Incentives for dishonesty: An experimental study with internal auditors

Loukas Balafoutas, Simon Czermak, Marc Eulerich and Helena Fornwagner [#]

Abstract

We conduct an experiment with professional internal auditors and evaluate their performance and objectivity, measured as the extent to which they truthfully report the performance of other participants in a real-effort task. It has been suggested in the literature that incentive-based compensation for auditors has the potential to lead to dishonest behavior, for instance when their payoff depends on the performance of the unit they are auditing. In line with our hypotheses, we find that incentive-based compensation increases dishonest behavior among internal auditors: competitive incentives lead to under-reporting of other participants' performance, while collective incentives lead to over-reporting of performance. In addition, we find that moving from an environment with objective performance evaluation towards a peer evaluation scheme reduces performance.

Keywords: dishonesty; incentives; sabotage; internal audit; experiment

JEL classifications: C93; M40; M49

[#] Balafoutas: Department of Public Finance, University of Innsbruck, Universitätsstrasse 15, 6020 Innsbruck, Austria (email: loukas.balafoutas@uibk.ac.at); Czermak: Management Center Innsbruck, Universitätsstrasse 15, 6020 Innsbruck, Austria (email: simon.czermak@mci.edu). Eulerich: Mercator School of Management, University of Duisburg-Essen, Lotharstraße 65, 47057 Duisburg (email: marc.eulerich@uni-due.de) (Corresponding Author). Fornwagner: Department of Public Finance, University of Innsbruck (email: helena.fornwagner@uibk.ac.at). We are grateful to the German Institute of Internal Auditors (DIIR) for supporting this research project. We would like to thank our research assistants Joel Behrend, Selina Fröhle, Lisa Hauf, Ronja Krane, Sarah Stuefer and Matthias Walter for their valuable help in conducting the experimental sessions in Munich and Frankfurt. We also thank Bram Cadsby, Marie Claire Villeval, participants at the ESA European meeting 2016 in Bergen and Asia Pacific meeting 2017 in Taipei, and seminar participants at the University of Queensland, Queensland University of Technology, Max Planck Institute for Tax Law and Public Finance, GATE Lyon and the University of Innsbruck for useful comments. This project has received a Certificate of Good Standing (21/2016) from the Board for Ethical Questions in Science at the University of Innsbruck.

1. Introduction

Honest and ethical behavior within organizations is of paramount importance, not only for the organizations involved but also for stakeholders and society as a whole. It therefore comes as no surprise that the popular press as well as the academic literature abound with stories of unethical conduct, fraud or deception and with proposed mechanisms to constrain such phenomena. One aspect of unethical conduct which has received particular attention in the literature is sabotage, defined as '*any (costly) actions that one worker takes that adversely affect the output of another*' (Lazear, 1989). This definition is particularly relevant in the context of tournaments and personnel economics, because it embodies the idea that employees within an organization very often compete – implicitly or explicitly – against each other for promotions, bonuses, and the like. It has been shown that, in an environment in which workers have the opportunity to evaluate each other and hence to engage in strategic sabotage, efficiency is compromised due to a combination of output destruction through sabotage and lower performance by workers in anticipation of sabotage by others (Carpenter et al., 2010).[1]

Sabotage in the form of destroying, reducing or *under-reporting* the output of one's co-workers is not the only potential threat to ethical behavior and in particular truthful information transmission in organizations. Misreporting can also take the form of inflating one's performance, for instance when individual piece rates or targets are used to determine payments (Cadsby et al., 2010; Conrads et al., 2013). Moreover, let us consider the possibility that, in certain cases, workers may have incentives to *over-report* the output of their colleagues. This kind of behavior may entail severe efficiency losses, arising if the management must make decisions based on false information about the performance, risk outlook, or financial viability of various teams, units, or divisions of a firm.

The propensity of an organization's employees to engage in untruthful reporting is likely to depend on the incentives – typically in the form of incentive-based compensation – provided by the management. Broadly speaking, one can distinguish between three classes of incentive schemes: (i) *individual scheme*s, which include flat payments but can also condition one's compensation on her own performance through piece rates or target-based payments, (ii) *competitive schemes*, in which part of an employee's compensation depends on her performance relative to other employees using tournament-like incentives, typically within a team or organizational unit, and (iii) *collective or team-based schemes*, which condition part

[1] See Chowdhury and Gürtler (2015) for a recent survey on sabotage in tournaments.

of one's payment on the performance of a group using team incentives like bonuses based on achieving team targets, or even payment schemes directly tying individual payments on the overall success of (a unit of) the organization.

In this paper we examine experimentally how cheating behavior in the form of under- and over-reporting varies depending on the prevailing incentive scheme. We do this using a sample of professional internal auditors recruited during two large conferences of the German Institute for Internal Audit. We measure two main outcomes, namely the performance of participants in a simple real effort auditing task as well as their evaluation (report) on the performance of one other, randomly matched participant. We differentiate between an individual, a competitive and a collective, or team-based, payment scheme. We build on the design by Carpenter et al. (2010), which features peer evaluation and sabotage opportunities in a tournament and a piece rate environment, and we add to this design a collective payment scheme based on team performance.

The novelty of our study lies, on the one hand, in the fact that we extend the literature on incentives driving untruthful information transmission about others' performance to include not only sabotage in the spirit of Lazear (1989), but also strategic over-reporting of peer performance. We thus study both types of misreporting and their response to incentives in a unified setting and under two experimental conditions: one in which performance is objectively evaluated by the experimenter and one that relies on peer evaluation. Moreover, our work studies these questions using a sample of professionals who are typically faced with (a combination of) the three classes of incentives outlined above and whose job consists of evaluating other units within their firm, creating ample opportunities for strategic under- or over-reporting. We will argue in the following section that this sample is particularly well suited for the examination of our research questions because incentive-based compensation and the particular performance indicators used often create misaligned incentives between auditors and higher management.

