



# Competition, Information, and the Erosion of Morals

Julien Benistant, Fabio Galeotti, Marie Claire Villeval

## ► To cite this version:

Julien Benistant, Fabio Galeotti, Marie Claire Villeval. Competition, Information, and the Erosion of Morals. Journal of Economic Behavior and Organization, 2022, 204, pp.148-163. <10.1016/j.jebo.2022.10.008>. <hal-03805532>

**HAL Id: hal-03805532**

**<https://hal.science/hal-03805532v1>**

Submitted on 7 Oct 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

# Competition, Information, and the Erosion of Morals \*

Julien Benistant<sup>†</sup>, Fabio Galeotti<sup>‡</sup> and Marie Claire Villeval<sup>§</sup>

## Abstract

We study the impact of competition on morals using a dynamic variant of the die-under-the-cup task. Players can repeatedly misreport the outputs of consecutive die rolls to earn more money, either under an individual piece-rate pay scheme or in a two-player tournament. In this dynamic setting, we disentangle the effect of the incentive scheme and the effect of information provision about one's relative performance, by comparing settings with continuous *vs.* final *ex post* feedback on the counterpart's reported outcome. We find that individuals lie more under competitive rather than non-competitive incentive schemes, but only if both players can cheat in the tournament. Continuous feedback on the counterpart's reports does not increase cheating in the tournament, while it does under the piece-rate scheme. These findings shed light on the effects that different competition and information policies have on morals in occupational settings.

**Keywords:** Dishonesty, feedback, peer effects, competitive incentives, experiment

**JEL Codes:** C92, M52, D83

---

\* *Acknowledgments:* We thank the Editor, Ginny Choi, and two reviewers, whose comments and suggestions significantly improved the paper. We are also grateful to A. Hopfensitz, A. Kajackaite, M. Lefebvre, as well as the participants of the 2022 ASFEE conference in Lyon, for useful comments. Finally, we thank Q. Thevenet for research assistance. This research has benefited from the support of the project DECISION of the French National Research Agency (ANR-19-CE26-0019), the support of IDEXLYON from Université de Lyon (project INDEPTH) within the Programme Investissements d'Avenir (ANR-16-IDEX-0005), and it was performed within the framework of 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. 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.

<sup>†</sup>Univ Lyon, CNRS, ISC Marc Jeannerod, UMR 5229, Bron, France; julien.benistant@isc.cnrs.fr

<sup>‡</sup>Univ Lyon, CNRS, GATE, UMR 5824, F-69130 Ecully, France; fabio.galeotti@cnrs.fr

<sup>§</sup>Univ Lyon, CNRS, GATE, UMR 5824, F-69130 Ecully, France; IZA, Bonn, Germany; villeval@gate.cnrs.fr

# 1 Introduction

The question of whether markets erode morals has attracted the attention of social thinkers for centuries, and the debate is still ongoing today (*e.g.*, Montesquieu 1748; Smith 1776; Marx 1867; Schumpeter 1942; Hirschman 1977; Shleifer 2004; McCloskey 2006; Satz 2010; Sandel 2012; Bruni and Sugden 2013; Bowles 2016; Storr and Choi 2019).<sup>1</sup> Market critics often argue that markets undermine moral values by shifting people’s attention from their preexisting moral commitments to self-interested cost-benefit considerations (Bowles, 2011). This effect may be caused or exacerbated by competition, which is typically considered as an essential component of how markets work. Other often cited explanations for the erosion of morals in markets are the diffusion of responsibility, which reduces the expected costs of acting immorally, and the transmission of social information about the acceptability of certain unethical behaviors (Falk and Szech, 2013; Sutter et al., 2020).

Market defenders contrast this perspective by stressing the fact that markets are primarily places of cooperation and not ruthless competition. In markets, participants can realize mutual benefits, improve their well-being, and achieve society-wide coordination of resources, goods and services. In this conception, competition and social information can actually be good for morals because they allow individuals to build a moral reputation and force out those who do not live up to the society’s moral standards.

Inspired by this debate, our study explores how different pay schemes (competitive *vs.* non-competitive), coupled together with different feedback policies on the individuals’ performance, may affect morality in the workplace.<sup>2</sup> Our goal is not to test directly whether markets *per se* erode morals but rather to explore under which conditions certain instruments that are created by the market to improve the productivity of the workplace can be harmful for morals. More specifically, we study how social information about others’ conduct influences individual cheating behavior under a tournament pay scheme, and whether its effects are different under an individual piece-rate scheme.<sup>3</sup> Additionally, we investigate whether the level of cheating in competitive pay settings results from the material incentives produced by the competition itself or from the anticipation of an unfair competition due to the opponent’s cheating behavior. We study this in the context of a laboratory experiment, which allows us to examine the impact of social information and competition on cheating behavior over multiple periods.

---

<sup>1</sup>For recent empirical contributions on this topic see, for example, Falk and Szech (2013); Al-Ubaydli et al. (2013); Sutter et al. (2020); Choi and Storr (2020); Bartling et al. (2021).

<sup>2</sup>We focus on a specific moral value: honesty. We do that by studying cheating behavior. This contrasts with previous studies that look at sabotage. While sabotage is about increasing one’s chances of winning by reducing the opponent’s performance (*e.g.*, Lazear, 1989; Carpenter et al., 2010; Harbring and Irlenbusch, 2011), cheating consists of artificially improving one’s output, for example, by taking performance-enhancing drugs or forging data (Charness et al., 2014).

<sup>3</sup>Both tournaments and piece rates are common means of rewarding workers for their performance. 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. Tournaments are commonly employed to determine promotions, hiring, raises and bonuses (Lazear, 2018).

Although they also try to induce cooperation among employees, many firms in market economies provide competitive incentives to their employees in the form of a promotion or a bonus to motivate them. Similarly, companies use highly competitive selection processes to recruit the most qualified candidates for job openings. Such market instruments have a positive impact on employees' work performance or on the ability of companies to sort employees according to their talents (Prendergast, 1999). However, they might also have adverse consequences regarding workers' morality; for example, exposure to sabotage can discourage effort and cooperation (*e.g.*, Carpenter et al., 2010; Berger et al., 2013). They may induce individuals to adopt fraudulent practices, such as résumé inflation, fabricating positive results, or deceiving customers to win a tournament. Examples of this abound (Association of Certified Fraud Examiners, 2020), especially in the domain of credence goods: for example, salesmen who lie to customers about the quality of their products in order to sell more and win a sales contest (Poujol, 2009). There are also examples regarding public officers cheating in national promotion tests (Fitch, 2013), or applicants exaggerating or misrepresenting details about their experience or skills in order to increase their chances of being hired in a company (Weiss and Feldman, 2006; Levashina and Campion, 2007).

Unfortunately, company data on employees' cheating behaviors are rare and thus, the identification of the precise role of competition and social information on morals is almost impossible with field data. Most of what we know about the effects of competitive incentives on cheating behavior comes from laboratory experiments. These studies confirm that individuals tend to cheat more under competitive pay schemes (*e.g.*, Cadsby et al. 2010; Schwieren and Weichselbaumer 2010; Belot and Schröder 2013; Faravelli et al. 2015; Rigdon and D'Esterre 2015; Dutcher et al. 2016; Aydogan et al. 2017; for a survey, see Gangadharan et al. 2020).<sup>4</sup> However, little is known about what aspects of the competition are responsible for the erosion of moral values. Is cheating behavior motivated by the direct material incentives of the competition, by peer effects (in terms of conformity to a norm and rivalry) induced by repeated comparison with a counterpart, or by the fear of a biased competition due to the possible unethical behavior of the competitor? Disentangling these mechanisms is crucial to understanding how competitive incentives affect morality in occupational settings.

Another underexplored question concerns the impact of different feedback policies on cheating behavior in competitive *vs.* non-competitive settings. As is standard in the personnel economics literature, by feedback we mean the information that agents receive regarding their output relative to that of other employees. A company typically has dis-

---

<sup>4</sup>We only consider the literature on tournaments here, as this is the main focus of our study. Other studies have also examined the effects of different market structures on unethical behavior (*e.g.*, Falk and Szech, 2013; Rabanal and Rud, 2018; Feltovich, 2019), the spillover effects of competition on post-competition behaviors (*e.g.*, Buser and Dreber, 2016; Schurr and Ritov, 2016), and the link between the strength of the competition and the occurrence of cheating (*e.g.*, Conrads et al., 2013; Cartwright and Menezes, 2014).

cretion regarding its feedback policy: it can continuously inform its employees about their relative output to favor emulation or, instead, provide limited feedback (such as annual performance evaluations) or even conceal information until the end of a competition.<sup>5</sup> Which of these feedback policies is more effective in limiting cheating?<sup>6</sup> If the information about an individual’s output reveals something about his morality, the provision of this information during competition could affect the moral conduct of the other employees. Previous studies found that, in non-competitive settings, individuals tend to cheat more when they are informed, even imperfectly, about the cheating behavior of other individuals (*e.g.*, Gino et al., 2009; Fosgaard et al., 2013; Robert and Arnab, 2013; Rauhut, 2013; Diekmann et al., 2015; Lefebvre et al., 2015; Soraperra et al., 2017; Dimant, 2019; Bäker and Mechtel, 2019; Lauer and Untertrifaller, 2019; Charroin et al., 2022). Are these peer effects—driven by rivalry or conformity—stronger in competitive settings because of the monetary stakes? On the one hand, under competitive incentives, individuals have a pecuniary interest to mimic their opponents’ cheating behavior, which could increase even more cheating than non-competitive settings, as it has been observed for sabotage. On the other hand, feedback on others’ behavior reduces uncertainty about the relative position of the opponent, which may lead to less cheating since the success of the competition does not depend on absolute performance levels but on outperforming the competitor. We investigate which of these effects predominates.

Finally, little is known about the *dynamics* of cheating behavior under different payment schemes and feedback policies. Previous experimental studies on the effects of incentives or feedback on dishonesty have mainly focused on static settings where participants make one-shot decisions or play over very few rounds. However, how does cheating behavior evolve over multiple periods, and how does this depend on the payment scheme and feedback policy in use? A recent meta-analysis of Abeler et al. (2019) suggests that dishonest behavior does not change over time. Meanwhile, 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. In addition to these conflicting results obtained in non-competitive settings, we know very little about how repetition affects cheating under competitive pay schemes and how the provision of social information affects the evolution of cheating behavior over time in such settings.

To fill the aforementioned gaps in the literature, we designed a laboratory experiment in which individuals were matched in pairs and played individually and repeatedly over several periods a variant of the die-under-the-cup task (Fischbacher and Föllmi-Heusi, 2013; Shalvi et al., 2011). In each period, participants had to roll a die in private with no scrutiny and report the outcome on their computer. Their final output was computed as the sum of

---

<sup>5</sup>See examples in Chapter 9 of Lazear and Gibbs (2014).

<sup>6</sup> Most of the literature on feedback policies examines their impact on employees’ performance (for a survey, see Villeval 2020). We depart from it by focusing on cheating behavior rather than effort.

their individual reports in all periods. Hence, this output was only a function of luck and honesty like in other organizational experiments on cheating behavior (*e.g.*, d’Adda et al., 2017). This was done to capture the pure effects of feedback and incentives on cheating behavior and isolate them from the effect on effort and other possible confounds (*e.g.*, entitlement effects). This design reflects real-world settings where the employees’ job is to report output or quality measures (*e.g.*, the results of an emission test or some financial results) rather than to exert a real effort. It captures situations in which performance does not depend on the agents’ effort choices but, for example, on their skill. We varied between participants whether they were paid according to a piece rate or a tournament payment scheme and whether they received continuous feedback about their counterpart’s reports at the end of each period or only at the end of the task. Such feedback provides participants with a noisy signal about the prevailing dishonesty since the reports of several high outputs by the counterpart may indicate luck or misreporting.

We also designed a variant of the tournament treatment, under each feedback condition, where it was common knowledge that 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 to what extent the difference in cheating between the individual piece rate and the tournament pay scheme, if any, is driven by the competitive nature of the incentive or by the individuals’ beliefs about the dishonest behavior of the counterpart, and thus about the unfairness of the competition. If the 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 competitive incentive itself. If not, this would suggest that the competition itself generates misbehavior, regardless of the counterpart’s behavior, for monetary and possibly also non-monetary reasons, such as the pure joy of winning by a larger margin (*e.g.*, Parco et al., 2005; Cason et al., 2018).

Our results show that participants over-reported 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 final feedback under the piece-rate scheme, suggesting the existence of peer effects, but not in tournaments. Despite the latter result, participants reacted to their counterpart’s perceived level of honesty both in tournaments and in the piece-rate pay scheme.

