural analysis of disappointment aversion in a real effort competition. *American Economic Review 102*(1), 469–503.

Gino, F., S. Ayal, and D. Ariely (2009). Contagion and differentiation in unethical behavior: the effect of one bad apple in the barrel. *Psychological Science 20*(3), 393–398.

Glaeser, E. L., B. Sacerdote, and J. A. Scheinkman (1996). Crime and social interactions. *The Quarterly Journal of Economics 111*(2), 507–548.

Gneezy, U., A. Kajackaite, and J. Sobel (2018). Lying Aversion and the Size of the Lie. *American Economic Review 108*(2), 419–153.

Gneezy, U. and J. Potters (1997). An experiment on risk taking and evaluation periods. *The Quarterly Journal of Economics 112*(2), 631–645.

Harbring, C. and B. Irlenbusch (2011). Sabotage in Tournaments: Evidence from a Laboratory Experiment. *Management Science 57*(4), 611–627.

Harbring, C., B. Irlenbusch, M. Kräkel, and R. Selten (2007). Sabotage in Corporate Contests – An Experimental Analysis. *International Journal of the Economics of Business 14*(3), 367–392.

Herbst, D. and A. Mas (2015). Peer effects on worker output in the laboratory generalize to the field. *Science 350*(6260).

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

Khalmetski, K. and D. Sliwka (2019). Disguising Lies – Image Concerns and Partial Lying in Cheating Games. *American Economic Journal: Microeconomics 11*(4), 79–110.

Klimm, F. (2019). Suspicious success – Cheating, inequality acceptance, and political preferences. *European Economic Review 117*, 36–55.

Konrad, K. A. (2000). Sabotage in Rent-seeking Contests. *Journal of Law, Economics & Organization 16*(1), 155–165.

Kuhnen, C. M. and A. Tymula (2012). Feedback, Self-Esteem and Performance in Organizations. *Management Science 58*(1), 94–113.

Lauer, T. and A. Untertrifaller (2019). Conditional dishonesty. *Mimeo*.

Lazear, E. P. (1989). Pay equality and Industrial Politics. *Journal of Political Economy 97*(3), 561–580.

Lazear, E. P. (2018). Compensation and incentives in the workplace. *Journal of Economic Perspectives 32*(3), 195–214.

Lefebvre, M., P. Pestieau, A. Riedl, and M. C. Villeval (2015). Tax evasion and social information: an experiment in Belgium, France, and the Netherlands. *International Tax and Public Finance 22*(3), 401–425.

Levashina, J. and M. A. Campion (2007). Measuring Faking in the Employment Interview: Development and Validation of an Interview Faking Behavior Scale. *Journal of Applied Psychology 92*(6), 1638–1656.

Ludwig, S. and G. K. Lünser (2012). Observing your competitor - The role of effort information in two-stage tournaments. *Journal of Economic Psychology 33*(1), 166–182.

Mas, A. and E. Moretti (2009). Peers at Work. *American Economic Review 99*(1), 112–145.

Mazar, N., O. Amir, and D. Ariely (2008). The Dishonesty of Honest People: A Theory of Self-Concept Maintenance. *Journal of Marketing Research 45*(6), 633–644.

Parco, J. E., A. Rapoport, and W. Amaldoss (2005). Two-stage contests with budget constraints: An experimental study. *Journal of Mathematical Psychology 49*(4), 320–338.

Patacchini, E. and Y. Zenou (2008). The strength of weak ties in crime. *European Economic Review 52*(2), 209–236.

Patacchini, E. and Y. Zenou (2009). Juvenile delinquency and conformism. *The Journal of Law, Economics, & Organization 28*(1), 1–31.

Piskorski, T., A. Seru, and J. Witkin (2015). Asset quality misrepresentation by financial intermediaries: Evidence from the rmbs market. *Journal of Finance 70*(6), 2635–2678.

Poujol, J. F. (2009). Management of sales advisers and service climate: an experiment. *Journal of Service Management 20*(3), 274–289.

Prendergast, C. (1999). The provision of incentives in firms. *Journal of Economic Literature 37*(1), 7–63.

Rabanal, J. P. and O. A. Rud (2018). Does competition affect truth telling? An experiment with rating agencies. *Review of Finance 22*(4), 1581–1604.

Rauhut, H. (2013). Beliefs about lying and spreading of dishonesty: Undetected lies and their constructive and destructive social dynamics in dice experiments. *PLoS ONE 8*(11).

Rigdon, M. L. and A. D'Esterre (2015). The Effects of Competition on the Nature of Cheating Behavior. *Southern Economic Journal 81*(4), 1012–1024.

Robert, I. and M. Arnab (2013). Is dishonesty contagious? *Economic Inquiry 51*(1), 722–734.

Rosaz, J., R. Slonim, and M. C. Villeval (2016). Quitting and peer effects at work. *Labour Economics 39*, 55–67.

Schurr, A. and I. Ritov (2016). Winning a competition predicts dishonest behavior. *Proceedings of the National Academy of Sciences 113*(7), 201515102.

Schwieren, C. and D. Weichselbaumer (2010). Does competition enhance performance or cheating? A laboratory experiment. *Journal of Economic Psychology 31*(3), 241–253.

Shalvi, S., J. Dana, M. J. Handgraaf, and C. K. De Dreu (2011). Justified ethicality: Observing desired counterfactuals modifies ethical perceptions and behavior. *Organizational Behavior and Human Decision Processes 115*(2), 181–190.

Soraperra, I., O. Weisel, R. Zultan, S. Kochavi, M. Leib, H. Shalev, and S. Shalvi (2017). The bad consequences of teamwork. *Economics Letters 160*, 12–15.

Van Dijk, F., J. Sonnemans, and F. Van Winden (2001). Incentive systems in a real effort experiment. *European Economic Review 45*(2), 187–214.

Villeval, M. C. (2020). Performance feedback and peer effects. In K. Zimmermann (Ed.), *Handbook of Labor, Human Resources and Population Economics*. Springer.

Weiss, B. and R. S. Feldman (2006). Looking good and lying to do it: Deception as an impression management strategy in job interviews. *Journal of Applied Social Psychology 36*(4), 1070–1086.

A Instructions

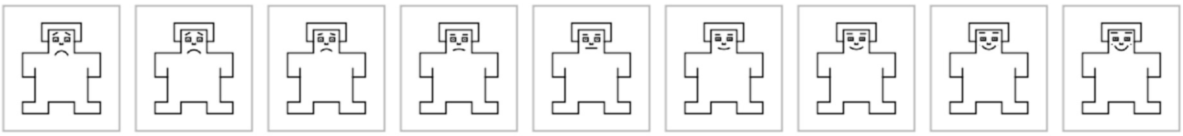
Instructions have been translated from French. All the instructions were displayed on the computer screen.