In line with our hypotheses we find that, when payments are determined on the basis of mutual evaluations and competitive incentives are present, there is substantial under-reporting of the matched partner's performance: on average, reported points are 22% lower than actual points. On the contrary, we find substantial over-reporting in the presence of team incentives: reported performance is inflated by about 16% on average. These findings suggest that, even in our sample which is subject to very high honesty standards in their job, misreporting of other participants' performance is relative widespread and responds to

incentives in a systematic way. We further examine the share of untruthful reports and find that those are relatively low, meaning that most participants submit truthful reports but there are those who misreport in a systematic way in order to increase their monetary payoff. Finally, we find that average performance is lower under peer than under objective evaluation, given that participants anticipate that their performance may be distorted by their peers. This result has implications for the appropriate design of evaluation schemes in organizations.

Our work relates to various strands of the literature. We have already mentioned Carpenter et al. (2010) who have shown that competitive incentives in organizations lead to more sabotage and lower performance compared to a piece rate scheme. That competitive incentives generally increase sabotage has also been shown in Harbring and Irlenbusch (2005, 2011), while Charness et al. (2014) find increased sabotage and lower performance in tournament environments even in the absence of monetary incentives. The above studies consider sabotage in the form of peer evaluation, but tournament incentives have also been shown to increase dishonesty when individuals are asked to report their own performance: Faravelli et al. (2015) show that lying about one's own performance takes place with a higher probability in competitive (compared to non-competitive) situations and that persons with a tendency to behave dishonestly often select themselves into competitive environments. Schwieren and Weichselbaumer (2010) also find that competitive pressure increases cheating on one's performance in a task, although interestingly this applies only to low performers. Lying about one's performance has further been found to increase with the prize spread in a tournament (Conrads et al., 2014) and to be more pronounced when individual performance targets are in place (Schweitzer et al., 2004; Cadsby et al., 2010).[2]

The literature on team incentives is more limited, but there is enough evidence to suggest that team incentives have the potential to increase dishonest behavior. For instance, Conrads et al. (2013) find that lying in the die-rolling task of Fischbacher and Föllmi-Heusi (2013) is more prevalent when participants are paired in teams of two and paid according to an equal revenue-sharing rule and Wiltermouth (2011) reports a similar pattern in a word unscrambling task. Conrads et al. (2013) discuss possible reasons for this result, which is to some extent counter-intuitive because individual monetary incentives for cheating are weaker under team than under individual incentives. One possible reason, for which the authors find support, is that under team incentives there is some diffusion of responsibility since payments

[2] In addition to targets, bonus payments have also been found to increase cheating in a subsequent stage (Gill et al., 2013).

are determined by the sum of individual reports. Another possible reason is that misreporting in this environment benefits not only the player who gives the report, but also one more player: misreporting can therefore be classified as a Pareto white lie (Erat and Gneezy, 2012). To this discussion we add the possibility that unethical behavior is affected by social norms: in Gino et al. (2009) as well as in Kroher and Wolbring (2015) cheating becomes more frequent when participants have strong indication that another participant has also cheated, while in Gino et al. (2013) experimental participants are less likely to opt for a non-regulation environment where cheating is possible when the choice of environment is public. This is in line with the theory of how social norms of dishonesty can spread (Cialdini et al., 1990; Keizer et al., 2008).[3] Compared to this strand of the literature, we study how team incentives shape untruthful reports not about a die roll or own performance, but about the performance of another participant.

On a more general level, investigating behavior in connection with misreporting and sabotage focuses on ethical aspects of human decision making and is also related to a recent and expanding literature on dishonesty.[4] Although subjects in experiments generally act rationally when taking into account their own and others' costs and benefits from lying (Gneezy, 2005; Sutter, 2009), there is empirical research to indicate that many people do not lie (Abeler et al., 2014) or at least that they do not fully exploit possibilities to maximize profits by acting dishonestly (Abeler et al., 2016). In a task often used to measure dishonesty in economic experiments, participants are found to systematically over-report a die roll in order to increase their payoff, although often not to the full extent (Fischbacher and Föllmi-Heusi, 2013; Shalvi et al., 2011). Kajackaite and Gneezy (2015) argue that individuals experience a fixed rather than convex cost of lying, which would imply that – in the absence of considerations about being exposed as a liar – some participants would never lie while others would lie to the full extent. Gneezy et al. (2013) examine individual heterogeneity with respect to lying behavior and classify experimental subjects into one of eight types based on their behavior in a sender-receiver game with opportunities for lying. They report that about one third of subjects always tell the truth, 28% act in line with pure own money maximizing behavior, and 23% lie when the incentives to do so are high, but not when they are low. Finally, the study by Rosaz and Villeval (2012) is related to ours since it combines the

[3] In a similar spirit, Cohn et al. (2014) show that dishonest behavior can also be influenced by business culture. They conduct an experimental study in which they measure the honesty of employees of large, international banks. As it turns out, the probability of dishonest behavior increases when the professional identity of bankers is made salient compared to a control treatment.

[4] Irlenbusch and Villeval (2015) provide a comprehensive review of the economic literature on sabotage and dishonesty seen from an ethics-based approach.

literature on performance evaluation and lying by examining the frequency of different types of lies (using the classification of Erat and Gneezy, 2012) of a supervisor about a worker's performance. Like in our study, Rosaz and Villeval (2012) find that supervisors bias their evaluations both upwards and downwards depending on the nature of economic incentives. Our setting is quite different, using a sample of professionals and placing hierarchically equal employees either in a direct winner-takes-all competition to each other or in a team. Nevertheless, it is interesting that the findings of the two studies qualitatively point towards the same direction.