To better understand the role of beliefs, we conducted a follow-up experiment where we elicited the participants’ beliefs about the prevalence of cheating behavior in each period under tournament and piece-rate incentives, respectively. The results of this experiment reveal that individuals underestimated more the extent of cheating when they were paid a piece-rate, compared to when they competed in a tournament. They adjusted their perception of the norm and their behavior depending on the information received about

their counterparts.

Overall, our findings show that competitive incentives and continuous feedback should be used with care to limit the contagion of dishonesty, but also that the erosion of morals in competitive settings can be significantly curbed when it is made clear that the opponent cannot cheat. Importantly, the stronger negative effect of continuous feedback on honest reporting observed under non-competitive incentives suggests that conformity to a norm and/or rivalry plays an important role in morality besides incentive schemes. Therefore, it is particularly important to include both social norms and individuals' behavioral motivations when discussing the responsibility of market institutions in the erosion of morals.

The following section describes the design and procedure of the main experiment. Section 3 presents the results of this study. Section 4 introduces our follow-up experiment and summarizes the main results. Section 5 discusses the findings of both experiments and concludes the study.

## 2 Design and Procedures of the Main Experiment

This section introduces the design of the main experiment and describe the procedure.

### 2.1 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), as in Dai et al. (2018). The possible outcomes of the die roll were “red,” “yellow,” or “blue” with equal probability (1/3). The die was placed inside a sealed cup to increase privacy, as done in Shalvi et al. (2011). In each period, the participants had 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. The participants' output in this task was computed as the sum of points earned across the 24 periods.<sup>7</sup>

We purposely chose a task where the output is only determined by luck and honesty in order to isolate the pure effect of information and incentives on cheating behavior.<sup>8</sup> Particularly, we get rid of several possible confounds that would arise if the output was determined by effort. First, our results are not influenced by the heterogeneity of the participants' innate abilities or effort, which is known to vary across payment schemes and feedback policies (see Villeval, 2020); second, participants have to form beliefs about

---

<sup>7</sup>Individuals performed the same task across treatments. Thus, the potential effects of a lucky or unlucky draw in the first few periods apply equally to all conditions and cannot drive differences across treatments.

<sup>8</sup>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 in cheating behavior.

the honesty of their counterpart, but not on their ability and effort; third, we prevent subjective feelings of entitlement from arising. Our task mimics real-world settings where an agent overreports an output measure (*e.g.*, the quality of a product, the profitability of a portfolio, company earnings), which is determined in part or fully by factors out of the agent’s control (*e.g.*, the volatility of markets, the agent’s innate abilities, other employees’ work). 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 entire duration of the experiment. One dimension varied the payment scheme: participants were either paid based on an individual piece-rate scheme or entered 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 counterparts 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 identifying the role of beliefs about the counterpart’s dishonesty. We ran two versions of this tournament with asymmetric cheating opportunities, one where the feedback on relative output was continuous and one where it was given only at the end of the 24 periods. Hence, we ran six treatments.

### 2.1.1 Payment Schemes

In the *Piece Rate* treatment (PR), 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 reports. In the *Tournament* treatment (TR), the pair member with the highest number of points accumulated at the end of the 24 periods earned €0.36 per point, whereas the counterpart earned €0. As such, the expected payoffs were the same in the two payment schemes, assuming a 50% chance of winning the tournament.

We paid participants based on the sum of their reports in all 24 periods, instead of one or a few rounds selected at random, for different reasons. First, it reflects what typically occurs in the field. For example, in sales contests, salesmen usually compete over cumulative sales for several weeks (*e.g.*, Delfgaauw et al., 2013). Second, it prevented frequent ties in the tournament (two competitors with the same total output). Third, this procedure did not add background risk to the decision-making setting.



### 2.1.2 Information

In the *Final Feedback* treatment (FF), participants were only informed at the end of the experiment about the reports made by their counterparts in each of the 24 periods. Thus, this treatment kept uncertainty about one’s relative standing and the honesty of the counterpart until the end of the experiment.

In the *Continuous Feedback* treatment (CF), participants were informed about the reports made by their counterparts 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 C.1 in Appendix C). As soon as both participants entered their reports, the box was updated with new information, and a new period began. Thus, in this treatment, participants were continuously informed about the difference in output between themselves and their counterparts, and their evolution over time, in each period. This feedback provided noisy information about the honesty of the counterpart. Over time, participants could perceive whether the counterpart was dishonest by comparing the distribution of reports with the theoretical uniform distribution, but they could never be certain whether a report of the counterpart in a given period was honest.

### 2.1.3 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), only one pair member was requested to roll the die and report the outcome. The other competitor was passive. The passive player 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 the roles were kept fixed throughout the 24 periods. Both participants within each pair were aware of the profile of their competitor (passive or active). All other aspects of the task were identical to those of 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 anticipate that their counterpart will cheat. This additional treatment manipulation allowed us to mute the second channel. Table 1 summarizes our treatments.

Table 1: Treatments and Conditions

Treatments	Piece-Rate (PR)	Tournament (Two-Sided) (TR)	Tournament (One-Sided) (TR1)
Final Feedback (FF)	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 parentheses for each condition. Differences in the number of participants reflect (a) different attendance rates across sessions; (b) the need for more observations in the Continuous Feedback conditions is because the independent observation is the pair and not the participant.

#### 2.1.4 Additional Measures

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).<sup>9</sup> Before the main task, we also elicited participants' risk preferences, using the procedure of Gneezy and Potters (1997).<sup>10</sup> Finally, we elicited participants' beliefs about others' reports at the end of the experiment (after the second round of questions about emotions). 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.<sup>11</sup> We elicited the participants' beliefs for the first, the thirteenth, and the last period. The objective was to obtain a rough idea about the participants' perceptions 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 this measure is biased since beliefs were elicited after participants learned the distribution of reports made by their counterparts. In fact, we chose not to elicit beliefs during the task to prevent subjects from focusing on those beliefs and how to best respond to them. Indeed, previous studies found that such elicitation may confound the results (see, e.g., Croson, 2000; Gächter and Renner, 2010, in the context of public good games). We elicited the beliefs between periods in the follow-up experiment, allowing these potential biases to manifest (see details in Section 5).

<sup>9</sup>Appendix A reports the instructions and the illustrations used to report one's feelings.

<sup>10</sup>Participants were endowed with 150 Euro cents and were asked to choose how much to invest in a lottery. This lottery had two possible outcomes with an 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. The participants received feedback on their earnings in this task only at the end of the session.

<sup>11</sup>Participants were told that the 24 participants could be from the current session and/or a previous one.

## 2.2 Procedures

We conducted the experiments at GATE-Lab, Lyon, France. A total of 356 participants were recruited using HROOT (Bock et al., 2014), of which 89.9% were students from local engineering, business, and medical schools; the other participants were unemployed individuals and employees (see Table B.1 in the Appendix for a summary of the participants' characteristics). We conducted two sessions for each treatment, except for the TR-CF and PR-CF treatments, for which we conducted three sessions. More sessions were needed for these treatments because the independent observation is the pair and not the participant.<sup>12</sup> The experiment was programmed with z-Tree (Fischbacher, 2007). After completing the mood and risk elicitation, the participants directly received instructions for the die task on their computer screen (see Appendix A). 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.

## 3 Behavioral Conjectures

Recent literature on cheating games has emphasized the existence of a moral cost of lying even in the absence of scrutiny and negative externalities of cheating (*e.g.*, Fischbacher and Föllmi-Heusi, 2013; Kajackaite and Gneezy, 2017; Abeler et al., 2019). Theoretical models have been developed in which the decision to lie results from the comparison of the expected monetary benefit of lying and its expected moral costs due to an intrinsic preference for honesty and the reputational cost of being perceived as a potential cheater (Dufwenberg and Dufwenberg, 2018; Gneezy et al., 2018; Abeler et al., 2019; Khalmetski and Sliwka, 2019). This can explain why individuals tend to lie only partially when actions cannot be scrutinized by the experimenter or by other participants.<sup>13</sup>

In our experiment, scrutiny by the experimenter and by the counterpart is limited because the true outcomes of the die rolls can only be observed by the player, and the distribution of reports over time only gives a probability of whether the individual cheated or not. Hence, we expect a significant fraction of the participants to cheat but mostly partially to lower suspicion.<sup>14</sup> To illustrate this, consider the case of the piece rate pay

---

<sup>12</sup>Using G-power (3.1), we ran a post-hoc power analysis on all comparisons between treatments. This analysis reveals that we have an achieved power ranging from 0.22 to 0.29 for a small effect size (Cohen's  $d$  of 0.3), from 0.68 to 0.81 for a medium effect size (Cohen's  $d$  of 0.6) and from 0.95 to 0.99 for a large effect size (Cohen's  $d$  of 0.9).

<sup>13</sup>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. Contrastingly, when the true outcome is not observable, this number drops to 27%. This is an estimate that 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.

<sup>14</sup>Note that in our repeated environment, an individual cheats partially if he misreports the outcome of the die roll in some periods but does not cheat maximally in all periods. Our definition of partial cheating is a natural extension of the one used in one-shot cheating games if one considers the 24 periods together. Partial cheaters are individuals whom neither tell the truth nor maximize their payoff (Fischbacher and

scheme with final feedback. To simplify the analysis, we consider the decision to cheat in the 24 periods together,<sup>15</sup> and assume the following parameterized utility function with a convex lying cost that depends on the amount cheated:<sup>16</sup>

$$U_i = \pi(x_i + l_i) - \theta \cdot ([l_i]^+)^2 \quad (1)$$

where  $[l_i]^+ = \max(l_i, 0)$  and  $0 \leq \theta < \infty$ .  $x_i$  is the value obtained by rolling the die 24 times and adding up the numbers, while  $l_i$  represents the overall level of lying throughout the 24 die rolls. This can be any value between 0 (no lying) and  $48 - x_i$  (the agent reports only reds in all periods).  $\pi$  is the monetary reward that the agent receives from reporting  $x_i + l_i$ . As in previous studies on cheating, we assume that no participant misreports a die roll to his disadvantage.<sup>17</sup> This is why an individual experiences disutility only when  $l_i > 0$ .  $\theta$  captures an individual's aversion to lie. We assume that this aversion depends on how much the individual fears to be perceived as a liar, and this fear increases in the incentives to lie (see Kajackaite and Gneezy, 2017). That is, the higher the monetary stakes, the higher the fear of being perceived as a liar when reporting a high outcome. We also assume that the aversion to lying depends on the agent's beliefs regarding the social appropriateness of lying. It is straightforward to show that under the simplest piece-rate scheme in which there is no continuous feedback (*i.e.*,  $\pi = x_i + l_i$ ), the individual chooses  $l_i^* = \min(\lfloor \frac{1}{2\theta} \rfloor, 48 - x_i)$ , where the notation  $\lfloor \dots \rfloor$  identifies the nearest integer function. For  $0 \leq \theta < 0.022$ , the agent should, in expectation, always cheats to the full extent. For  $0.022 \leq \theta \leq 1$ , the agent should, in expectation, cheat partially. For  $\theta > 1$ , the individual should never cheat. Based on previous studies, we expect participants to cheat, for the most part, partially. Hence, we hypothesize  $\theta$  to be between 0.022 and 1.<sup>18</sup> We state our first conjecture as follows:

**Conjecture 1 (Cheating Behavior):** Most participants over-report their actual out-

---

Föllmi-Heusi, 2013) throughout the 24 rounds. For example, this definition applies to participants who *sometimes* or *all the time* report yellow instead of blue or to participants who *sometimes* report red instead of yellow.

<sup>15</sup>We model the individuals' behavior as if they pre-committed to their lying strategy in advance before playing the task. This is a bit of a stretch because they can adjust their behavior in each period depending on the history of die rolls and, in the case of continuous feedback, of the counterpart's plays. Modelling these dynamics is challenging and beyond the scope of this study. It is thus left for future research. Our theoretical framework should be taken only as an approximate indicator of how individuals are expected to behave, on average, in the experiment.

<sup>16</sup>For a similar parameterization see, *e.g.*, Dufwenberg and Dufwenberg (2018), Cartwright and Xue (2020) or Grundmann (2020).

<sup>17</sup>Previous studies also assume risk neutrality, and we follow this assumption. We measured risk aversion and control for it in our regression analysis.