A.1 Two-Sided treatments

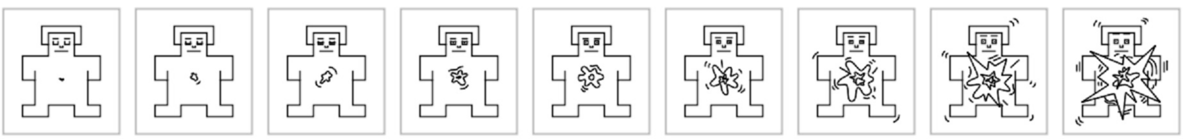
Welcome screen Hello. Thank you for participating in this experiment on decision making. Please turn off your phone. It is forbidden to talk to the others participants for the whole duration of the experiment. All your decisions are anonymous. If you have any questions during the experiment, please press the red button and an assistant will come to answer your questions in private.

The experiment is divided into two independent parts. In addition, you will be asked some questions at the end of the experiment. At the beginning of each part, you will receive the corresponding instructions. At the end of the experiment, you will receive the earnings for the first and second part as well as a show-up fee of €5. You will be paid in private in a separate room by someone who does not know the content of the experiment in order to maintain confidentiality.

Questionnaire Before receiving the instructions of the first part, please answer sincerely the following two questions. We remind you that all your answers are anonymous.
Please select the picture that best describe how you feel **right now**. For example, if you feel completely happy or contented, you could indicate it by selecting the picture at the rightmost part of the screen. On the other hand, if you feel completely unhappy or unsatisfied, you could indicate it by selecting the picture at the leftmost part of the screen. The figures also allow you to describe intermediate feelings of happiness by selecting any of the other pictures.



Again, please select the figure that best describes how you feel **right now**. For example, if you feel completely stimulated, excited, or aroused, you could indicate that by selecting the rightmost picture. On the other hand, if you feel completely relaxed, calm, or unaroused, you could indicate that by selecting the leftmost picture. The figures also allow you to describe intermediate feelings of arousal by selecting any of the other pictures.



Part 1

In this part, you will take one decision which will determine your earnings of this part. You are endowed with €150 cents (€1.5) and you have to decide which part of this amount (between 0 and €150 cents, included) you wish to invest in a lottery. In this lottery, the amount invested will be either multiplied by 2.5 or entirely lost. The probability of each of these events is 50%. Your final payoff will be equal to the following sum: amount not invested in the lottery + outcome of the lottery.

Part 2: Main Task

The task

At the beginning of this part, you will be paired with another participant. You will stay paired with the same participant throughout this part.
This part is composed by 24 periods. {*Two-sided conditions*: In each period, you are asked to roll the die that is contained in the cup on your desk without taking off the lid from the cup. Your task is to report the color which appears on the top of the die when looking through the lid. If you wish, you can roll the die more than once in each period. However, you have to report *only the outcome of your first roll.*

The points obtained in each period depend on the outcome of the roll as indicated below:}

{*One-sided conditions*: Each one of you has a cup on the desk but only your or your counterpart's cup contains a die. In each period, the one with the die is asked to roll it in the cup without taking off the lid from the cup. The task of this person is to report the color which appears on the top of the die when looking through the lid.
The task of the person without the die is to roll an electronic die by clicking on the button "roll the die" on the computer screen. Both for the electronic die and the die in the cup, there are three possible outcomes which have the same chance to be drawn.

If you wish, you can roll the die more than once in each period. However, if you have the die in the cup, you have to report *only the outcome of your first roll.* If you have the electronic die, the computer automatically records the outcome of the first roll.

The number of points obtained in each period depends on the outcome of the roll (electronic or not) as indicated below:

Die outcome	Points
Red	2 points
Yellow	1 point
Blue	0 points

{*Final Feedback condition*}

At the beginning of each period, you are reminded the number of points obtained in the previous periods as well as the total number of points accumulated so far.
At the end of the period 24 (and only at that moment), you will be informed of the number of points obtained by your counterpart in each period as well as the total number of points that she or he has accumulated.

{*Continuous Feedback condition*}

At the beginning of each period, you are reminded the number of points obtained in the previous periods as well as the total number of points accumulated so far.
At the end of each period and at the end of period 24, you will be informed of the number of points obtained by your counterpart in each period as well as the total number of points that she or he has accumulated until then.

Earnings

At the end of the 24 periods, you will have accumulated a certain number of points.

{Piece-rate condition} Your earnings will be determined by the number of points that you have accumulated. Your points will be converted to euros according to the following conversion rate: 1 point = €18 cents, and they will be paid to you in cash.

{Tournament condition} If you have accumulated more points than your counterpart, your earnings will be determined based on the number of points that you have accumulated. Your points will be converted to euros according to the following conversion rate: 1 point = €36 cents, and they will be paid to you in cash.

If you have accumulated less points than your counterpart, your earnings will be equal to €0.

In case of a tie in the number of points accumulated at the end of period 24 between you and your counterpart, the program will randomly select who will receive €0 and who will receive his or her accumulated points converted to Euros.

Predictions

Before proceeding to the end of the experiment, you will have the opportunity to earn an additional €1.5. You will have to predict the colors that 24 other participants (excluding your counterpart)*{One-sided conditions*: 24 other participants who received a die in the cup (excluding your counterpart or yourself)*}* have reported in the first, thirteenth and last period, respectively. These 24 participants are chosen at random by the computer from the current session or previous sessions of this experiment. Here is the procedure:

For each of these periods (1, 13 and 24), you will have to indicate how many participants, among these 24, reported the color red (which gives 2 points), the color yellow (which gives 1 point) and the color blue (which gives 0 points).

At the end of the session, the program will randomly select one of these three periods and one of these three colors. You will be paid for your prediction regarding the selected period and color.

- If your prediction is correct (meaning that your predicted number of participants for this period and color is equal to the true number), you earn €1.5;
- If your prediction is incorrect by one participant (meaning that your predicted number of participants for this period and color differs from the true number by one participant), your earn €1;
- If your prediction is incorrect by two participants (meaning that your predicted number of participants for this period and color differs from the true number by two participants), your earn €0.5;
- If your prediction is incorrect by three participants or more (meaning that your predicted number of participants for this period and color differs from the true number by three participants or more), your earn €0.

Example: Imagine that the program randomly select period 13 and the color yellow. This implies that you will be paid for your prediction on the number of participants who reported the color yellow in period 13. Imagine that your prediction was 12 (out of 24 participants) while the correct number is 14. In that case, your prediction is incorrect by two participants and your earnings are equal to €0.5.