Finally, our findings on performance under objective and peer evaluation can be linked to several previous studies on this topic. Building on theoretical considerations drawn from agency theory (Ross, 1973), compensation schemes based on fixed payments are inferior with respect to incentive and sorting effects to schemes that are continuously tied to the performance of a worker (Lazear, 2000a). These effects have found empirical support from both field data (Lazear, 2000b) and controlled laboratory settings including a real effort task (Cadsby et al., 2007). An essential element in performance-based payment schemes – such as those considered in the context of our study – is the evaluation or measurement of performance, which can be based on subjective or objective criteria (depending on the relevance of other persons' judgements in the assessment process). As relevant dimensions of performance are often not open to an objective evaluation or can be influenced by factors outside the control of evaluated individuals (Bol, 2008), a considerable string of literature points to the advantages of subjective measures or suggests a combined application of objective and subjective measures (for an overview see Kunz, 2015). Despite such limitations to the exclusive use of objective measures, recent studies in part overturn prior consensus. For instance, Ahn et al. (2010) show that objective performance evaluation includes a higher degree of discriminability, which in turn has a positive effect on employees' performance improvements. Similarly, Kunz (2015) finds that the inclusion of subjective elements into evaluation systems can result in lower work effort than in situations of purely objective evaluation, but only when those subjective elements introduce imprecision and when individuals' autonomous motivation is low.

The rest of the paper is structured as follows. Section 2 introduces some important aspects of internal auditing that are relevant for this study, including potential conflicts of interest and the role of incentive-based compensation schemes. Section 3 outlines the

experimental design and procedures and formulates our hypotheses, while section 4 presents the findings of the experiment. Section 5 concludes.

2. Internal auditing and auditors' objectivity

Although internal auditing is one of the main pillars of good corporate governance (IIA, 2016), "research on the internal audit function (IAF) is still in its infancy" (DeFond andZhang, 2014). From a practical perspective, the Institute of Internal Auditors (henceforth IIA), the worldwide professional association of internal auditors, defines in its professional practice framework the fundamental purpose and scope of internal auditing as follows: '*Internal auditing is an independent, objective assurance and consulting activity designed to add value and improve an organization's operations. It helps an organization accomplish its objectives by bringing a systematic, disciplined approach to evaluate and improve the effectiveness of risk management, control, and governance processes*' (IIA, 2016). All internal auditors and internal audit functions who are members of the IIA or auditors who hold a professional certification of the IIA have to follow this definition and the professional practice framework. To create value for an organization through the different consulting and assurance activities, the objectivity of internal audit functions and the personal objectivity of each internal auditor is existential. The IIA specifies this attribute in the standard 1120, 'Individual Objectivity': '*Internal auditors must have an impartial, unbiased attitude and avoid any conflict of interest*.' In their interpretation, the worldwide standard-setter explains that conflicts of interests may arise when an internal auditor has a competing professional or personal interest and cannot fulfill his or her duties objectively. Hence, the questions of objective reporting are essential in the field of internal audit.

Drawing on empirical insights from psychology and economics, Bazerman et al. (1997) suggest that auditors' independent and objective decision making is hardly possible, due to inherent conficts of interest in the auditing profession. In this particular context the authors mainly argue on the basis of the self-serving bias (for an overview see Babcock and Loewenstein, 1997). By conducting a case study with practicing members of the IIA, Ahlawat and Lowe (2004) show that judgements of internal auditors are influenced by their advocacy position, thus compromised by interests of their company (in case of in-house auditing) or clients (in case of outsourced auditing) and additionally observe that the corresponding effect is more severe in case of in-house auditing.

One stream of research focusing on the objectivity of internal auditors finds potential conflicts of interests when companies use internal auditing as a so-called Management Training Ground (henceforth MTG), e.g., Messier et al., 2011; Christ et al., 2015; Abbott et al., 2016. MTG describes a staffing arrangement, which rotates new or experienced employees into the internal audit functions for a certain time, before promoting them into management positions outside internal audit after a successful period as an internal auditor.[5] Only the best and overperforming internal auditors will be promoted into promising positions, while underperforming auditors will stay for a longer period in the internal audit functions or will never get a promotion. Findings show that, especially in this arrangement, a loss of objectivity (Hoos et al., 2014; Rose et al., 2013) and efficiency (Anderson et al., 2012) can be identified.

Another potential conflict of interest closely related to our experimental design arises from the compensation of internal auditors. In general, through an incentive-based compensation individuals should be motivated to align their own interests with the interests of the company in order to minimize agency problems (Baker et al., 1988; Chong and Eggleton, 2007). However, a risk of biased measures of performance in order to maximize the individual welfare can also be recognized (Watts and Zimmermann, 1990). This type of behavior is not only dysfunctional, but also ethically questionable and not objective. In particular, incentive-based compensation can influence the objectivity and pose a threat for the professional judgement of internal auditors. Prior research has shown that incentive-based compensation is a common practice in many organizations (e.g., Dezoort et al., 2000; Dickins and O'Reilly, 2009) The typical types of performance indicators with influence on the individual payment are overall company performance, performance of the internal audit function, and individual performance (Dezoort et al., 2000). On the one hand, this structure should increase the productivity and motivation of auditors and recruit talented new employees. On the other hand, negative findings of an audit can decrease the company's performance and thus indirectly reduce the individual salary of an internal auditor, which can affect their objective judgement (Mutchler, 2003).