<sup>18</sup>We can use the results of Dai et al. (2018) to provide a rough estimate of  $\theta$ . They use the exact same task as us on a similar student population but for only one period. We assumed that the utility is described by (1) and that  $\theta$  is drawn from a normal distribution, censored at 0, with mean  $\bar{\theta}$  and variance  $\sigma_{\theta}^2$ . We ran a simulation with 10,000 observations for different values of  $\bar{\theta}$  and  $\sigma_{\theta}^2$ . The values that best match the distribution of reports in Dai et al. (2018) (*i.e.*, those minimizing the sum of squared deviations from Dai et al.'s distribution) are  $\bar{\theta} = 0.7$  and  $\sigma_{\theta} = 0.4$ , respectively.

comes partially.

As payoffs depend on the ability to outperform the competitor, we expect that individuals who face competitive incentives are more likely to behave dishonestly than those who are paid an individual piece rate. In a tournament, the expected monetary reward of agent  $i$  is equal to  $2(x_i + l_i)P$ , where  $P$  is the probability that the agent wins against contestant  $j$ . We can write this probability as:

$$\begin{aligned} P &= \text{prob}(x_i + l_i > x_j + l_j) = \text{prob}(l_i - l_j > x_j - x_i) \\ &= \text{prob}(l_i - l_j > \xi) = G(l_i - l_j) \end{aligned}$$

where  $\xi = x_j - x_i$ ,  $\xi \sim g(\xi)$  with  $E(\xi) = 0$  and  $E(\xi^2) = 2\sigma_\xi^2$ , and  $G(\cdot)$  is the cumulative distribution function of  $g(\xi)$ .<sup>19</sup> Agent  $i$  maximizes his expected utility with respect to  $l_i$ . Assuming that each competitor chooses his optimal level of cheating by taking the choice of the rival as given, the reaction function of agent  $i$  is:

$$(l_i + x_i)g(x_j - x_i) + G(l_i - l_j) - \theta l_i = 0 \quad (2)$$

where  $g(l_i - l_j)$  is derived from  $\frac{\partial G(l_i - l_j)}{\partial l_i}$ . Assuming that both competitors have the same moral cost of lying, the reaction function of agent  $j$  is identical to (2), except that the subscripts  $i$  and  $j$  are reversed. The symmetry of the reaction functions implies that, in equilibrium,  $l_i = l_j$  and  $G(0) = 1/2$ , assuming a solution exists.<sup>20</sup> Additionally, in expectation,  $x_i = x_j = 24$ . Substituting all this into (2) and solving for  $l$  yields the optimal level of cheating in the tournament:

$$l^* = \min \left( \left\lfloor \frac{1 + 48g(0)}{2\theta - 2g(0)} \right\rfloor, 48 - x_i \right)$$

Since  $g(0) > 0$ , it is straightforward to see that the optimal level of cheating in the tournament is higher than the optimal level of cheating in the piece-rate scheme.<sup>21</sup>

Furthermore, we expect that when the opponent cannot cheat in the tournament, participants are less likely to over-report their outcome than when the opponent can cheat. To show this, note that, in expectation, individual  $i$  will always win the tournament and receive a higher piece rate if  $l_i > 0$  since  $x_i = x_j = 24$ . Hence, the optimal level of cheating maximizes  $U_i = 2(x_i + l_i) - \theta \cdot ([l_i]^+)^2$ , and is equal to  $\min(\lfloor \frac{1}{\theta} \rfloor, 48 - x_i)$ . This level of cheating is smaller than the optimal level of cheating under the two-sided tournament but larger than the optimal level of cheating under the piece-rate scheme, if  $\theta$  does not

<sup>19</sup> $g(\xi)$  approximates the normal distribution. This is because the probability distribution of the sum of several dice approaches the normal distribution as the number of die rolls increases (Chavan, 2014).

<sup>20</sup>In the Online Appendix, we show that this is the case for  $\theta > 0.2$ . Note that a solution typically exists if  $\sigma_\xi^2$  is sufficiently large (see Lazear and Rosen, 1981).

<sup>21</sup>It is also possible that participants believe that cheating is more common and less socially inappropriate in the tournament than under a piece-rate scheme, which implies a lower  $\theta$ . This may further increase cheating behavior in the tournament.

vary. However, we also expect  $\theta$  to increase in the one-sided tournament because stakes are higher than in the individual piece-rate scheme (the marginal benefit of cheating doubles). Thus, the individual may fear more to be perceived as a liar and, consequently, will reduce lying. Therefore, overall, we expect participants to roughly 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.<sup>22</sup> This would be in line with previous studies showing that the marginal benefit of cheating does not influence the willingness to cheat (see, *e.g.*, Fischbacher and Föllmi-Heusi, 2013; Kajaite and Gneezy, 2017; Abeler et al., 2019). Our second conjecture is as follows:

**Conjecture 2 (Competition):** (a) Participants cheat more under a tournament payment scheme than under an individual piece-rate scheme; (b) when their opponent is not allowed to cheat, they cheat less than in a standard tournament; and (c) they are equally likely to cheat in a one-sided tournament and under the piece rate scheme.

Previous studies have reported a contagion effect of dishonesty in settings comparable to our individual piece-rate environment (*e.g.*, Gino et al., 2009; Diekmann et al., 2015; Charroin et al., 2022). The mechanisms behind such peer effects may be emulation (or rivalry) and a preference for conformity that may be channeled by changes in beliefs (Rauhut, 2013): individuals who receive a signal about others' cheating behavior may update their beliefs about the appropriateness of cheating and adjust their behavior accordingly by imitating others.<sup>23</sup> In our theoretical framework, this can be captured by a change in  $\theta$ . Indeed, the moral cost of lying may be reduced (increased) when discovering that the counterpart is likely cheating (reporting honestly).<sup>24</sup> Such a contagion effect may be asymmetric, as bad examples have more influence than good ones (*e.g.*, Lefebvre et al., 2015; Dimant, 2019). A possible explanation is that most people tend to underestimate the extent of cheating (Diekmann et al., 2015). Assuming that this also applies to our setting (we provide evidence of this in a follow-up experiment), we expect participants who are paid an individual piece rate to lie more when they are continuously informed about their counterpart's reports rather than at the end.

---

<sup>22</sup>Based on the same logic,  $\theta$  should also increase in the two-sided tournament. However, the new  $\theta$  should be more than four times larger to make cheating in the two-sided tournament equal to cheating under the piece rate. We provide the proof of this in the Online Appendix.

<sup>23</sup>Note that information flows bidirectionally in our setting: players receive social information and share information on their behavior with the counterparts. Disentangling the effect of sharing *vs.* receiving social information on behavior 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 that of sharing.

<sup>24</sup>One could model the rivalry behind peer effects by assuming that the agent experiences additional utility from outperforming the opponent. Agent  $i$ 's utility function can, for example, be written as:  $U_i = \pi(x_i + l_i) + CF \cdot \alpha \cdot P - \theta \cdot ([l_i]^+)^2$ , where  $P$  is the probability of outperforming the opponent,  $CF$  is a dummy equal to 1 if feedback is continuous (and 0 otherwise), and  $\alpha$  is the marginal intrinsic benefit of being ahead. However, our results do not seem to support this interpretation of peer effects: as we will see later, when we introduce continuous *vs.* final feedback, cheating only increases under the piece-rate scheme and not under the tournament scheme.

Similar contagion effects channeled by changes in beliefs may be at work in the tournament setting with continuous feedback. If individuals underestimate (overestimate) the extent of cheating behavior under a tournament, they will cheat more (less) when they receive continuous feedback than when they receive final feedback, whereas they will not cheat more if their beliefs are accurate. This may result from both peer effects induced by conformity or emulation, as described for the piece-rate scheme, and from the innate strategic nature of tournaments. Indeed, unlike the individual piece-rate scheme, not mimicking the opponent's reporting behavior bears the risk of losing the tournament and earning nothing. Hence, we expect participants to stick more to their opponent's performance when they are continuously informed about this performance in the tournament than under the individual piece-rate scheme.

If the opponent cannot cheat, the signal individuals receive through their counterpart's reports is uninformative about the prevalence of cheating in the population. Thus, we expect no difference in the intensity of cheating between the one-sided tournaments with continuous and final feedback.<sup>25</sup> The continuous feedback only allows the participants to adjust the level of their cheating such that they are sure to outperform their opponent. Our third conjecture is as follows:

**Conjecture 3 (Feedback):** (a) Under a piece-rate pay scheme, individuals cheat more when they receive continuous rather than final feedback because of peer effects; (b) in a tournament, the difference between the two types of feedback depends on the players' beliefs about the prevalence of cheating in the population. Moreover, to avoid losing the tournament, when feedback is continuous, participants are more likely to adjust their level of cheating to the reports of their counterparts than individuals compensated by an individual piece-rate scheme; (c) in one-sided tournaments, both continuous feedback and final feedback have the same effect on misreporting.

Regarding the effect of feedback, our main experiment allows us to examine whether individuals cheat more when they receive continuous feedback rather than final feedback. However, we cannot identify whether this is because individuals underestimate more the extent of cheating behavior in the population. We explored this in a follow-up experiment in which we elicited the individuals' beliefs during the task.

## 4 Results of the Main Experiment

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

---

<sup>25</sup>This holds under the assumption that individuals pre-commit their cheating strategies, which may not be true, especially as the tournament progresses. Furthermore, it only holds for risk-neutral individuals. Risk-averse individuals might cheat more when feedback is final rather than continuous to increase their chance of winning against possibly very lucky opponents.

## 4.1 Cheating Behavior Across Treatments

Our first result is in line with previous studies:

**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	Passive player reports
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%	0.96 (0.04)
TR1-CF	18.5%	30.5%	51%	1.32 (0.05)	9.7%	1.02 (0.02)
Total	16.6%	27.3%	56.1%	1.38 (0.02)	9.4%	0.99 (0.02)

*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 two points, a yellow one pays one point, and a blue one pays zero 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 next column reports the percentage of participants who reported the maximum number of points (48 points) in the task (full extent cheating). The last two columns report the average and standard error of the reports from the passive players in the TR1-FF and TR1-CF treatments; sign-rank tests show that they do not significantly differ from 1. Passive participants who could not cheat in the TR1 treatments are excluded from the statistics reported in this table except in the last two columns.

**Support to Result 1.** For each treatment, Table 2 presents 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 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<sup>26</sup> comparing the actual and expected reports for each outcome show that participants significantly over-reported the red outcome and under-reported the blue outcome ( $p < 0.001$  in each treatment).<sup>27</sup> 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). Similar results are provided

<sup>26</sup>All non-parametric tests were 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 calls for an analysis at the pair level and prevents us from using binomial tests. In the FF treatments, binomial tests gave similar results to the Wilcoxon signed-rank test.

<sup>27</sup>All statistical analyses are performed at the pair level for the CF treatments and the participant level for the FF treatment. Participants who were passive in the TR1 treatments were excluded from the analysis. The last two columns of Table 2 display the passive players' reports, which do not differ from 1.



by Kolmogorov-Smirnov tests comparing the distribution of the reports with a uniform distribution ( $p < 0.001$  for all treatments).

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 outcome and the expected outcome show that participants over-reported significantly in all treatments ( $p < 0.001$  in all treatments). However, most participants did not 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. These percentages indicate that the participants did not lie in all periods.

To assess the occurrence of cheating throughout the experiment, we split the 24 periods in blocks of 4 periods.<sup>28</sup> 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 their outcome in the TR than in the PR treatments. **b)** They over-reported less of their outcome in the TR1 treatments than in the TR treatments. **c)** They did not over-report systematically more in TR1 than in the PR treatments.

**Support to Result 2.** Table 3 reports the  $p$ -values of a Dunn’s test.<sup>29</sup> 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 reported significantly more points in the pooled TR treatments than in the pooled PR treatments (Wilcoxon rank-sum test,  $p < 0.001$ ). This is in line with previous research showing that competitive incentives increase dishonesty.

Table 3 also indicates that the values reported are 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 also the case when

---

<sup>28</sup>Figure C.2 in 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 two and six periods, and the results were qualitatively the same. Furthermore, we report in Figure C.3 in Appendix C the average reported value in each period, by treatment, and Table B.3 in Appendix B reports the  $p$ -values from the sign-rank tests conducted on each period in each treatment.

<sup>29</sup>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.

comparing 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 behaved more honestly in a tournament when they faced an opponent who could not cheat, as predicted by Conjecture 2b. First, lying a little is sufficient to outperform his opponent; second, a passive opponent's reports are not informative on the prevailing honesty norm.