B Additional tables

Table B.1: Demographics of the Participants, by Treatment (Main Experiment)

Treatment	Average age (S.D.)	Males (%)	Nb. subjects
Piece rate - Final Feedback (PR-FF)	24.78 (11.24)	50.0%	42
Piece rate - Continuous Feedback (PR-CF)	23.07 (6.55)	52.8%	72
Tournament - Final Feedback (TR-FF)	21.41 (2.07)	52.3%	44
Tournament - Continuous Feedback (TR-CF)	21.50 (2.28)	47.4%	78
One-Sided Tournament - Final Feedback (TR1-FF)	22.67 (6.03)	44.8%	58
One-Sided Tournament - Continuous Feedback (TR1-CF)	26.68 (12.47)	45.2%	62
Total	23.29 (7.80)	48.6%	356

Table B.2: Demographics of the Participants, by Treatment (Follow-up Experiment)

Treatment	Average age (S.D.)	Males (%)	Nb. subjects
Piece rate with beliefs - Continuous Feedback (bPR-CF)	26.59 (10.15)	56.5%	46
Tournament with beliefs - Continuous Feedback (bTR-CF)	27.75 (10.14)	64.3%	56
Total	27.23 (10.12)	60.78%	102

Table B.3: Determinants of the Die Roll Report

	Blue (0 point)	Yellow (1 point)	Red (2 points)
PR-FF	*Ref*	*Ref*	*Ref*
PR-CF	-0.048*	0.020	0.028
	(0.025)	(0.027)	(0.043)
TR-FF	-0.113***	-0.084***	0.198***
	(0.026)	(0.030)	(0.048)
TR-CF	-0.094***	-0.059**	0.153***
	(0.026)	(0.029)	(0.046)
TR1-FF	-0.044	-0.044	0.088*
	(0.032)	(0.029)	(0.050)
TR1-CF	-0.061**	-0.018	0.079*
	(0.027)	(0.032)	(0.047)
Period	-0.001	-0.002***	0.002***
	(0.001)	(0.001)	(0.001)
Other controls		Yes	
Observations		7104	
Number of clusters		221	
Pseudo-R^2		0.019	
p>chi2		<0.001	

Notes: The table reports marginal effects of a multinomial logit regression. Standard errors (in parenthesis) are clustered at the subject (pair) level in the FF (CF) treatments. The baseline category is the PR-FF treatment. Other independent variables include risk attitude, age, male gender, student status, and a control for the month in which the data were collected. *** indicates significance at the 1% level, ** at the 5% level, and * at the 10% level.

C Additional Figures

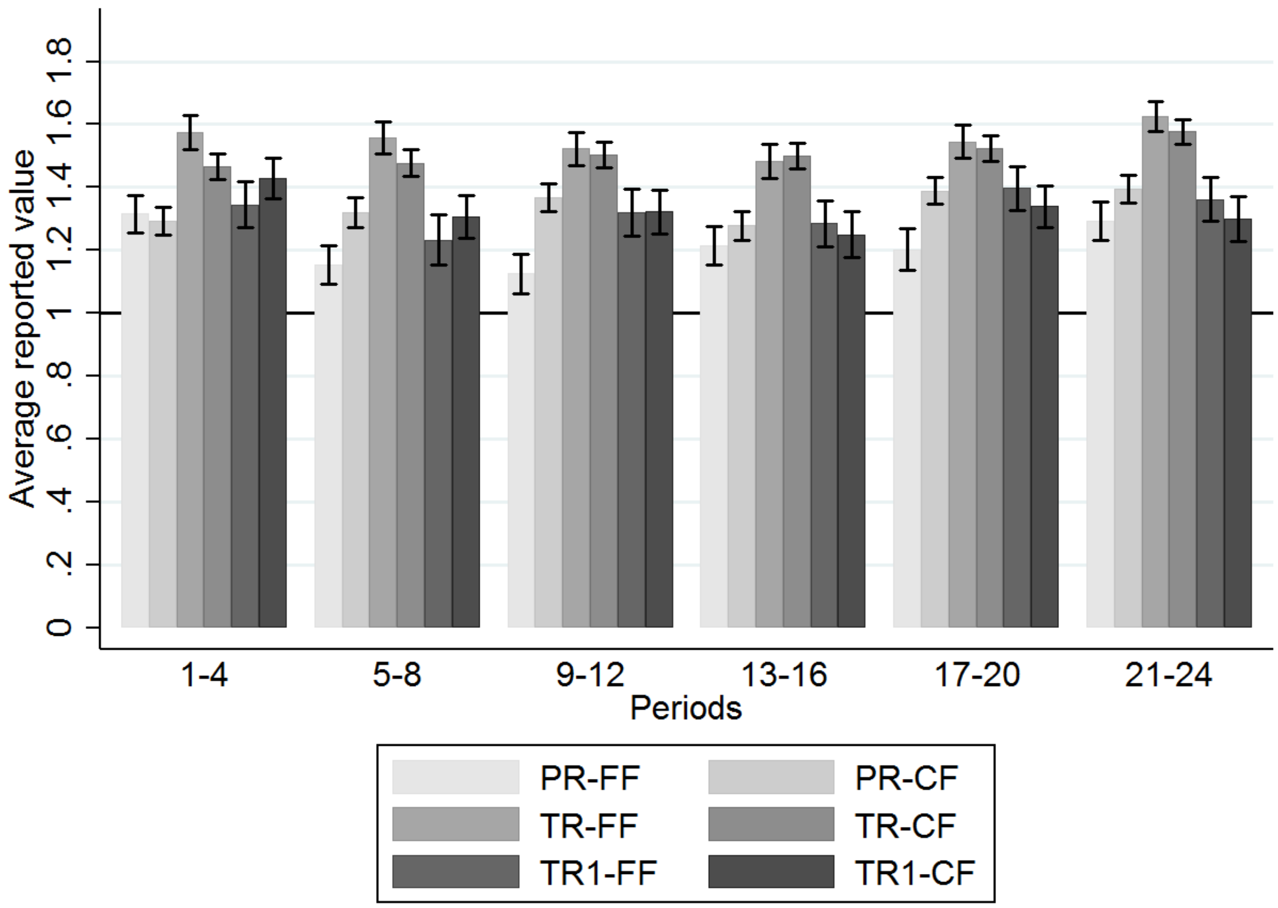


Figure C.1: Average Reported Value, by Treatment and Blocks of Four Periods *Notes*: The error bars represent standard errors of the mean. The horizontal black line corresponds to the expected value if participants were reporting honestly (average reported value = 1).