Results from existing empirical literature strongly suggest that incentive-based compensation has a negative impact on internal auditors' objectivity. This impact is either measured indirectly by external auditors' planning judgements regarding budgeted audit hours

[5] Internal auditing is a suitable function for this approach, because auditors perform a large range of activities across different departments within the organization and gain experiences with various parts of the business as a preparation for future positions.

or audit fees, assuming that less reliance on internal auditors objectivity' results in a higher amout of audit hours or higher audit fees (Dezoort et al., 2001; Chen et al., 2017), or measured directly by assessing internal auditors' perceptions in hypothetical scenarios (Schneider, 2003; Hanafi and Stewart, 2015). In the audit planning experiment of Dezoort et al. (2001), incentive-based compensation negatively affected external auditors' reliance on internal auditors' work, but only in situations where the relevant task was subjective. Chen et al. (2017) compare data from a survey about the use of incentive-based compensation for chief internal auditors with archival data about audit fees charged for corresponding audits, and find a positive relationship between the application of incentive-based compensation and the amount of audit fees, which is mainly driven by compensations fully or partially paid in form of stocks or stock options. The results of Schneider (2003) similary suggest an impairment of internal auditors' objectivity in cases where incentive-based compensations are tied to stock prices. Hanafi and Stewart (2015) show that incentive-based compensation based on company performance significantly affects internal auditors' objectivity, if the latter bias their decision to achieve a company's performance targets. They find no evidence that compensation based on individual performance influences objectivity, a result which could be explained by concerns about job security and the long-term career prospects of the auditor.[6]

Both practices, incentive-based compensation and MTG, are a common practice of today's internal audit functions. For instance, with respect to incentive-based pay, a recent survey among 450 internal auditors in Germany, Austria and Switzerland reveals that more than half of the respondents had a variable component in their wage. This variable component depends primarily on the success of the firm (in 43% of cases) and on the achievement of individual targets (in 45% of cases), and to a lesser extent (8% of cases) on the results of the audited unit (Eulerich, 2014). As a consequence, internal auditors are in a situation where outperforming their own objectives and measures can increase their chance of moving up into a future management position and also maximize their individual wealth. It is obvious that the motivation to misreport, cheat or sabotage in reporting audit results is high.

The implementation of competitive incentives and team incentives in different stages of our experimental design directly relates to the characteristics of incentive-based

[6] It is worth mentioning that there is also some literature on the impairment of objectivity among external auditors (e.g., Koch et al. 2012; King 2002). We focus our experiment and our discussion on internal auditors, noting that internal and external auditing are alternative governance and assurance mechanisms (Felix et al. 2001) with largely different tasks and objectives. Moreover, the incentive-based compensation or MTG arrangement is only present in internal auditing.

compensation contingent on individual and corporate performance. More precisely, competitive incentives implemented in our experiment mirror a situation in which the internal auditing function serves as a MTG,in the sense that the competitive aspect requires a promising candidate for future management positions to outperform potentially competing colleagues in the internal auditing department. Similarly, the team aspect of incentive-based compensation may urge an ambitious internal auditor to avoid decisions which could harm corporate performance. An important advantage of our experimental design in this context is the direct measurement of impaired objectivity in form of misreporting or sabotage activities conducted by experimental participants, which would be practically impossible in real-life contexts. Thus, we believe that this makes the setting of our experiment particularly well-suited to study the effect of conflicts of interest on (dis)honest behavior of internal auditors, stemming from compensation schemes and recruitment practices for top-management positions.

3. The experiment

3.1 Design and experimental procedures

We approach the issues discussed above by running a 'lab in the field' experiment with professional internal auditors who are members of the German Institute for Internal Audit ('Deutsches Institut für Interne Revision', DIIR). We take the study by Carpenter et al. (2010) as the benchmark for our design. Participants in our experiment performed a real effort task, namely the identification of incorrectly added sets of three two-digit random numbers between 11 and 99. The use of real effort tasks is widespread in economic experiments (Charness and Kuhn, 2011), and we chose this particular task because to some extent its nature (discovering errors) is related to everyday tasks and processes of our participants in their workplace.

In detail, participants received a list of 30 calculations and the information that out of 30 stated results 10 were wrong.[7,8] Their task was to identify, without using a calculator, as

[7] An example sheet can be found in Figure A.1 in the Appendix.
[8] We distributed the incorrect results across the list of calculations by always placing one incorrect calculation by chance somewhere within three consecutive calculations: for instance, this means that within the first four calculations at least one and at most two were wrong. This was not made known to participants. In addition, to avoid the possibility of some subjects identifying the mistakes very quickly by checking the sum of the last digits only, we chose the calculations such that, in every stage, only two out of the ten mistakes could be identified by looking at last digits only.

many of the mistakes as possible within three minutes and to mark the mistaken calculations with a cross. They received one point if they correctly identified a mistake and minus 0.5 points if they marked a correct result as false. This means that the highest possible score in this task is 10 points and the lowest possible score is -10 points, while if a participant crossed out all 30 calculations she would achieve a score of zero.

All participants had to go through three stages in which they performed the task outlined above. These stages correspond to the three different types of incentives discussed in the introduction, namely individual, competitive, and team incentives. Every participant received a new sheet of calculations for each stage, but we ensured by testing the real effort tasks with students that the difficulty of the calculations remained the same for each stage. We should also point out that participants received no feedback between stages regarding the reported points, or regarding whether they had won the competition (in Stage 2, see below). Feedback was only provided when subjects were paid their earnings after the end of the experiment. The incentives for the three stages were as follows:

Stage 1: Individual incentives. Participants received a fixed payment of €2, plus a piece rate of €2 for each point they achieved in the task.

Stage 2: Competitive incentives. Participants were randomly assigned into groups of two and competed against each other. The higher performer in each group (i.e., the person with the higher number of points in the task) received a fixed payment of €2 plus a payment that was double the one in the piece-rate scheme (€4 per achieved point), while the low performer received only the fixed payment. Participants had no information regarding the identity of their partner, nor did they know that they would evaluate the performance of the same person who would evaluate them. This was done in order to minimize concerns that reporting behavior could be driven by motives of conditional reciprocity.