Table 3: Dunn's Test: Pairwise Treatment Comparisons of Reports

Treatments	PR-FF	PR-CF	TR-FF	TR-CF	TR1-FF
PR-CF	<b>0.035</b>	-	-	-	-
TR-FF	<b>&lt;0.001</b>	<b>0.008</b>	-	-	-
TR-CF	<b>&lt;0.001</b>	<b>0.019</b>	0.383	-	-
TR1-FF	<b>0.068</b>	0.419	<b>0.006</b>	<b>0.015</b>	-
TR1-CF	<b>0.062</b>	0.424	<b>0.006</b>	<b>0.014</b>	0.494

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

The test reported in Table 3 also shows 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, we find no significant difference between the reports in the pooled PR treatments and the pooled TR1 treatments (Wilcoxon rank-sum test,  $p = 0.282$ ). The overall intensity of lying in a piece-rate scheme is similar to its intensity 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 feedback frequency on misreporting.

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

**Support to Result 3.** In the PR treatments, participants behaved more dishonestly when they received continuous 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. Contrastingly, 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:  $p = 0.494$ ).<sup>30</sup> For the TR1 treatments, this result sup-

<sup>30</sup>One may argue that the lack of difference between the TR-FF and TR-CF treatments could be due to a ceiling effect because the levels of cheating are already high in these two treatments. However, in

ports Conjecture 3c.<sup>31</sup> It also indicates that repeatedly observing a counterpart that does not cheat when honesty does not result from a personal choice does not make individuals more honest. To study the role that beliefs about others' cheating behavior play in the PR and TR treatments, we use our follow-up experiment that is introduced 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 run during the same months.<sup>32</sup> The marginal effects are reported in Table B.4 in Appendix B. Table 4 reports the results of multiple pairwise comparisons of the marginal effects across 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 blue outcome in the TR treatments than in the PR treatments. Reports differed between the TR-FF and TR1 treatments, and they were not sensitive to feedback in the TR and TR1 treatments. These results are consistent with those obtained from the non-parametric tests. However, reports did not differ significantly between the TR-CF and TR1 treatments, although the sign of the coefficients is in line with the non-parametric tests. Additionally, only reports of the blue outcome differed between PR-FF and PR-CF treatments. Overall, these results support Conjectures 2a, 2b, and 3c; they only partially support Conjectures 2c (as reports in TR1 sometimes differ from those in PR) and 3a (as only the reports of a blue outcome differ between PR-FF and PR-CF). Finally, we report in the Appendix further analyses exploring the heterogeneity of the effects of incentives and social information on behavior.<sup>33</sup>

## 4.2 Dynamics of Cheating Behavior

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

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

---

period 24 in the TR-FF and TR-CF treatments, only 68.18% and 64.10% of the participants, respectively, reported a red outcome, suggesting that there was still room for more cheating for a significant portion of the participants.

<sup>31</sup>Indeed, continuous feedback adds little information value compared to final feedback for risk-neutral individuals. However, we obtain the same findings when controlling for risk attitudes in the regressions presented below.

<sup>32</sup>This variable controls for the fact that the sessions were run during three distinct months.

<sup>33</sup>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 manipulation on happiness and nervousness. We found that participants were less happy and more nervous in the TR treatments, possibly because competition induced more stress.

Table 4: Pairwise Comparisons of Treatments for All Outcomes

Blue outcome (0 points)					
	PR-FF	PR-CF	TR-FF	TR-CF	TR1-FF
PR-CF	<b>-0.048**</b> (0.025)	-	-	-	-
TR-FF	<b>-0.113***</b> (0.026)	<b>-0.065***</b> (0.024)	-	-	-
TR-CF	<b>-0.094***</b> (0.026)	<b>-0.046*</b> (0.024)	0.019 (0.025)	-	-
TR1-FF	-0.044 (0.032)	0.005 (0.031)	<b>0.070**</b> (0.031)	0.051 (0.031)	-
TR1-CF	<b>-0.061**</b> (0.027)	-0.013 (0.026)	<b>0.052**</b> (0.026)	0.033 (0.026)	-0.018 (0.029)
Yellow outcome (1 point)					
PR-CF	0.020 (0.027)	-	-	-	-
TR-FF	<b>-0.084***</b> (0.030)	<b>-0.104***</b> (0.031)	-	-	-
TR-CF	<b>-0.059**</b> (0.029)	<b>-0.079***</b> (0.029)	0.025 (0.032)	-	-
TR1-FF	-0.044 (0.029)	<b>-0.064**</b> (0.031)	0.040 (0.032)	0.015 (0.032)	-
TR1-CF	-0.018 (0.032)	-0.038 (0.034)	<b>0.066*</b> (0.034)	0.041 (0.035)	0.026 (0.032)
Red outcome (2 points)					
PR-CF	0.028 (0.043)	-	-	-	-
TR-FF	<b>0.198***</b> (0.048)	<b>0.169***</b> (0.050)	-	-	-
TR-CF	<b>0.153***</b> (0.046)	<b>0.125***</b> (0.048)	-0.044 (0.052)	-	-
TR1-FF	<b>0.088*</b> (0.050)	0.059 (0.055)	<b>-0.110**</b> (0.056)	-0.066 (0.055)	-
TR1-CF	<b>0.079*</b> (0.047)	0.051 (0.052)	<b>-0.118**</b> (0.053)	-0.074 (0.052)	-0.008 (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.4 in 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.

**Support to Result 4.** Figure 1 reports the marginal effects of the period variable on the likelihood of reporting a blue, yellow, or red outcome in each treatment. Estimates are computed from a multinomial logit regression model similar to that reported in Table B.4 in Appendix B, but with the inclusion of an interaction term between each treatment dummy and the time trend variable. In both the PR-CF and TR-CF treatments, we observe a significant increase in 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 group ( $p = 0.001$ ). This reveals an escalation effect in the presence of social information.<sup>34</sup> Not surprisingly, in the TR1-FF and TR1-CF treatments 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 outcomes, respectively).

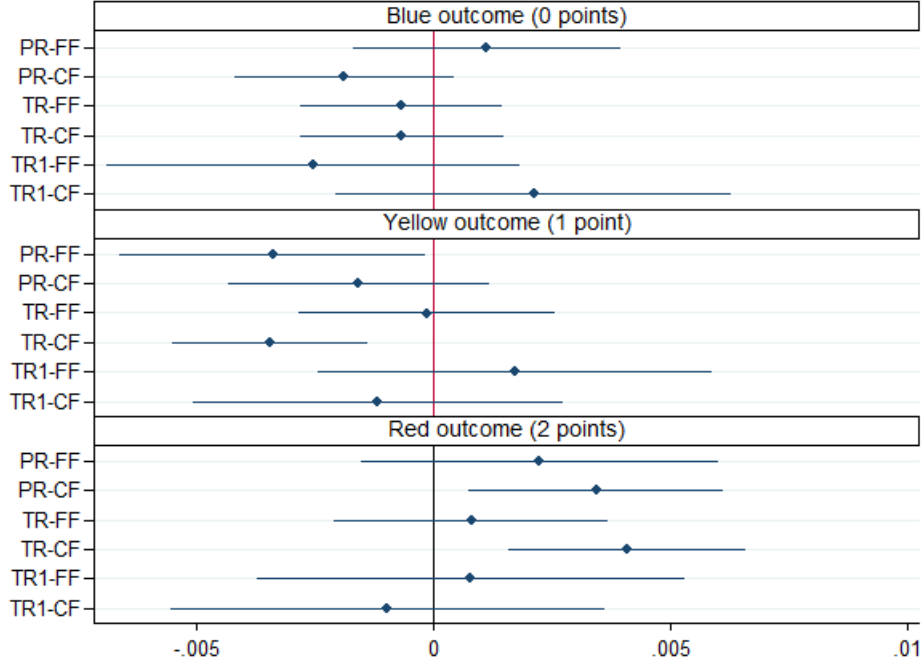


Figure 1: Marginal Effects of Time on the Probability to Report Each Outcome, by Treatment.

*Notes:* The figure displays the marginal effects of the period variable on the likelihood of reporting each possible outcome in each treatment. Estimates are computed using a multinomial logit regression model. The horizontal lines represent 95% confidence intervals.

To support the idea that the escalation of cheating in the PR-CF and TR-CF treatments over time results from social contagion,<sup>35</sup> we explore further peer effects, by following the strategy used by Falk and Ichino (2006) and replicated in Rosaz et al. (2016). We calculate the within-pair and between-pair standard deviations (S.D.) 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 FF treatments allows us to determine whether social interactions have a significant impact on the PR-CF and TR-CF treatments. If social interactions generate conformity 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.

<sup>34</sup>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.

<sup>35</sup>One might alternatively argue that this escalation reflects a decrease of the moral cost of cheating over repetitions, regardless of peer effects (e.g., Garrett et al., 2016). However, such a decrease in the moral cost of lying over time should be independent of the presence of continuous feedback. Our results do not support this interpretation since they show no significant effect of time in the FF treatments, except for PR-FF, as discussed in footnote 34.

Finally, the difference between the between-pair S.D. and 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 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 ( $p < 0.001$ ).

We used the actual pairs of participants to calculate the standard deviations in the PR-FF and TR-FF treatments. As a robustness test, we also ran a simulation-based approach.<sup>36</sup> The results of this simulation are displayed in Figure C.4 in Appendix C; they confirm the conclusions of the non-parametric tests.<sup>37</sup> 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.5 in Appendix C). We also compare the difference between the between-pair S.D. and 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 was significantly larger for the pairs in the CF treatments than for the simulated pairs in the FF treatments (PR treatments:  $p < 0.001$ ; TR treatments:  $p < 0.001$ ) (see also Figure C.6 in Appendix C).

Overall, these analyses of within- and between-pair standard deviations support the existence of peer effects induced by 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.

Finally, we test whether the effect of feedback over time is larger in the TR-CF treatment than in the PR-CF treatment (Conjecture 3b). This brings us to our last result.

**Result 5** (*Treatment Differences in Escalation*): The contagiousness of dishonesty was not significantly higher 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 that the within-pair average standard deviation differed between the two treatments ( $p = 0.217$ ).

---

<sup>36</sup>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.

<sup>37</sup>Not surprisingly, we found no difference between the within- and between-pair S.D. in the TR1-FF and TR1-CF treatments.

Thus, we reject Conjecture 3b.

## 5 A Follow-up Experiment to Identify the Role of Beliefs

### 5.1 Design and Procedures

The main experiment revealed that participants cheated more when they received continuous rather than final feedback under piece-rate incentives, but not in tournaments. We argue that this is because uninformed participants underestimated the extent of cheating behavior more under a piece-rate pay scheme than in a tournament where the anticipation of the opponent’s behavior has strategic value. We ran a second experiment with 102 new participants from the same pool of subjects to test this interpretation. The experiment comprised of two conditions: the belief-PR-CF and belief-TR-CF treatments (bPR-CF and bTR-CF, hereafter), involving 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 randomly selected participants from past sessions reported a red, a yellow and a blue outcome in a given period. At the end of the experiment, the program randomly drew one period and one color for payment. Participants received €1.5 for a correct prediction, €1 if their prediction was incorrect by  $\pm 1$ , €0.5 if it was incorrect by  $\pm 2$ , and €0 otherwise. All the other procedures were equivalent to those used in the main experiment.<sup>38</sup>

### 5.2 Results

As in the main experiment, we find that participants over-reported the die rolls.<sup>39</sup> On average, participants over-reported more in the bTR-CF treatment than in the bPR-CF one. 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 others and 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 treatments with behavior in the bPR-CF and bTR-CF treatments shows that participants

<sup>38</sup>See Appendix A for the instructions.

<sup>39</sup>Pooling all periods, 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). Additionally, they 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.

reported significantly fewer reds (58% *vs.* 64%,  $p = 0.091$ ) and more yellows (27% *vs.* 22%,  $p = 0.027$ ) in the main experiment compared to the follow-up experiment.<sup>40</sup> We also replicate the dynamics that we observe in the main treatment under continuous feedback.<sup>41</sup>

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. Particularly, we can compute to what 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 the bPR-CF (bTR-CF) treatment. The difference is marginally statistically significant (Wilcoxon rank-sum test,  $p = 0.055$ ). More precisely, participants in the bPR-CF treatment underestimated the proportion of reds in period 1 by 19.39%, whereas participants in bTR-CF treatment overestimated it by 3.69% ( $p = 0.011$ ). Furthermore, participants in bPR-CF (bTR-CF) treatment 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%) than participants in the bTR-CF treatment (by 23.26%), but the difference is not significant ( $p = 0.571$ ). These findings indicate that before receiving any feedback, the participants who were paid a piece rate underestimated the extent of cheating more than participants who competed in a tournament. This supports our explanation for Result 3: compared to final feedback, the effect of continuous feedback on dishonesty is positive and significant only if individuals initially underestimate the extent of cheating.

In Appendix D.4, we also look at how participants updated their beliefs over time once they began to receive information about their counterpart’s behavior and whether these beliefs impacted behavior. We find that participants expected a higher (lower) average report in the group of 24 if they previously underestimated (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 because of potential endogeneity issues.