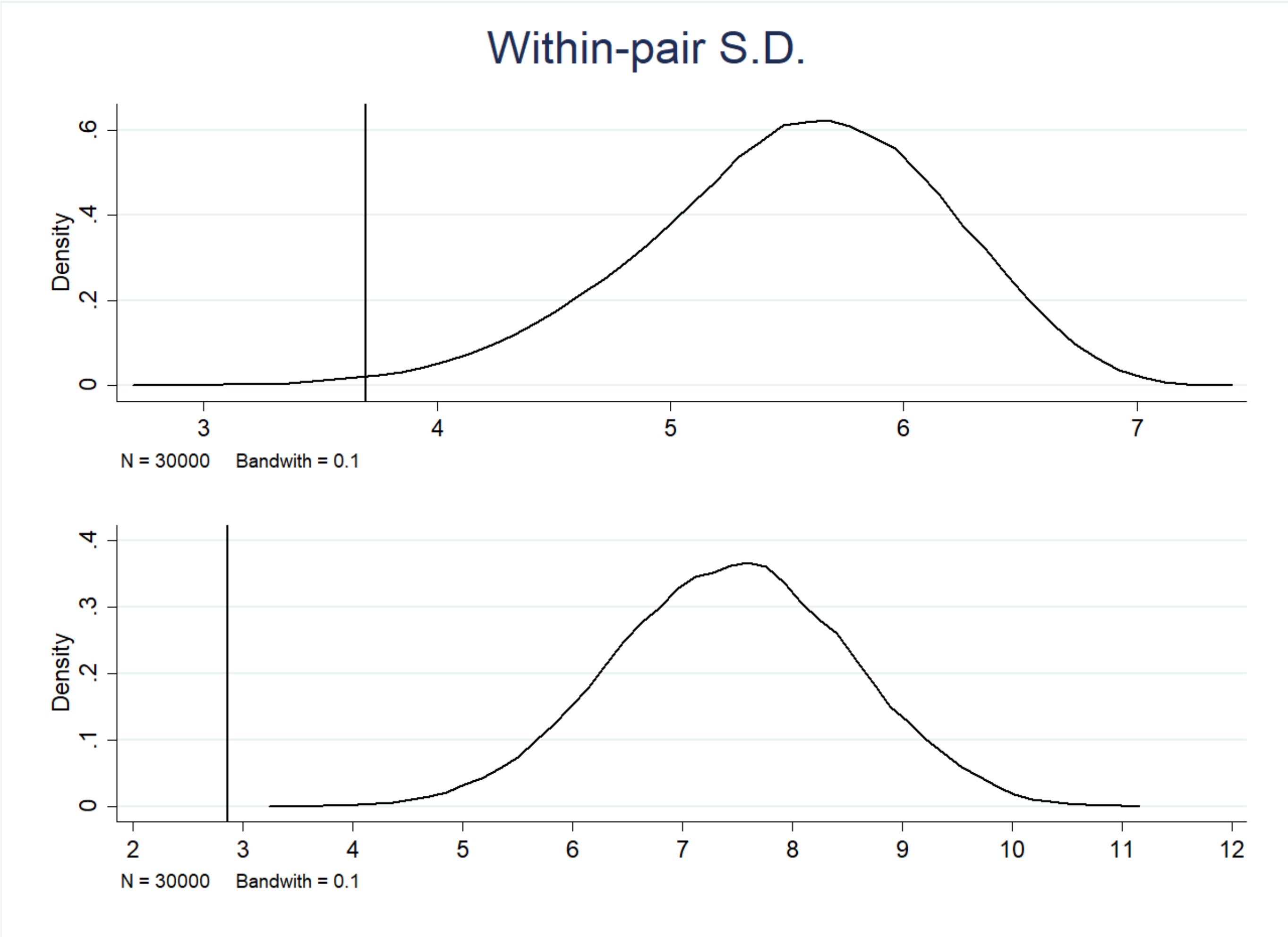


Figure C.2: **Top panel:** Within-Pair Standard Deviation in Reported Value in the PR-CF and PR-FF Treatments. The vertical line represents the average within-pair S.D. in the PR-CF treatment while the curve represents the within-pair S.D. density of the simulated pairs in the PR-FF treatment. **Bottom panel:** Within-pair standard deviation of the value reported in the TR-CF and TR-FF treatments. The vertical line represents the average within-pair S.D. in the TR-CF treatment while the curve represents the within-pair S.D. density of the simulated pairs in the TR-FF treatment..