Stage 3: Team incentives. Participants were again in groups of two and were paid according to a simple equal sharing rule: each group member was paid the fixed payment of €2, plus €1 for each point achieved in the task by the two group members together (i.e., half the piece rate payment). Participants had no information regarding the identity of their partner and their were told that they would be again be randomly matched with some (meaning that their partner could be the same as in stage two, but did not have to).

To control for order effects we let half of the participants go through the three stages in the order outlined above, while for the other half we reversed the order between the

competitive and the team stage. The individual stage was always played first, to give participants some experience with the task and to have a benchmark against which the effects of introducing compeititve and team incentives are evaluated. After performing the task and handing back the answer sheet to the experimenters, each participant received a sheet from the experimenters, showing the correct answers and where the 10 mistakes lay. He or she also received the answer sheet from another person and was asked to report the number of correctly solved tasks of this matched participant.[9] This was done in all three stages. It was made clear to participants that they were free to report any number they wanted for the performance of their matched participant.[10,11]

We implemented two different treatments in a between-subjects design, varying the evaluation protocol that was used to determine payment-relevant performance in the task. The two treatments were the following:

Peer evaluation (henceforth *Peer*). In this treatment, payments were solely determined on the basis of participants' reports about each other's performance.

Objective evaluation (henceforth *Objective*). In this treatment, reports made by participants about each other's performance had no consequences and all payments were based on an objective evaluation of each participant's performance by the experimenters. Although reports had no consequences, we still elicited them to keep the design and the two treatments comparable, see also Carpenter et al. (2010).

At the end of the experiment every participant was asked to fill out a questionnaire with questions about their gender, age, level of education, working experience, working position, salary range, experience, compensation scheme, risk attitudes, and the perception of how competitive their occupation is. We ran our experiments during two large conferences of the DIIR, in Frankfurt (June 2016) and in Munich (October 2016). Sessions lasted for about 45 minutes and the average payment was €46.15 per participant.

[9] One possible concern could be that some participants might falsify the performance of their matched partner by changing the partner's responses ex post, in order to bring them in line with his or her report. We ruled out this possibility by handing out two different colors of pens for the performance and the reporting part in each stage. Thus, no subject could have added or subtracted points from their partner without the experimenters noticing.
[10] To make participants feel as unobserved as possible, we installed mobile separating walls between the seats.
[11] Whenever a subject privately asked whether there could be sanctions by the experimenter for misreporting a performance, we made it clear that this was not the case.

3.2 Hypotheses

Pure own money maximizing concerns lead to the straightforward prediction of under-reporting (sabotage) in the competitive stage and over-reporting in the team stage in the peer evaluation treatment. In particular, a participant whose only concern is to maximize her expected payoff would report a performance of -10 for her matched partner in the competitive stage and a performance of 10 in the team stage. In the individual stage, reports about others do not matter for own payoffs and hence any report is consistent with own money maximization. Likewise in treatment *Objective*, since points are objectively determined by the experimenters and reports about others have no consequences whatsoever, there is no reason to expect systematic under- or over-reporting and reports can take any value.

The evidence from existing experiments described in the introduction indicates that some people are unwilling to cheat, some people cheat to the full extent if it is in their interest to do so, while others cheat but do not exploit the full extent of misreporting opportunities. Partial cheating is consistent with theories of social or self-image concerns as well as with the existence of psychological costs of lying. Such costs of lying may stem from an inherent aversion to misrepresenting the truth through telling a lie, or they may stem from guilt feelings directed towards the affected person in line with the theory of guilt aversion (Charness and Dufwenberg, 2006; Battigalli and Dufwenberg, 2007; see also Battigalli et al., 2013, for a discussion on the role of guilt in explaining lying aversion). Regardless of the nature of the psychological costs of lying, on average we should observe at least some degree of misreporting under competitive and team incentives in treatment *Peer*. Accordingly, we formulate our two predictions on reporting decisions as follows:

Hypothesis 1: In the competitive stage of treatment *Peer*, we expect systematic under-reporting such that reported points lie below actual points on average.

Hypothesis 2: In the team stage of treatment *Peer*, we expect systematic over-reporting such that reported points lie above actual points on average.

There is an important distinction to be made between under- and over-reporting: while the former hurts the matched partner (at least in expectation), the latter benefits both players. In the taxonomy of Erat and Gneezy (2012) this is a Pareto white lie. Moreover, if agents are motivated by guilt aversion, it is likely that they expect their matched partner to expect a high report even if this means over-reporting their score; in this case, pure monetary considerations

as well as guilt aversion both work towards increasing reported points. Therefore, one might expect over-reporting to be more pronounced than under-reporting in our experiment.

Regarding actual performance, the relevant comparison is between performance in the peer evaluation treatment (in which peer reports determine payoffs) and the objective evaluation treatment (in which payoffs are determined solely on the basis of actual performance). If individuals anticipate full under-reporting (over-reporting) of their points by their matched partner in the competitive (team) stage based on pure own money maximizing behavior, then – given a positive cost of effort – they should reduce their performance to zero. If they anticipate some positive probability of misreporting, or if they expect only partial misreporting, then we would still expect them to reduce their performance in these stages on average, as for instance in Carpenter et al. (2010) under competition. In the individual stage the prediction is somewhat less obvious and depends on the expectation of participants about how their peers will evaluate them. There are a number of reasons why participants may expect at least some misreporting, for instance due to mistakes by the evaluators who have no monetary stakes in this stage, or due to altruistic or spiteful motives. Therefore, we also expect a lower performance in this stage under peer than under objective evaluation..

Hypothesis 3: We expect actual performance to be lower in every stage in treatment *Peer* than in treatment *Objective*.

4. Results

4.1 Descriptive statistics

We begin by providing in Table 1 some background information about our participants based on the post-experimental questionnaire. The statistics are also broken down by treatment, in order to show that the two samples are comparable in terms of a number of relevant characteristics.