## 6 Discussion and Conclusion

This study investigated how feedback about others’ reported outcomes influences individuals’ cheating behavior in 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 the anticipation of an unfair competition due to the opponent’s dishonest reporting.

---

<sup>40</sup>The difference is larger under piece-rate incentives (red: 51% *vs.* 61%,  $p = 0.067$ ; yellow: 31% *vs.* 25%,  $p = 0.026$ ).

<sup>41</sup>The likelihood of over-reporting 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 Appendix D.3).



The first important result is that competitive incentives make individuals more likely to cheat than an individual piece-rate pay scheme. This replicates in a new dynamic setting previous findings obtained in studies using real-effort tasks (*e.g.*, Schwierien and Weichselbaumer, 2010; Belot and Schröder, 2013; Faravelli et al., 2015; Rigdon and D’Esterre, 2015; Dutcher et al., 2016). We also found that individuals cheat less in tournaments when it is common knowledge that their opponents cannot misreport. Cheating differs only weakly in this context compared to the 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 continuous feedback on their counterpart’s reports, became more dishonest over time. This is in line with the idea of the behavioral contagiousness of dishonesty (*e.g.*, Gino et al., 2009; Robert and Arnab, 2013; Lauer and Untertrifaller, 2019; Charroin et al., 2022). 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, peer effects may be driven by pure conformity with a dishonesty norm, and rivalry in terms of status attached to relative ranking.

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

Our study provides new evidence on the role of feedback on relative outputs in cheating behavior under competitive and non-competitive payment schemes. It shows that competitive incentives, compared to non-competitive incentives, encourage dishonesty when there is uncertainty about the opponents’ degree of dishonesty and thus, about the actual fierceness of the competition. It also stresses the importance of individuals’ expectations in driving cheating behavior. In non-competitive settings, providing relative feedback generates peer effects with deleterious effects in terms of cheating. Future studies could investigate the exact nature of the effect of social information in non-competitive settings to disentangle the role of conformity to norms and rivalry. A related research question is whether the fierceness of the competition is something that is only relevant when individuals compete for money or whether it also drives dishonesty when people compete for status. For example, one could compare two PR conditions in which either both participants 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, particularly through the perception of what is socially acceptable.

Although one needs to remain cautious before extrapolating evidence from the lab, our findings have three main implications. First, market instruments that emphasize com-

petitiveness and continuous comparisons with peers should be introduced with care in organizations when agents have some discretion in reporting, as both favor the contagion of dishonesty. Second, our results show that when it is made clear to the competitors that their opponent cannot lie, there is less cheating, even if the individual himself is not monitored. This suggests that to avoid being accused of being responsible for the erosion of morals, competitive incentives *per se* have to be accompanied by policies that explicitly guarantee the actual fairness of the competition. Third, peer effects driven by the willingness to conform to the norm and/or by rivalry may be as responsible as incentive schemes for the erosion of morals in the society. Indeed, in our experiment, individuals cheat even in the absence of competitive monetary incentives. This pleads for inviting social norms and personal responsibility into the debate on the interactions between the market and morality.

From the general perspective of the effect of markets on morals, what can be derived from our results is that whether markets erode moral values depends on the specifics of the market institutions at play. In the context of employment relationships, the market provides several instruments and institutional arrangements that companies can adopt to attract certain types of employees and to incentivize the work of their employees. Our results suggest that the erosion of moral values in these contexts depends on how these different market instruments and institutional arrangements are combined together and on how individuals react to them. One limit of this analysis is that we only focused on one particular dimension of morality, namely honesty, and on two instruments. Future studies could explore the effect of these workplace incentives and arrangements on other moral preferences, such as trust, reciprocity, fairness, and generosity. Finally, another important extension could investigate how other market instruments, namely those used to support cooperation, also influence morality and its evolution.

## References

- Abeler, J., D. Nosenzo, and C. Raymond (2019). Preferences for Truth-Telling. *Econometrica* 87(4), 1115–1153.
- Al-Ubaydli, O., D. Houser, J. Nye, M. P. Paganelli, and X. S. Pan (2013). The causal effect of market priming on trust: An experimental investigation using randomized control. *PloS one* 8(3), e55968.
- Association of Certified Fraud Examiners (2020). Report to the nations – 2020 global study on occupational fraud and abuse. Technical report, Association of Certified Fraud Examiners – <https://acfepublic.s3-us-west-2.amazonaws.com/2020-Report-to-the-Nations.pdf>.
- 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.
- Bartling, B., E. Fehr, and Y. Özdemir (2021, 02). Does Market Interaction Erode Moral Values? *The Review of Economics and Statistics*, 1–32.
- 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.
- Berger, J., C. Harbring, and D. Sliwka (2013). Performance appraisals and the impact of forced distribution—an experimental investigation. *Management Science* 59(101), 54–68.
- Bock, O., I. Baetge, and A. Nicklisch (2014). Hroot: Hamburg Registration and Organization Online Tool. *European Economic Review* 71, 117–120.
- Bowles, S. (2011). Is liberal society a parasite on tradition? *Philosophy & Public Affairs* 39(1), 46–81.
- Bowles, S. (2016). *The Moral Economy: Why good incentives are no substitute for good citizens*. Yale University Press.
- Bruni, L. and R. Sugden (2013). Reclaiming virtue ethics for economics. *Journal of Economic Perspectives* 27(4), 141–64.
- 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.
- Cartwright, E. and L. Xue (2020). Lie aversion, anonymity and social distance: are people more willing to lie using a mobile phone? *SSRN Working Paper (available at <https://ssrn.com/abstract=3598023>)*.
- 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 (2022). Homophily, peer effects and dishonesty. *Journal of Economic Behavior & Organization* 200, 618–637.
- Chavan, R. R. (2014). The probability distribution of the sum of several dice approach to a normal distribution. *International Journal of Scientific & Engineering Research* 5(12).

- Choi, G. S. and V. H. Storr (2020). Market interactions, trust and reciprocity. *PloS one* 15(5), e0232704.
- Conrads, J., B. Irlenbusch, R. M. Rilke, and G. Walkowitz (2013). Lying and team incentives. *Journal of Economic Psychology* 34, 1–7.
- 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, 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. (2019). Contagion of pro- and anti-social behavior among peers and the role of social proximity. *Journal of Economic Psychology* 73, 66–88.
- Dufwenberg, M. and M. A. Dufwenberg (2018). Lies in disguise – A theoretical analysis of cheating. *Journal of Economic Theory* 175(175), 248–264.
- Dutcher, E. G., D. Glätzle-Rützler, and D. Ryvkin (2016). Don’t hate the player, hate the game: Uncovering the foundations of cheating in contests. Technical report, Working Papers in Economics and Statistics.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- Grundmann, S. (2020). Do just deserts and competition shape patterns of cheating? Technical report, Passauer Diskussionspapiere - Volkswirtschaftliche Reihe No. V-79-20.
- Harbring, C. and B. Irlenbusch (2011). Sabotage in Tournaments: Evidence from a Laboratory Experiment. *Management Science* 57(4), 611–627.
- Hirschman, A. O. (1977). *The Passions and the Interests: Political Arguments for Capitalism before Its Triumph*. Princeton University Press.
- 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.
- 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.
- Lazear, E. P. and M. Gibbs (2014). *Personnel Economics in Practice, 3rd edition*. Wiley.
- Lazear, E. P. and S. Rosen (1981). Rank-order tournaments as optimum labor contracts. *Journal of political Economy* 89(5), 841–864.
- 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.
- Marx, K. (1867). *Capital: A critique of political economy*. Duke University Press.
- McCloskey, D. N. (2006). *The Bourgeois Virtues: Ethics for an Age of Commerce*. University of Chicago Press.
- Montesquieu, C. d. (1748). *The spirit of the laws*. Cambridge University Press.
- 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.
- 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.
- Sandel, M. J. (2012). *What money can’t buy: the moral limits of markets*. Macmillan.

- Satz, D. (2010). *Why some things should not be for sale: The moral limits of markets*. Oxford University Press.
- Schumpeter, J. A. (1942). *Capitalism, socialism and democracy*. Routledge.
- 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.
- Shleifer, A. (2004). Does competition destroy ethical behavior? *American Economic Review* 94(2), 414–418.
- Smith, A. (1776). *An inquiry into the nature and causes of the Wealth of Nations*. Clarendon Press.
- 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.
- Storr, V. H. and G. S. Choi (2019). *Do markets corrupt our morals?* Palgrave Macmillan.
- Sutter, M., J. Huber, M. Kirchler, M. Stefan, and M. Walzl (2020). Where to look for the morals in markets? *Experimental Economics* 23(1), 30–52.
- 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.*

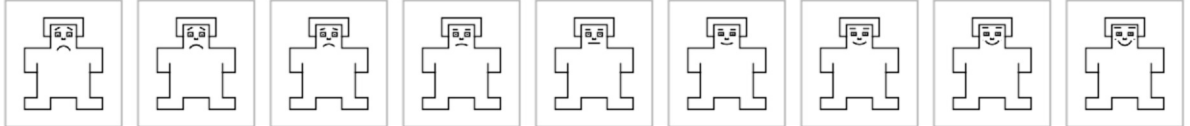
### A.1 Original treatments

**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.

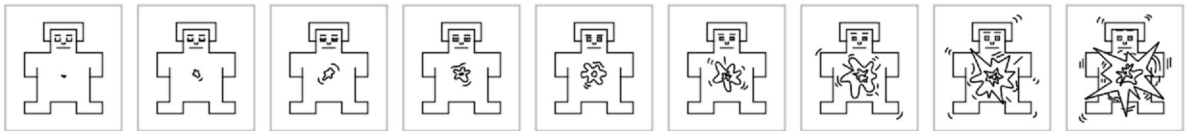
The experiment is divided into two independent parts. In addition, you will be asked some questions at the end of the experiment. At the beginning of each part, you will receive the corresponding instructions. At the end of the experiment, you will receive the earnings for the first and second part as 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.

## A.2 Follow-up treatments

The instructions of the bPR-CF and bTR-CF treatments are similar to the instructions for the PR-CF and TR-CF treatments, respectively. The only changes concern the Predictions part at the end of the instructions of the original treatments:

### Predictions

At the beginning of each period, before performing the reporting task, you will have to predict the colors that 24 other participants reported during the period. These 24 participants are randomly selected by the computer from previous sessions of the same experiment. In these sessions, the participants did not have to predict the behavior of others participants. They only had to perform the reporting task. Here is the procedure:

For each period, 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).

- 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.

## 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: P-values of Sign-Rank Tests in Every Period, by Treatment.

Period	1	2	3	4	5	6	7	8	9	10	11	12
PR_FF	0,002	0,028	0,068	0,004	0,590	0,433	0,068	0,061	0,004	1,000	0,549	0,095
PR_CF	0,083	0,007	0,004	<0.001	0,004	0,018	<0.001	<0.001	<0.001	0,004	0,004	<0.001
TR_FF	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0,001	0,001
TR_CF	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
TR1_FF	0,237	0,555	0,078	0,423	0,093	0,139	0,546	0,170	0,011	0,884	0,879	0,555
TR1_CF	0,065	0,024	0,283	0,298	0,824	0,723	0,002	0,161	0,183	0,696	0,071	<0.001
Period	13	14	15	16	17	18	19	20	21	22	23	24
PR_FF	0,353	0,003	0,257	0,059	0,353	0,014	0,724	0,003	0,004	0,117	0,056	0,011
PR_CF	0,059	<0.001	0,020	0,019	<0.001	<0.001	0,001	<0.001	0,001	0,021	<0.001	<0.001
TR_FF	<0.001	<0.001	<0.001	0,006	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
TR_CF	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
TR1_FF	0,343	0,343	0,879	0,009	0,011	0,005	0,004	0,873	0,114	0,042	0,013	0,423
TR1_CF	0,065	0,359	0,139	0,151	0,038	0,08	0,151	0,094	0,363	0,001	0,055	0,021

*Notes:* The table reports the p-values of sign-rank tests comparing the average reported values with the expected values in every period and all treatments. The average is computed at the pair-level for the FF treatments and at the individual level of the CF treatments.