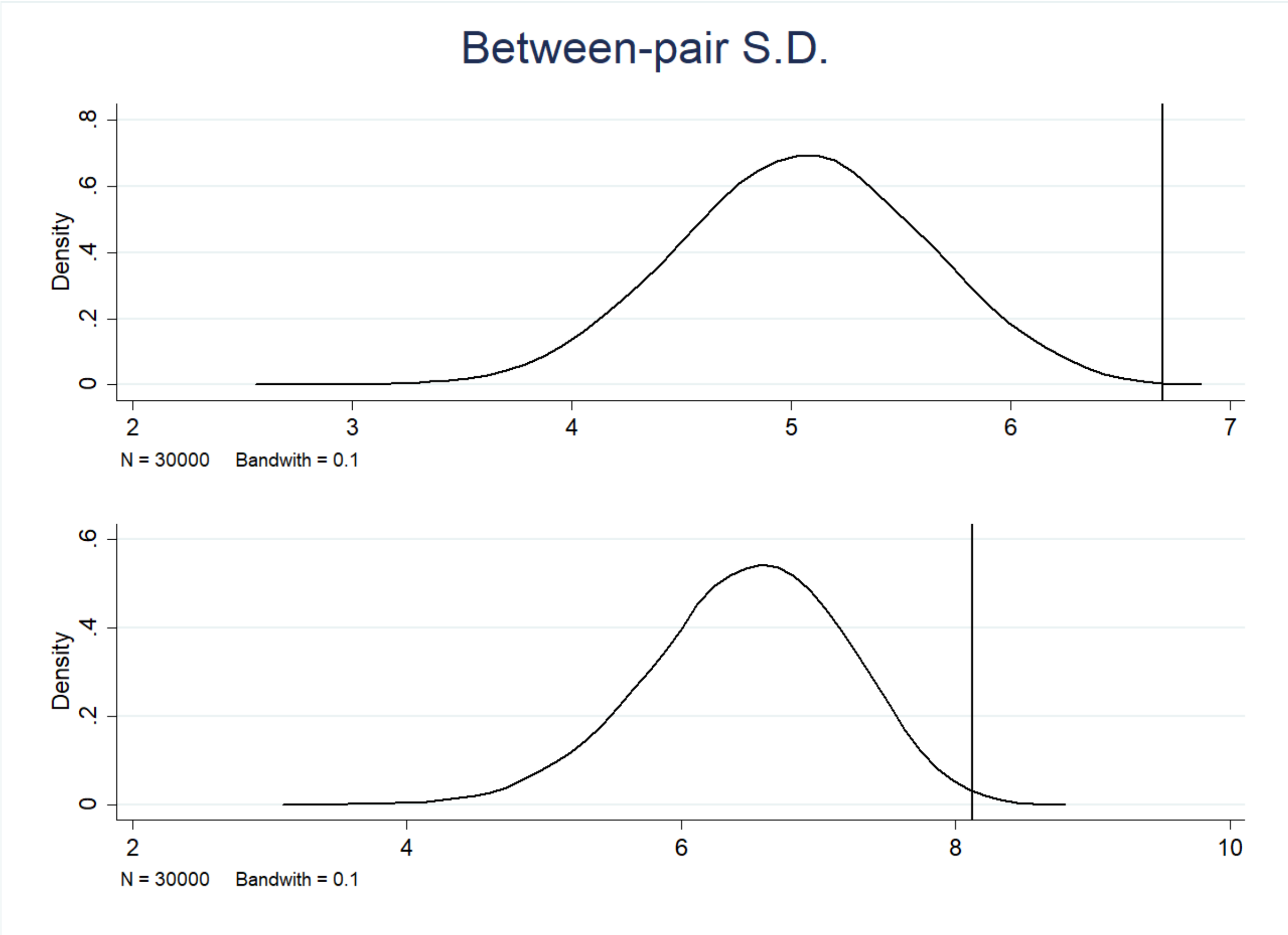


Figure C.3: **Top panel:** Between-Pair Standard Deviation in Reported Value in the PR-CF and PR-FF Treatments. The vertical line represents the average between-pair S.D. in the PR-CF treatment while the curve represents the between-pair S.D. density of the simulated pairs in the PR-FF treatment. **Bottom panel:** Between-pair standard deviation of the value reported in the TR-CF and TR-FF treatments. The vertical line represents the average between-pair S.D. in the TR-CF treatment while the curve represents the between-pair S.D. density of the simulated pairs in the TR-FF treatment.

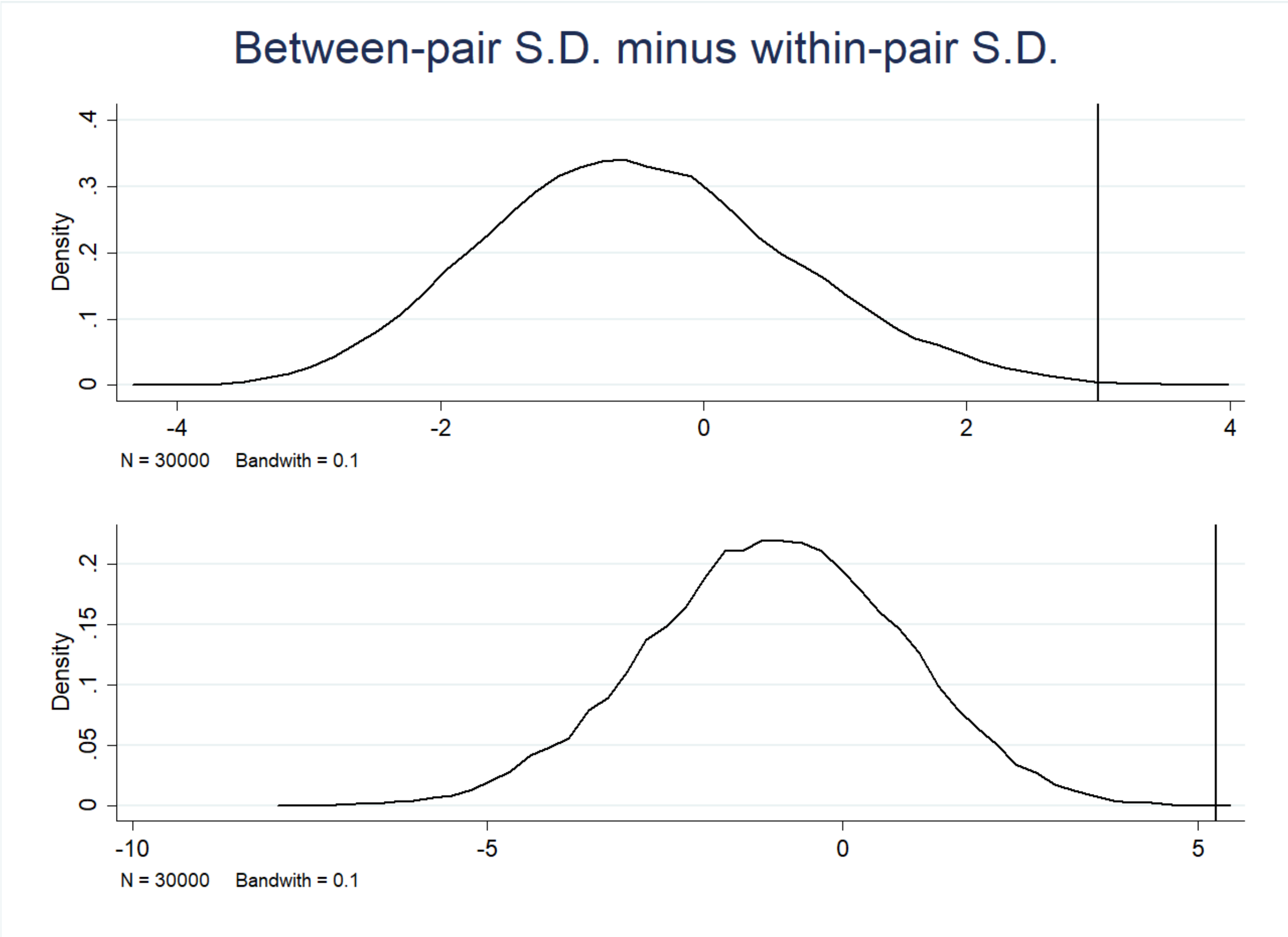


Figure C.4: **Top panel:** Between-Pair Minus Within-Pair Standard Deviation in the PR-CF and PR-FF Treatments. The vertical line represents the average between-pair minus within-pair S.D. in the PR-CF treatment while the curve represents the density of the difference between the between pair and the within-pair S.D. of the simulated pairs in the PR-FF treatment. **Bottom panel:** Between-pair minus within-pair standard deviation in the TR-CF and TR-FF treatments. The vertical line represents the average between-pair minus within-pair S.D. in the TR-CF treatment while the curve represents the density of the difference between the between pair and the within-pair S.D. of the simulated pairs in the TR-FF treatment.