We ran the experiment with a total of 95 participants[12], 45 in the objective evaluation and 50 in the peer evaluation treatment. 34.74% of our sample were female (33 out of 95) and the average age was 42.4 years. Only 11 participants work for a public or non-profit company,

[12] We collected data for 98 people in total but had to exclude three of them from the analysis: one was using a mobile phone calculator even though we kindly asked him to refrain from doing so, and two participants left before finishing Stage 3, therefore we have no reporting decisions from these participants for Stage 3. All three did not receive any payment.

whereas the majority holds a position in a company of the financial or non-financial private sector. In our participant pool we find primarily Certified Internal Auditors, CRMAs (Certification in Risk Management Assurance), and Certified Public Accountants. Most participants occupy higher level positions (e.g., Head of Internal Auditing), which is also reflected in the reported amount of staff which works for them (36.84 % have at least one person who is hierarchically positioned underneath them).

<Table 1 about here>

Table 1 reveals that our sample is indeed characterized by a very high prevalence of variable payment schemes: over 75 % report having a variable wage component in addition to their fixed wage. This characteristic is not significantly different between treatments. Whenever wages include a variable component, the most common indicators for this component are the auditor's personal performance (62.1% of responses) and the success of the company they are working for (60% of responses).[13] Moreover, a substantial share of participants (30.8%) report being part of the Management Training Ground scheme, which as we have discussed means that they are implicitly in a tournament against other internal auditors within their organization. This proportion does not differ significantly between treatments (see Table 1). Moreover, more than half of our subjects evaluate their profession as at least moderately competitive.

We also included a question about self-assessed risk attitudes. Specifically, we asked participants if they are generally a risk-seeking person or if they try to avoid any kind of risk. The mean reported value is 5.21 on a scale from 0 (completely risk averse) to 10 (fully risk seeking), and is not significantly different between the two treatments. Finally, we also asked participants how many other people in the room they knew personally. Only one person knew more than 20 % of the present persons, while on average the reported percentage was very low at 5.5%.

Table 2 summarizes the actual and reported performance and the differences between the two, for each of the three incentive schemes (stages) and in each of the two treatments

[13] Reported annual pre-tax wages lie over €90,000 for 39 participants, between €70.000 and €90.000 for 23 participants and below €70.000 for 19 participants, while 14 participants did not report their wage range.

(*Peer* treatment in panel a, *Objective* in panel b). We begin by analyzing reported performance and the extent of misreporting in order to test our Hypotheses 1 and 2; we then turn to actual performance in section 4.3 in order to test Hypothesis 3. We note that our two main outcomes, the extent of misreporting and actual performance, do not differ significantly between the two orders with which stages 2 and 3 were played in any of the two treatments. Therefore, in what follows we pool the two orders. Moreover, for ease of exposition, from now on we refer to Stage 2 as the environment with competitive incentives and to Stage 3 as the environment with team incentives.

<Table 2 about here>

4.2 Reporting decisions

Under individual incentives, reported performance (6.07 points) and actual performance (5.96 points) in treatment *Peer* are very close to each other and statistically indistinguishable (p=0.68, Wilcoxon signed-ranks test).[14] Hence, participants on average report each other's performance truthfully in the absence of monetary incentives to misreport. Under competitive incentives, however, participants can increase their chances of winning the tournament by under-reporting the performance of their partner. In line with Hypothesis 1, we find that these incentives introduce a large difference of 1.3 points between actual and reported points (5.87 vs. 4.57, p=0.04, Wilcoxon signed-ranks test), meaning that performance is under-reported by 22.1% on average. This pattern reverses under team incentives, where participants can increase their payoff by over-reporting the performance of their partner. In this stage we find that reported performance is on average about one point higher than actual performance, which corresponds to an increase of 16.4% (p=0.001, Wilcoxon signed-ranks test). This finding supports Hypothesis 2. Comparing the difference in differences across the three stages, we can confirm that under-reporting (in the competitive stage) and over-reporting (in the team stage) are significantly more extensive compared to the differences observed in the individual stage (p=0.01, p=0.001 respectively, Wilcoxon signed-ranks tests).

It is worth comparing actual and reported points also in treatment *Objective*, in which reports have no bearing on payoffs regardless of the incentive scheme in place. As expected, we find no significant difference between actual and reported points in the individual stage

[14] All reported p-values are based on two-tailed tests.

and in the team stage (p=0.63, p=0.19 respectively, Wilcoxon signed-ranks tests). However, we do find some evidence of under-reporting in the presence of competitive incentives with reported points being slightly lower (by 0.13 points) than actual points (p=0.06). While at first this seems like a puzzling result given that reports in the objective evaluation treatment do not matter for payments, it can be natural for participants to engage in some limited under-reporting in order to enhance their self-image or status perception. Interestingly, Carpenter et al. (2010) and Charness et al. (2014) also find that participants engage in sabotage when they compete against each other even in the absence of material incentives to do so. They say that this kind of behavior '*could be symptomatic of a simple, maybe even unconscious or affective/emotional, response to competition*' (Carpenter et al., 2010; p. 510).

The above analysis indicates that there is on average substantial misreporting. But how widespread is such behavior among participants? Table 3 shows the number of truthful and untruthful reports by stage and treatment, while Figure 1 plots for each treatment the distribution of actual output, reported output, and differences between the two (negative differences indicate under-reporting and positive differences over-reporting).