Table B.4: Determinants of the Die Roll Report

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

Part 2

Period 1 out of 24

Your number of points obtained in each period:

Period	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Points																								

Your total number of points accumulated in past periods: 0

Number of points obtained by your co-participant in each period:

Period	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Points																								

Total number of points accumulated by your co-participant in past periods: 0 points

Please now shake the cup to roll the die and report the outcome. Validate by pressing OK.  
Blue pays 0 point, Yellow pays 1 point, Red pays 2 points

My outcome:

☐ Blue  
☐ Yellow  
☐ Red

OK

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

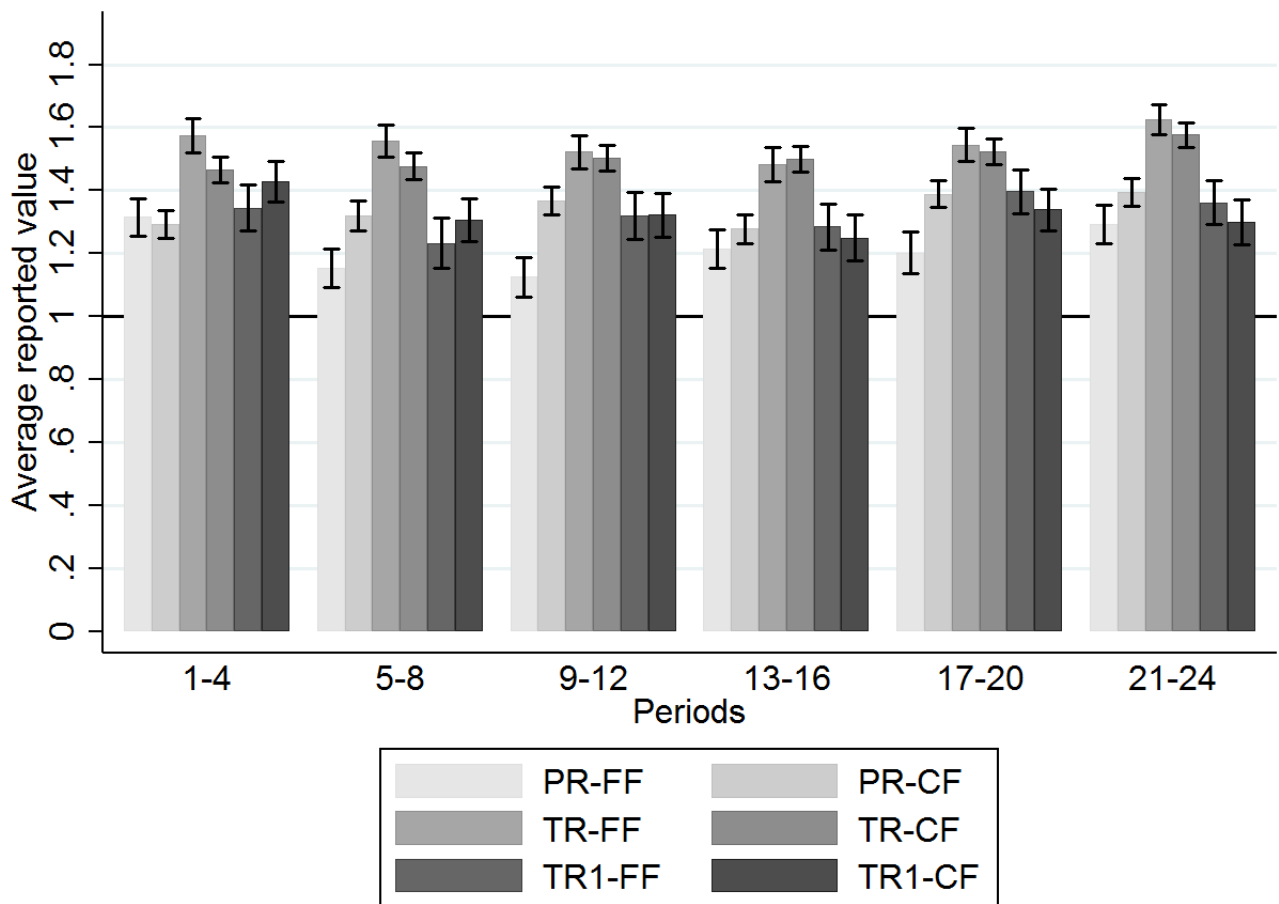


Figure C.2: Average Reported value by Treatments and Blocks of Four Periods

*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).

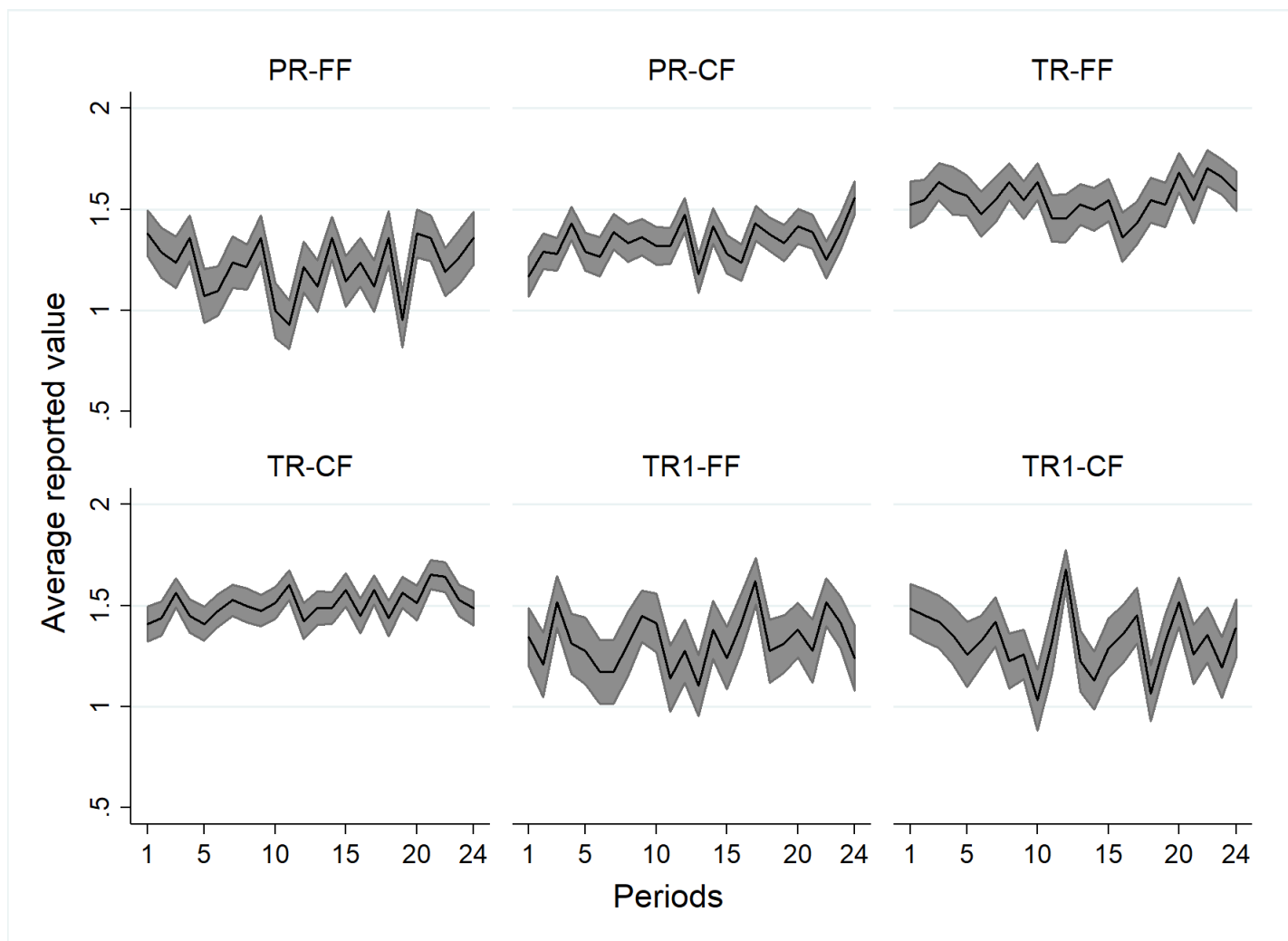


Figure C.3: Average Reported Value over Periods, by Treatment. *Notes:* The shaded areas represent standard errors of the mean.

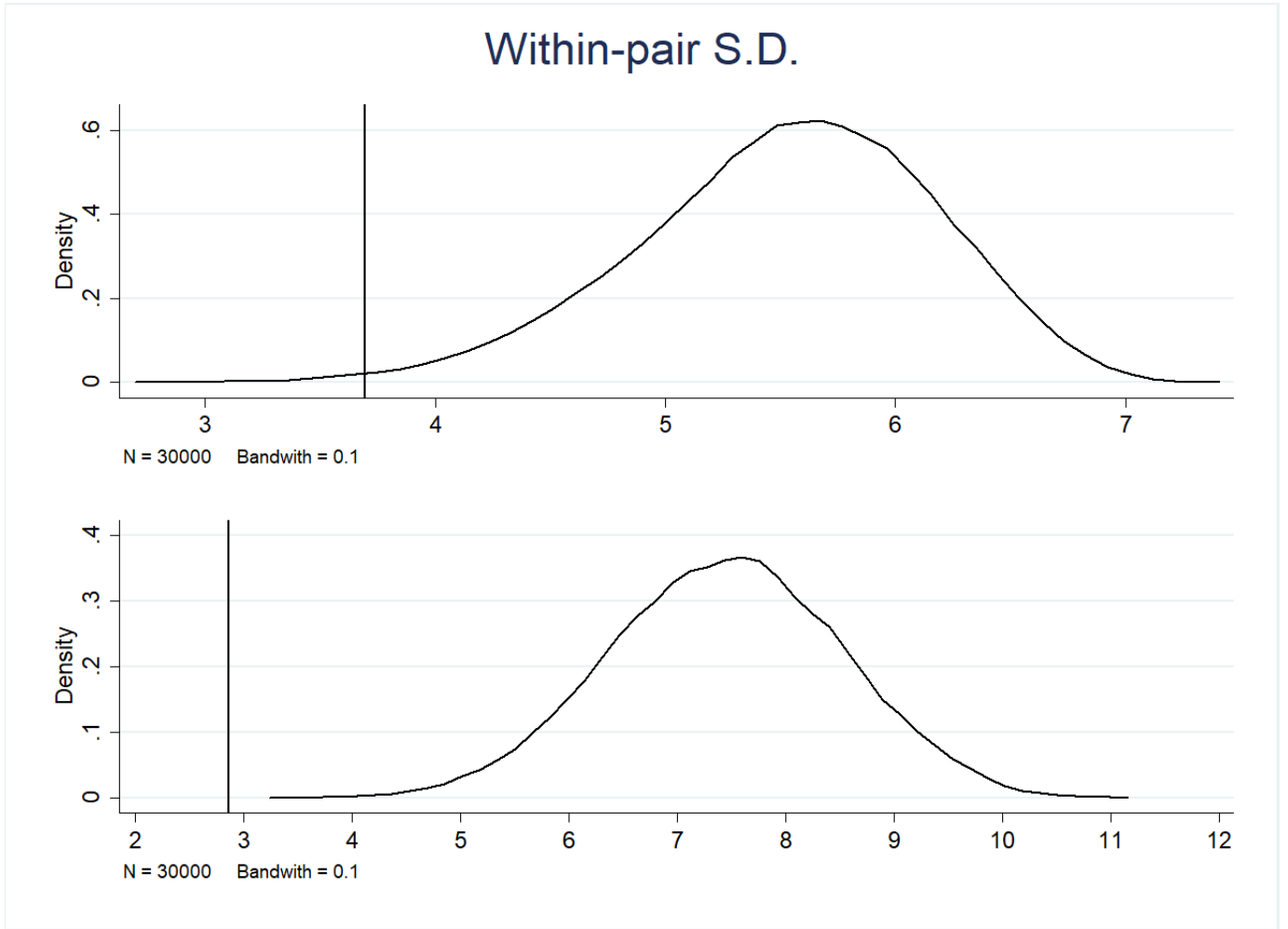


Figure C.4: **Top panel:** Within-Pair Standard Deviation in Reported Value in the PR-CF and PR-FF Treatments. The vertical line represents the average within-pair S.D. in the PR-CF treatment while the curve represents the within-pair S.D. density of the simulated pairs in the PR-FF treatment. **Bottom 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..