D Additional Results

D.1 Effect of Gender

Figure D.1 displays the probability to report a red outcome, by gender and by treatment. Estimates are computed from a multinomial logit regression model similar to the one reported in Table B.3, but with the inclusion of an interaction between each treatment dummy and the male variable. Figure D.1 shows that males were significantly more likely to report a red outcome than females in the TR-FF treatment, suggesting a higher dishonesty of males in this environment ($p < 0.001$).[32] In line with this result we also find that 47.8 % of the males in the TR-FF treatment cheated to the full extent while only 4.8% of the females did. This gap in the extent of lying is not present in any other treatment.[33]

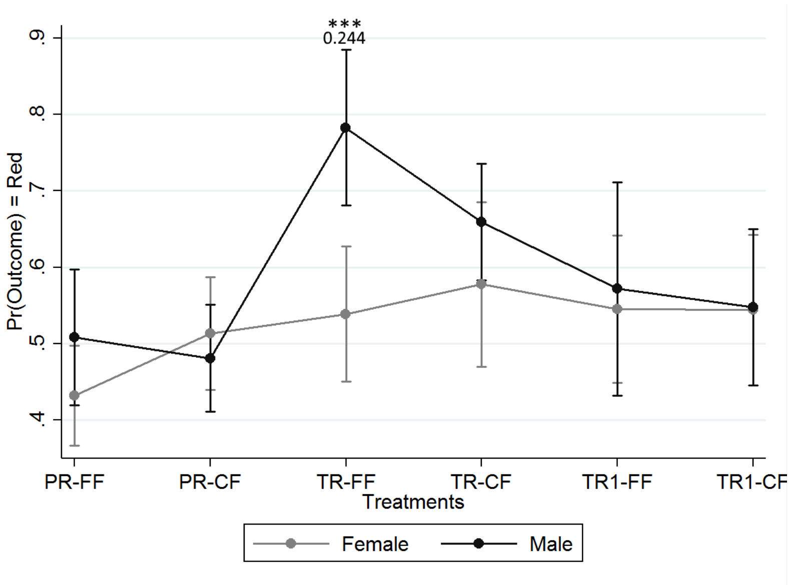


Figure D.1: Probability of Reporting a Red Outcome, by Gender and Treatment
Notes: Bars represent 95% confidence intervals. The numerical value is the coefficient of the marginal effect of being a male on the probability of reporting a red outcome. Stars represent significance of the marginal effect of the male variable in a given treatment. *** indicates significance at the 1% level.

These results cannot be explained by a higher competitiveness of males because we should observe that males cheated more than females in all treatments with competition, which is not the case. An alternative explanation is that males formed different beliefs about the honesty of their counterpart. They might have expected a higher level of dishonesty from their counterpart in the TR-FF treatment, compared to females. We cannot test this hypothesis directly because we elicited the beliefs about the empirical norm after the participants received feedback on their counterpart's reports (to hold this constant across treatments with and without continuous feedback). Thus, participants have been able to update their beliefs, as illustrated by the significant correlation between their beliefs and their opponent's performance in the TR-FF treatment ($\rho = 0.530, p <$

[32]Males were also less likely to report a blue or a yellow outcome than females in the TR-FF treatment ($p = 0.019$ and $p < 0.001$, respectively).
[33]This is also the case in the follow-up experiment. The analysis is available from the authors upon request.

0.001).

D.2 Happiness and Nervousness

Wilcoxon signed rank tests show that in the piece-rate treatments of the main experiment participants' happiness and nervousness did not change significantly over time (difference in happiness: -0.088, $p = 0.471$; difference in nervousness: -0.471, $p = 0.164$). In contrast, in the tournament treatments participants reported less happiness and more nervousness at the end compared to the beginning of the session (difference in happiness: -0.746, $p < 0.001$; difference in nervousness: 0.738, $p = 0.001$). Similar effects are observed in the TR1 treatments (difference in happiness: -0.725, $p < 0.001$; difference in nervousness: 1.125, $p < 0.001$).[34] In both the TR and TR1 treatments, the drop in happiness was larger for underdogs than front runners, though the difference is not significant or only marginal significant (-1.098 vs. -0.608 , $p = 0.240$ for the TR treatments; -1 vs. -0.466 $p = 0.055$ for the TR1 treatments). Also, nervousness increased only for the underdogs (2.157, $p < 0.001$ in TR; 2.621, $p < 0.001$ in TR1) while it decreased for the front runners (-0.471, $p = 0.003$ in TR; -0.339, $p = 0.022$ in TR1). Competition increases stress (*e.g.,* Buser et al., 2017), and this may explain the drop in happiness. The effect on nervousness depends instead on the result of the competition.

If we pool the different payment schemes together, we find that the continuous provision of social information did not affect the level of happiness (difference in happiness when feedback is continuous vs. final: -0.599 vs. -0.424, $p = 0.390$) but it reduced nervousness (0.387 vs. 0.833, $p = 0.027$).

In the follow-up experiment, we find that happiness significantly decreased (increased) over time under tournament (piece-rate) incentives (difference in happiness: -0.446 and 0.543, $p = 0.032$ and 0.023 in bTR-CF and bPR-CF, respectively). Nervousness remained instead relatively stable (difference in nervousness: -0.089, $p = 0.713$ in bTR-CF; -0.152, $p = 0.447$ in bPR-CF). The drop in happiness in bTR-CF was mainly driven by underdogs (difference in happiness: -0.769, $p = 0.007$ for underdogs; 0, $p = 0.478$ for front runners). Also, like in the main experiment, under tournament incentives, nervousness increased for underdogs (0.846, $p = 0.016$), while it decreased for front runners (-0.962, $p = 0.021$). If we consider the bPR-CF treatment, the increase in happiness was driven by front runners (difference in happiness: 1.1, $p = 0.005$ for front runners; 0.05, $p = 0.938$ for underdogs), while nervousness remained stable over time for both front runners (-0.5, $p = 0.126$) and underdogs (0.15, $p = 0.357$). The fact that happiness increased for front runners only under piece-rate incentives suggests that outperforming the counterpart gives pleasure to subjects as long as it does not hurt economically the other player. Interestingly, we do not find a similar increase in happiness for front runners in the PR-CF of the main experiment. In fact, for these subjects, happiness slightly decreased (by 0.133). A possible explanation is that subjects in bPR-CF had a stronger desire to outperform the counterpart than subjects in PR-CF, and this is because they were constantly reminded to think about the others in the belief elicitation.