<Table 3 about here>

<Figure 1 about here>

A closer look at the distribution of differences between reported and actual output reveals, at first, that most participants report truthfully independent of treatment and stage. This is also confirmed in Table 3: the share of truthful reports ranges from 70% (35 out of 50 in treatment *Peer* and team incentives) to 86% (46 out of 50, in treatment *Peer* under individual incentives). Although those participants who report untruthfully are a minority, we have already seen that they lead to statistically significant differences in mean outcomes between stages, and it also turns out to be the case that the share of untruthful reports is higher under competitive and team incentives than under individual incentives in treatment *Peer*. This share increases from 14% in Stage 1 (individual incentives) to 26% and 30% under

competitive and team incentives, respectively (individual vs. competitive stage: p=0.109; individual vs. team stage: p=0.021; competitive vs. team stage: p=0.77; McNemar's tests).[15]

In the objective treatment misreporting rates are identical under competitive and team incentives (15.6%), but they are significantly higher with 26.7% in the individual stage (p<0.001, McNemar's test). While this finding seems surprising, we note that the twelve cases of misreporting in the individual stage are almost symmetrically split between five cases of under-reporting and seven cases of over-reporting (see Figure 1), so that – as already noted above – there is no significant difference between actual and reported performance on average in this stage. Moreover, given that reporting in treatment *Objective* has no consequences whatsoever for payment, it is not at all clear that cases of misreporting in this treatment should be classified as cheating and it is quite likely that at least some cases of misreporting are simply the result of lack of attention or reluctance to exert any effort in order to report the correct performance. For this reason our focus lies on misreporting in treatment *Peer*, in which reports determine payments.

Since we have three reporting decisions for each participant, we can identify how often – and in which stage – a person misreported the performance of his or her partner in the peer evaluation treatment.[16] We begin by noting that 29 out of 50 participants reported the performance of their matched partner truthfully in all three stages. Further, we find that three participants gave a truthful report in Stage 1, but untruthful reports in Stages 2 and 3. Given the monetary incentives for dishonesty in Stages 2 and 3, these participants' behavior responds to both types of incentives in the predicted way. Five participants gave untruthful reports in Stage 2 but not in Stage 3 (all of them also reporting truthfully in Stage 1), responding to competitive but not to team incentives, while seven gave untruthful reports in Stage 3 but not in Stage 2 (all but one of them reporting truthfully in Stage 1), responding to collective but not to competitive incentives. Finally, five participants gave untruthful reports in all three stages and one participant only misreported in Stage 1.

Turning to the extent of misreporting in the peer evaluation treatment, we see in Figure 1 that only one participant resorted to full sabotage by reporting a performance of -10 for their matched partner in the competitive stage, while three participants chose to report -5 points and three chose to report 0. All but one of these reports were untruthful (the truthful

[15] A probit regression with untruthful reporting as the dependent variable also shows that the likelihood of misreporting is significantly higher in the competitive and team stage compared to the individual stage.

[16] We do not report here this analysis for the objective evaluation treatment, given the irrelevance of reporting decisions for payoffs.

report was one of the 0's). This pattern of incomplete cheating is in line with the experimental literature discussed in the introduction. In the case of team incentives, it appears to be the case that people had weaker reservations to cheat 'all the way': eleven participants reported a perfect performance of 10 for their matched partner, when in fact only two had achieved that score. It must be acknowledged, however, that an asymmetry exists between under- and over-reporting: given an average score of approximately 6 points (pooled for Stages 2 and 3 in treatment *Peer*) and a range of possible reported scores between -10 and 10, there is much more room for under-reporting than for over-reporting. This can explain the presence of more outliers with very strong misreporting in the competitive than in the team stage.

In Table 4 we show the results of a regression analysis for reporting decisions (in columns 1 to 4) and actual performance (in columns 5 and 6). The dependent variable in the reporting regressions is the difference between reported and actual points, so that positive (negative) values indicate over- (under-) reporting. The independent variables are dummy variables for the competitive and the team stage, a dummy variable for the *Peer* treatment, interactions between the stage and treatment dummies, as well as own performance and performance of the matched partner. In addition, specifications (2), (4) and (6) further include a number of control variables collected in the post-experimental survey, namely whether the participant is in a Management Training Ground (variable MTG), whether his or her wage includes a variable component, the salary range, gender, risk and competitive attitudes. To account for the fact that each participant performs three times and also makes three reporting decisions (one in each stage), all regressions include random effects and standard errors are clustered at the participant level.

<Table 4 about here>

The first two specifications on reporting decisions pool data from both treatments (*Objective* and *Peer*), while in columns (3) and (4) we consider each treatment separately. All reported results confirm the insights gained from the non-parametric statistical tests. In the pooled regressions, the effects of different incentives on reporting in treatment *Objective* are captured by the insignificant coefficients on the competitive and the team stage. The negative coefficients on the interaction term *Competitive*Peer* show that the effect of competitive incentives is significantly different in treatment *Peer* compared to *Objective*, capturing the

difference in differences, and the same argument applies to the positive coefficients on *Team*Peer*. In order to formally test the hypotheses that competitive (team) incentives lead to statistically significant under-reporting (over-reporting), we need to test the joint coefficients (*comp + comp x peer*) and (*team + team x peer*). F-tests on the restrictions that the joint coefficients are equal to zero yield $p<0.01$ for both restrictions in specifications (1) and (2). In the regressions broken down by treatment, we again see that reporting decisions are significantly affected by the incentive system in treatment *Peer* (column 4), but not in *Objective* (column 3). The above analysis, taken together with the non-parametric tests, allow us to formulate our first two results.

Result 1: Compared to individual incentives, competitive incentives lead to statistically significant under-reporting of other participants' performance in *Peer*. This supports Hypothesis 1.

Result 2: Compared to individual incentives, team incentives lead to statistically significant over-reporting of other participants' performance in *Peer*. This supports Hypothesis 2.