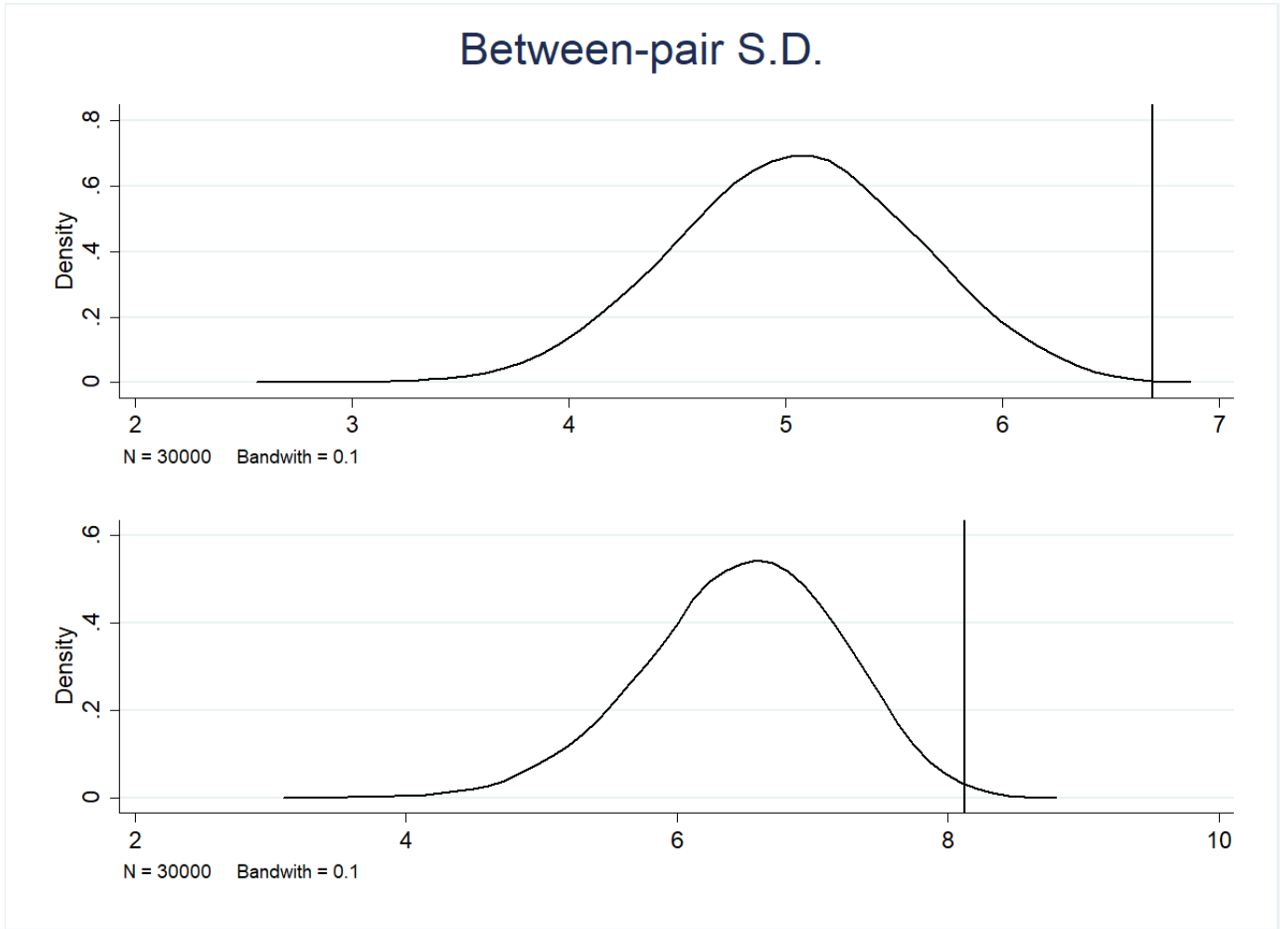


Figure C.5: **Top panel:** Between-Pair Standard Deviation in Reported Value in the PR-CF and PR-FF Treatments. The vertical line represents the average between-pair S.D. in the PR-CF treatment while the curve represents the between-pair S.D. density of the simulated pairs in the PR-FF treatment. **Bottom 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.

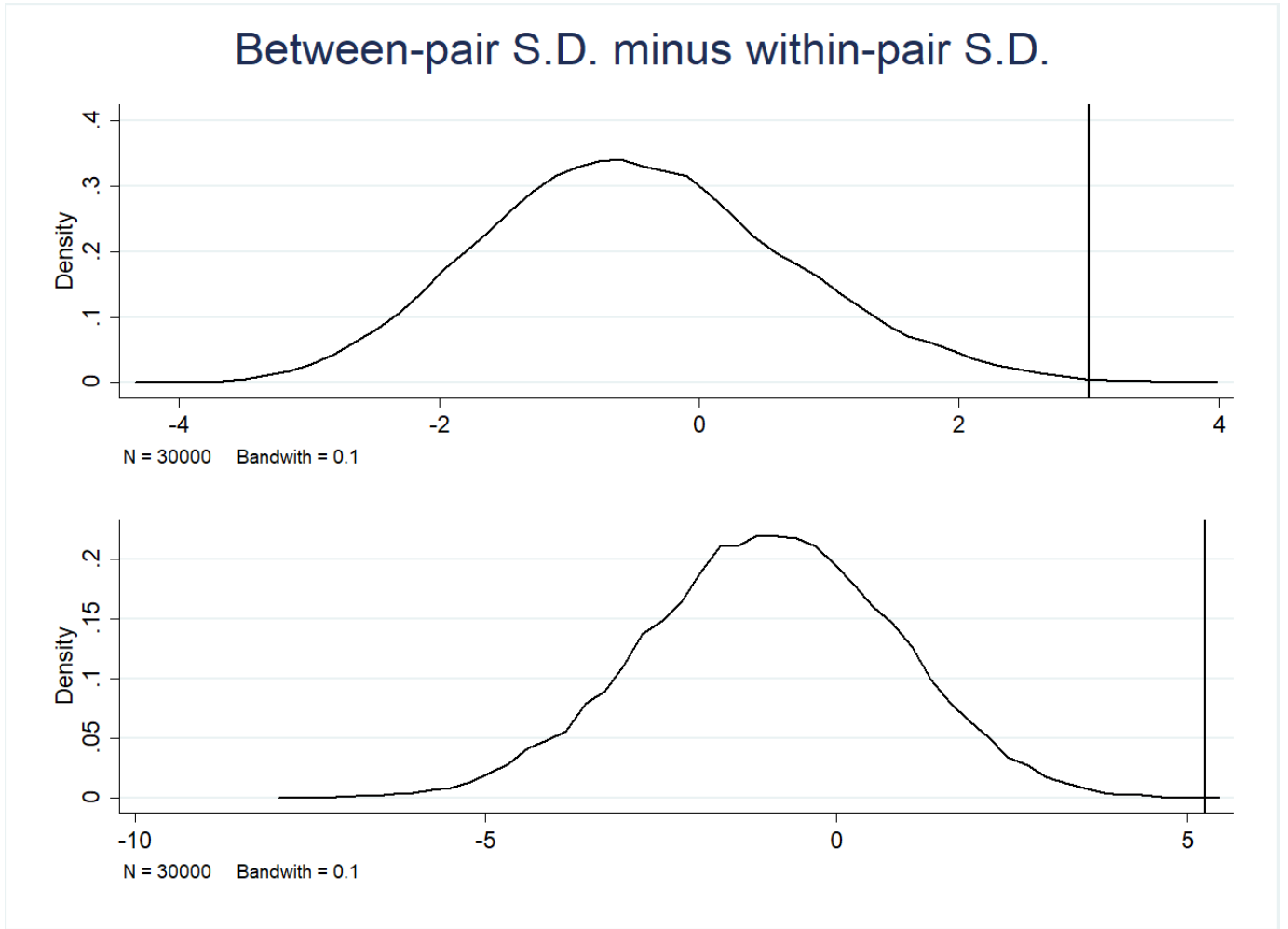


Figure C.6: **Top panel:** Between-Pair Minus Within-Pair Standard Deviation in the PR-CF and PR-FF Treatments. The vertical line represents the average between-pair minus within-pair S.D. in the PR-CF treatment while the curve represents the density of the difference between the between pair and the 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.4, but with the inclusion of an interaction term 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 ( $p < 0.001$ ).<sup>42</sup> 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.<sup>43</sup>

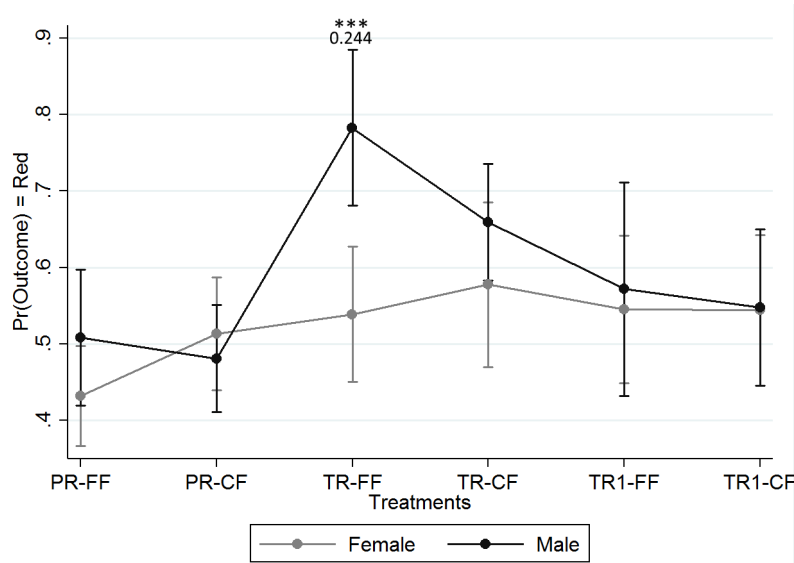


Figure D.1: Probability of Reporting a Red Outcome, by Gender and Treatment

*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 ( $\rho = 0.530, p <$

<sup>42</sup>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).

<sup>43</sup>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:  $1.125$ ,  $p < 0.001$ ).<sup>44</sup> 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 participants, happiness slightly decreased (by  $0.133$ ). A possible explanation is that participants in bPR-CF had a stronger desire to outperform the counterpart than participants in PR-CF, and this is because they were constantly reminded to think about the others in the belief elicitation.

---

<sup>44</sup>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.4 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	-0.026 (0.030)	-0.043 (0.036)	0.068 (0.060)
Period	-0.001 (<0.001)	-0.004*** (<0.001)	0.005*** (<0.001)
Other controls	Yes		
Observations	2448		
Number of clusters	51		
Pseudo-R <sup>2</sup>	0.020		
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:  $p = 0.543$ ).<sup>45</sup> 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:

$$\Delta B_i^t = \beta_0 + \beta_1 \times t + \beta_2 \times \text{bTR-CF} + \beta_3 \times t \times \text{bTR-CF} \quad (3)$$

$$+ (\beta_4 + \beta_5 \times \text{bTR-CF}) \times \left( \max \left\{ 0, \frac{\sum_{n=1}^{t-1} R_j^n}{t} \right\} - B_i^{t-1} \right) \quad (4)$$

$$+ (\beta_6 + \beta_7 \times \text{bTR-CF}) \times \left( \max \left\{ 0, B_i^{t-1} - \frac{\sum_{n=1}^{t-1} R_j^n}{t} \right\} \right) \quad (5)$$

$$+ (\beta_8 + \beta_9 \times \text{bTR-CF}) \times \left( \max \left\{ 0, \frac{\sum_{n=1}^{t-1} R_i^n}{t} - B_i^{t-1} \right\} \right) \quad (6)$$

<sup>45</sup>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.