[34] We pooled together the continuous and final feedback conditions to keep the number of comparisons low. Also, in the TR1 treatments we do not distinguish between active and passive subjects. A more disaggregated analysis is available from the authors upon request.

D.3 Contagiousness of Cheating in the Follow-up Experiment

Table D.1 displays the results of a multinomial logit regression similar to the one reported in Table B.3 for the main experiment. The probability of reporting a red (yellow) outcome increased (decreased) over time. We also find no significant differences between the bTR-CF and bPR-CF treatments. If we interact the trend variable (Period) with the treatment dummy bTR-CF, we find a similar effect of time in the two treatments (see Figure D.2).

Table D.1: Determinants of the Die Roll Report

	Blue (0 point)	Yellow (1 point)	Red (2 points)
bTR-CF	-0.019	-0.043	0.062
	(0.029)	(0.035)	(0.059)
Period	-0.001	-0.004***	0.005***
	(0.001)	(0.001)	(0.002)
Other controls		Yes	
Observations		2448	
Number of clusters		51	
Pseudo-R^2		0.018	
p>chi2		<0.001	

Notes: The table reports marginal effects of a multinomial logit regression. Standard errors (in parenthesis) are clustered at the pair level. Other independent variables include male gender, risk attitude, age, and student status. *** indicates significance at the 1% level.

We also compute the within and between-pair SD of the total value reported at the end of the 24 periods. The average within-pair SD is not significantly different between the two treatments (bPR-CF: 5.257; bTR-CF: 4.467; Wilcoxon ranksum: $p = 0.543$).[35] Like in the main experiment, we reject the hypothesis that the contagiousness of cheating was stronger under tournament than piece-rate incentives.

D.4 Belief Updating in the Follow-up Experiment

To study how participants updated their beliefs over time, we estimate the following OLS regression model:

$$\Delta B_i^t = \beta_0 + \beta_1 \times t + \beta_2 \times \text{bTR-CF} + \beta_3 \times t \times \text{bTR-CF} \quad (1)$$

$$+(\beta_4 + \beta_5 \times \text{bTR-CF}) \times \left(\max\left\{0, \frac{\sum_{n=1}^{t-1} R_j^n}{t}\right\} - B_i^{t-1} \right) \quad (2)$$

$$+(\beta_6 + \beta_7 \times \text{bTR-CF}) \times \left(\max\left\{0, B_i^{t-1} - \frac{\sum_{n=1}^{t-1} R_j^n}{t}\right\} \right) \quad (3)$$

$$+(\beta_8 + \beta_9 \times \text{bTR-CF}) \times \left(\max\left\{0, \frac{\sum_{n=1}^{t-1} R_i^n}{t} - B_i^{t-1}\right\} \right) \quad (4)$$

[35] The between-pair SD is equal to 7.482 in the bPR-CF and 7.356 in the bTR-CF. Qualitatively similar values were obtained in the main experiment for PR-CF and TR-CF.

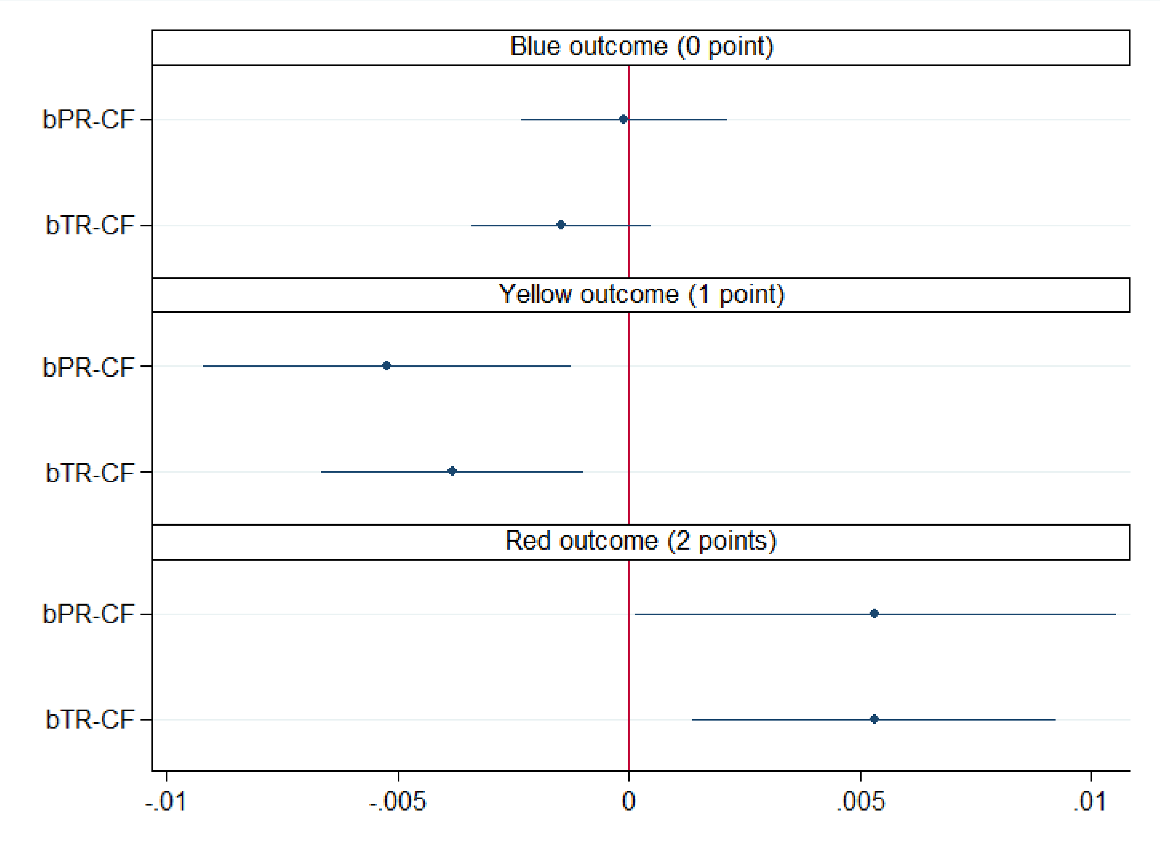


Figure D.2: Marginal effects of time on the probability to report each outcome, by treatment. *Notes:* The horizontal lines represent 95% confidence intervals.

$$+(\beta_{10} + \beta_{11} \times \text{bTR-CF}) \times \left(\max \left\{ 0, B_i^{t-1} - \frac{\sum_{n=1}^{t-1} R_i^n}{t} \right\} \right) \tag{5}$$