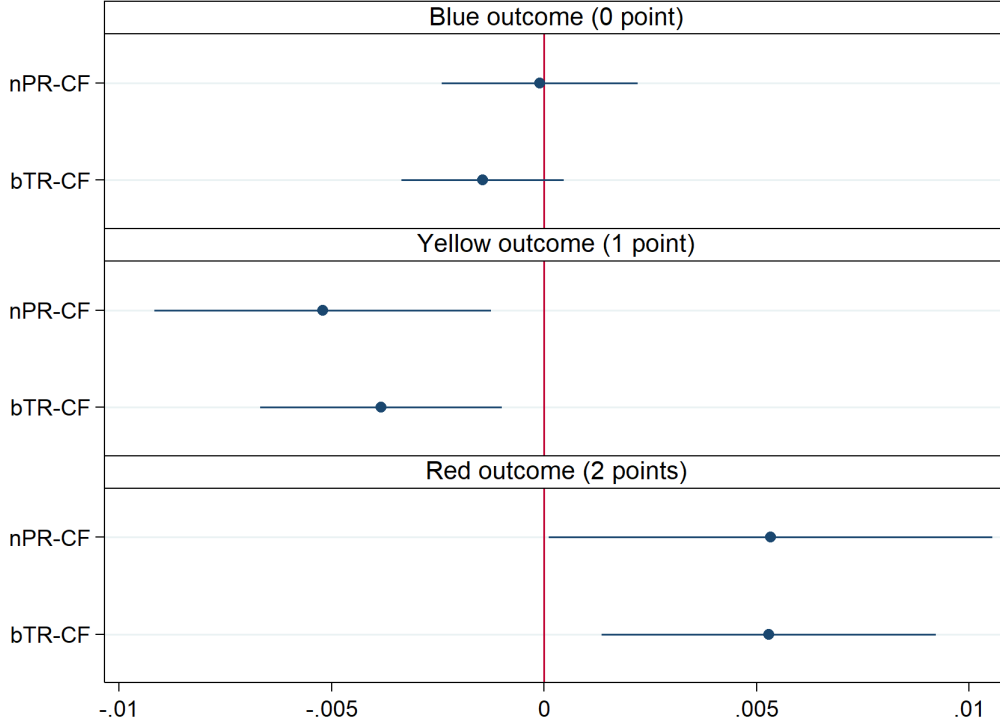


Figure D.2: Marginal effects of time on the probability to report each outcome, by treatment. *Notes:* The horizontal lines represent 95% confidence intervals.

$$+(\beta_{10} + \beta_{11} \times \text{bTR-CF}) \times \left( \max \left\{ 0, B_i^{t-1} - \frac{\sum_{n=1}^{t-1} R_j^n}{t} \right\} \right) \quad (7)$$

The dependent variable ( $\Delta B_i^t$ ) is the difference in the average belief of subject  $i$  between periods  $t$  and  $t-1$ . It is positive (negative) if a participant updated his 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 empirical 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.<sup>46</sup> In particular, we expect participants to adjust their beliefs upward (downward), if they have over-reported (under-reported) in the previous periods compared to their perceived empirical 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	$\Delta B_i^t$
Period ( $\beta_1$ )	-0.001 (0.001)
bTR-CF ( $\beta_2$ )	-0.003 (0.035)
Period $\times$ bTR-CF ( $\beta_3$ )	-0.001 (0.001)
Negative deviation of belief re norm from $j$ 's average report ( $\beta_4$ )	0.086** (0.032)
Negative deviation of belief re norm from $j$ 's average report $\times$ bTR-CF ( $\beta_5$ )	-0.045 (0.046)
Positive deviation of belief re norm from $j$ 's average report ( $\beta_6$ )	-0.144** (0.058)
Positive deviation of belief re norm from $j$ 's average report $\times$ bTR-CF ( $\beta_7$ )	0.051 (0.065)
Negative deviation of belief re norm from own average report ( $\beta_8$ )	0.014 (0.052)
Negative deviation of belief re norm from own average report $\times$ bTR-CF ( $\beta_9$ )	0.085 (0.093)
Positive deviation of belief re norm from own average report ( $\beta_{10}$ )	-0.305*** (0.066)
Positive deviation of belief re norm from own average report $\times$ bTR-CF ( $\beta_{11}$ )	0.234*** (0.081)
Other controls	Yes
Observations	2346
Number of clusters	51
p>chi2	0.001

*Notes:* The table reports the coefficients of a random-effects 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. Participants 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

<sup>46</sup>For example, consider a participant who, at a given period  $t$ , thinks that not many past participants cheated in that period. He then rolls a blue and realizes that the temptation to report a red is too high. He thus reports a red. At the same time, he may anticipate that others will behave analogously and, therefore, 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 participant  $i$  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.

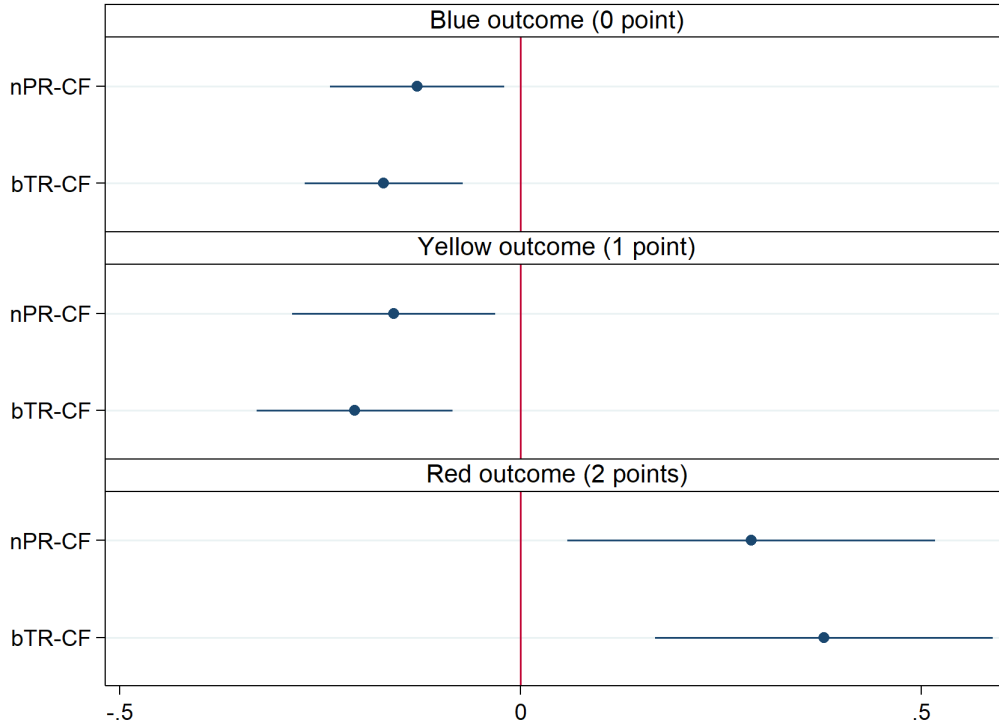


Figure D.3: Marginal Effects of Beliefs on the Probability to Report Each Outcome, by Treatment.

*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.



## E Comparison of $\theta$ in the Tournament and Piece Rate

In the tournament, the stakes associated with lying are higher than in the piece rate. Higher stakes make lying more attractive. At the same time, they also increase the fear of getting caught if one lies (Kajackaite and Gneezy, 2017). In our theoretical framework, this is captured by increasing  $\theta$ . The two effects—lying more attractive and higher fear of getting caught—might cancel each other out, meaning that the net effect of increasing the incentives associated with lying on behavior might be null. In this section, we ask the question: how much should  $\theta$  increase in the tournament as a result of the raise in stakes to make the level of cheating equivalent in the tournament and under a piece rate? To answer this, we need to solve the following equation for  $\theta_T$ :

$$\frac{1}{2\theta_{PR}} = \frac{1 + 48g(0)}{2\theta_T - 2g(0)} \quad (8)$$

where, in left-hand side, we have the optimal level of cheating under the piece rate, and, in the right-hand side, the optimal level of cheating in the tournament. Note that we can approximate  $g(\xi)$ , where  $\xi = x_j - x_i$ , with the Normal distribution. This is because the probability distribution of the sum of several dices approaches to the Normal distribution as the number of die rolls increases (Chavan, 2014). This is shown in Figure E.1 where we report the frequency distribution of the sum of 24 three-outcome die rolls, obtained by running 1,000,000 simulations. The mean of the distribution is 24, and the standard deviation is around 4. Hence, we can say that  $x \sim N(24, 16)$ , and  $g(\xi) \sim N(0, 32)$ .<sup>47</sup> The probability density function of  $\xi$  evaluated at  $\xi = 0$  is:

$$g(0) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(0-\mu)^2/2\sigma^2}$$

where  $\mu = 0$  and  $\sigma \approx 5.66$ . Hence,  $g(0) \approx 0.071$ . Substituting this into (8) and solving for  $\theta_T$  yields:

$$\theta_T = 4.408\theta_{PR} + 0.071 \quad (9)$$

Equation (9) indicates that the value of  $\theta_T$  must increase by a factor larger than 4 to have a similar level of cheating in the tournament and under a piece rate.

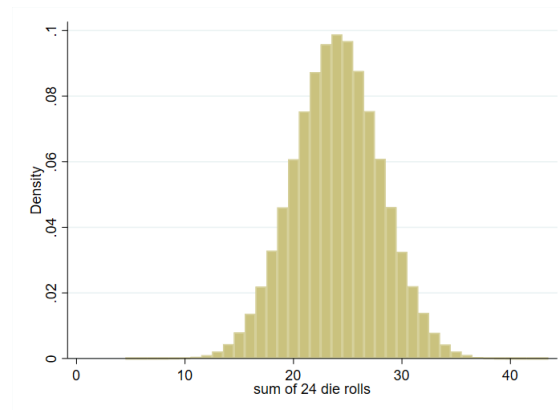


Figure E.1: Simulated frequency distribution of the sum of 24 die rolls

<sup>47</sup>Note that  $E(\xi) = E(x) - E(x)$ , while  $V(\xi) = V(x) + V(x) = 2V(x)$ .

## F Existence of a Unique Nash Solution in the Two-sided Tournament

In this section, we prove that, in the two-sided tournament, a unique Nash solution exists if  $\theta > 0.2$ . Recall that agent  $i$ 's reaction function is given by:

$$(l_i + x_i)g(x_j - x_i) + G(l_i - l_j) - \theta l_i = 0$$

In expectation,  $x_i = x_j = 24$ . Hence, we can rewrite the reaction function as:

$$(l_i + 24)g(0) + G(l_i - l_j) - \theta l_i = 0 \quad (10)$$

Note that  $g(\xi)$  approximates the Normal distribution with  $E(\xi) = 0$  and  $V(\xi) = 32$  (see Online Appendix E). Hence,  $g(0) \approx 0.071$  and  $G(l_i - l_j) \approx \frac{1}{2} \left[ 1 + \operatorname{erf} \left( \frac{l_i - l_j}{5.66\sqrt{2}} \right) \right]$ . Substituting this into (10), we can construct the reaction functions for both agent  $i$  and  $j$  (the opponent) for different values of  $\theta$  (recall that agent  $j$ 's reaction function is symmetric to that of agent  $i$ ). These reaction functions are shown in Figure F.1. It is clear from the figure that a unique Nash solution exists for  $\theta > 0.2$ . Note also that the solution implies partial cheating ( $0 < l^* < 24$ ), at least for the values of  $\theta$  relevant for this study (see 'Behavioral Conjectures' Section).

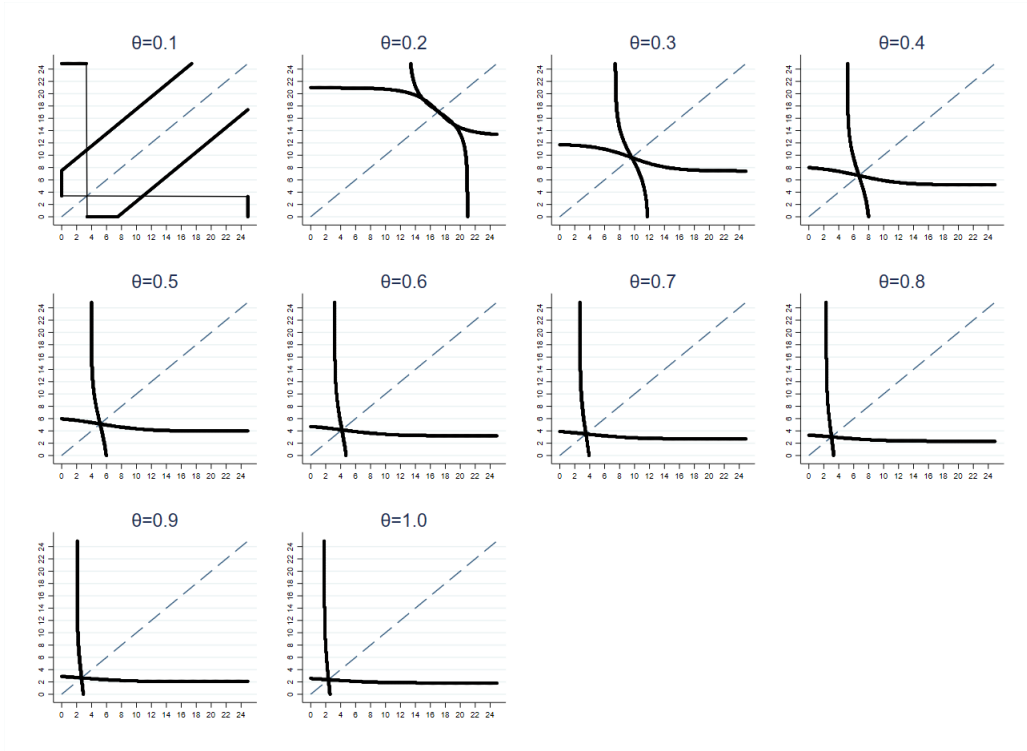


Figure F.1: Reaction functions of agents  $i$  and  $j$  for different values of  $\theta$