The dependent variable (ΔB_i^t) is the difference in the average belief of subject i between periods t and $t-1$. It is positive (negative) if a subject updated his or her beliefs upward (downward) from one period to the other. The model contains the following explanatory variables. In (1), we include a time variable (t), a dummy for the bTR-CF treatment, and the interaction between these two variables. In (2) and (3), we add two variables measuring the negative and positive deviations between i's average belief in $t-1$ (B_i^{t-1}) and the average report of the counterpart j until $t-1$ ($\frac{\sum_{n=1}^{t-1} R_j^n}{t}$), and their interactions with bTR-CF. A negative (positive) deviation captures the extent to which i underestimate (overestimate) the norm compared to the average signal received from j. We expect participants to update their perception of the norm (their beliefs about the average report of 24 past participants) depending on the signals they receive from the counterpart j. If they see that the counterpart is under-reporting compared to what they think is the norm (*i.e.*, $\frac{\sum_{n=1}^{t-1} R_j^n}{t} < B_i^{t-1}$), we expect them to adjust their beliefs downward. In contrast, if they see that the counterpart is over-reporting compared to their perception of the norm (*i.e.*, $\frac{\sum_{n=1}^{t-1} R_j^n}{t} > B_i^{t-1}$), we expect them to adjust their beliefs upward.

In (4) and (5), we add the negative and positive deviations between i's average belief about the norm in $t-1$ and i's average report until $t-1$ ($\frac{\sum_{n=1}^{t-1} R_i^n}{t}$), and their interactions with bTR-CF. These variables measure the extent to which i's behavior deviates from i's perceived norm in the previous period. If participants consider their own behavior equally informative about the norm, own deviations from the norm in one period may have an impact on the beliefs in the subsequent period.[36] In particular, we expect subjects to adjust their beliefs upward (downward), if they have over-reported (under-reported) in the previous periods compared to their perceived norm.

All the interaction terms in the model capture treatment differences in belief updating. We also control for risk attitude, age, gender and student status. Standard errors are clustered at the pair level.

Table D.2: Dynamics of Beliefs

Dependent variable	ΔB_i^t
Period (β_1)	−0.001 (0.001)
bTR-CF (β_2)	−0.003 (0.035)
Period × bTR-CF (β_3)	−0.001 (0.001)
Negative deviation of belief re norm from j's average report (β_4)	0.086** (0.032)
Negative deviation of belief re norm from j's average report × bTR-CF (β_5)	−0.045 (0.046)
Positive deviation of belief re norm from j's average report (β_6)	−0.144** (0.058)
Positive deviation of belief re norm from j's average report × bTR-CF (β_7)	0.051 (0.065)
Negative deviation of belief re norm from own average report (β_8)	0.014 (0.052)
Negative deviation of belief re norm from own average report × bTR-CF (β_9)	0.085 (0.093)
Positive deviation of belief re norm from own average report (β_{10})	−0.305*** (0.066)
Positive deviation of belief re norm from own average report × bTR-CF (β_{11})	0.234*** (0.081)
Other controls	Yes
Observations	2346
Number of clusters	51
p>chi2	0.001

Notes: The table reports the coefficients of an OLS regression on the difference between the participants' beliefs regarding the norm (i.e. the beliefs about the average report of 24 past participants) in one period and the previous one. Standard errors (in parentheses) are clustered at the pair level. Other independent variables include risk attitude, age, male gender and student status. *** indicates significance at the 1% level, and ** at the 5% level.

Table D.2 reports the results of this estimation. It confirms that participants expected a higher (lower) average report in the group of 24 if they had previously under-estimated (over-estimated) the norm compared to the signals they had received from j. This effect is analogous in the two treatments. Subjects updated their beliefs downward if they had previously under-reported the die rolls compared to their perception of the norm. However, this is statistically significant only

[36]For example, consider a subject who, at a given period t, thinks that not many past participants cheated in that period. (S)he then rolls a blue and realize that the temptation to report a red is too high. (S)he thus reports a red. At the same time, (s)he may anticipate that others will behave analogously and, therefore, (s)he update the norm in the subsequent periods.

for the bPR-CF treatment. Finally, we find no significant time trend on belief updating in both treatments.

Finally, we examine whether beliefs have an impact on behavior. We conduct a multinomial logit regression where the dependent variable is the outcome reported by subject i in period t. The independent variables are a treatment dummy for bTR-CF, the time trend, the average belief of subject i in period t, and interactions between the treatment dummy and the other two variables. We control for risk attitude, age, gender and student status. Standard errors are clustered at the pair level. Figure D.3 reports the marginal effects of the belief variable on the likelihood to report a blue, a yellow or a red outcome in each treatment. In both the bPR-CF and the bTR-CF treatments, participants were more likely to report a red outcome ($p = 0.014$ and $p < 0.001$, respectively) and less likely to report a yellow ($p = 0.014$ and 0.001) and a blue outcome ($p = 0.020$ and 0.001) if they expected higher reports. This suggests that participants tended to over-report more, the more they thought the others engaged in this behavior. Note that this could also result from a confirmation bias.

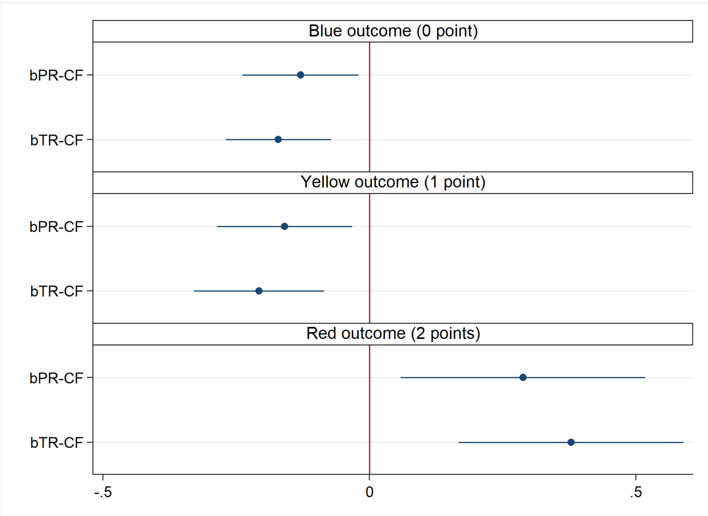


Figure D.3: Marginal Effects of Beliefs on the Probability to Report Each Outcome, by Treatment.

*Notes:*The figure displays the marginal effects of the belief variable on the likelihood to report a blue, a yellow or a red outcome in each treatment. Estimates are computed from a multinomial logit regression model. The horizontal lines represent 95% confidence intervals.