Regarding the effect of other variables, we begin by noting the coefficient on the performance of the matched partner, which is negative and significant in (1) and (2), suggesting that high performers are more likely to be the victims of sabotage.[17] In columns (2) and (4) we also see a positive coefficient for own performance, meaning that high performers also report higher numbers for their matched partners, ceteris paribus. Focusing on the data from treatment *Peer*, which drives most of the variation in our data, two more variables are marginally significant: reported points are negatively associated with the perceived competitiveness of the internal auditing profession (indicating that participants who perceive their job environment as more competitive report lower scores for their matched partners on average), and negatively associated with wage levels.

Looking at the difference between reported and actual performance in the Table 4 regressions does not take into account the fact that, depending on the actual performance of their matched partner, some participants have smaller or larger opportunities for cheating. For instance, if my matched partner's performance is nine points, I can only over-report her

[17] This result has been suggested in the theoretical model of Chen (2003), while Balafoutas et al. (2012) find empirical evidence to support it using field data from Judo tournaments.

performance by one point while I can under-report by up to 19 points. In the appendix (Table A.1) we report alternative versions of specifications 1 and 2 from Table 4, using *relative* over- or under-reporting as the dependent variable, i.e., the difference between reported and actual points as a percentage of the maximum possible difference in each case. Since this new variable is left-censored at -1 and right-censored at 1, we use the Tobit model. We confirm that Results 1 and 2 hold for this alternative specification. Moreover, comparing the effects of competitive and collective incentives in treatment *Peer* reveals that team incentives lead to much more misreporting than competitive incentives do once we take into account the extent of the available opportunities for cheating. For instance, in the second specification, over-reporting in the team stage relative to an individual's true performance is almost 20% higher than in the individual stage, while this difference is less pronounced for under-reporting in the competitive stage and lies at around 11 percentage points.

4.3 Performance

We begin by comparing performance across stages in order to examine how it responds to incentives within a given treatment. Casual inspection of the means in Table 2 as well as statistical testing shows that performance does not differ across stages. This finding suggests that performance does not respond to incentives and it is in contrast to much of the literature, which generally finds that performance increases in the presence of competitive incentives. Nevertheless, this finding can be explained if participants are already investing maximum effort in Stage 1 (for instance because the costs of effort are relatively low), so that performance cannot increase any further. In general, the response of performance to monetary and non-monetary incentives in real effort tasks in economic experiments is less clear-cut as one would think, and some recent papers show that participants in experiments often do not change their performance in response to stronger incentives (Eckartz et al., 2012; Araujo et al., 2016) or that this change depends critically on the particular task used (Takahashi et al., 2016). Using structural estimates based on a field experiment on gift exchange, Della Vigna et al. (2016) report a very low elasticity of effort with respect to monetary incentives.

Our Hypothesis 3 predicts that actual performance will be lower in treatment *Peer* than in *Objective*. Indeed, this is what we observe in the data (see Table 2): performance is lower in *Peer* than in *Objective* under individual incentives (5.96 vs. 6.73, p=0.04, Mann-Whitney test), competitive incentives (5.87 vs. 6.66 p=0.02) and team incentives (6.16 vs. 7.05,

p=0.08). Hence, Hypothesis 3 is confirmed in our dataset. In columns (5) and (6) of Table 4 we report the results of two regressions with actual output as the dependent variable. The *Peer* treatment dummy has a significant negative coefficient, capturing the fact that performance is higher in every stage in *Objective* compared to *Peer*. The difference in performance between treatments does not vary significantly by incentive scheme (as seen by the insignificant interaction terms). In terms of our controls, the only variable that does have an effect on performance is the wage level, with higher earners performing better in the experimental task on average.

Result 3: In line with Hypothesis 3, performance is lower in every stage in the peer evaluation compared to the objective evaluation treatment.

How can Result 3 be explained? Hypothesis 3 puts forward the idea that participants anticipate the possibility of untruthful reports in treatment *Peer*, which leads them to reduce performance compared to *Objective* (where their evaluation by the experimenters is fully accurate by design). In order to examine this explanation, we elicited the beliefs of internal auditors about the share of untruthful reports. The belief elicitation took place approximately 8 months after the experiment, during another conference of the DIIR in Düsseldorf in June 2017. A sample of 94 conference participants (all of whom were certified internal auditors) were given a precise description of the experimental design and asked to give their estimate of the share of untruthful reports in each of the three stages in treatment *Peer*.[18] The results reveal that participants indeed expected some misreporting in every stage: the expected share of untruthful reports was 14%, 29.2% and 27% in the invididual, competitive and team stage, respectively. All of these shares are significantly different from zero ($p<0.01$, Wilcoxon signed ranks tests), suggesting that indeed participants were likely to expect a positive probability that their true performance would be misreported in the peer evaluation treatment.[19]

5. Concluding remarks

[18] Out of these 94 participants, 21 indicated that they had participated in the experiment in 2016. Excluding these 21 participants does not affect our findings.

[19] It is also interesting to note that these expectations are very close to actual misreporting rates (14%, 26% and 30%, respectively). This indicates that beliefs were very well calibrated.

The aim of this study has been to examine experimentally how dishonest behavior in the form of misreporting the performance of another participant reacts to incentives in the form of different compensation schemes. Using a sample of professional internal auditors recruited during two large conferences of the German Institute of Internal Auditing, we have shown that individuals respond to changes in incentives in the predicted way: on average, they under-report each other's performance under a competitive compensation scheme and over-report it under a team-based compensation scheme. These findings complement the existing literature and draw attention to the potential perils associated with incentive-based compensation as it is often encountered in organizations. We have argued that professional internal auditors are particularly well suited for the study of these issues, given that they are employed in a sector where honest and objective behavior is among the central guiding principles, and also given that they typically face at least some combination of the type of incentives examined in our experiment.

Our results can make a number of important contributions to the profession of internal auditing. First, we extend prior literature regarding the impairment of internal auditors' objectivity by demonstrating that the compensation scheme and personal economic interest directly affects the performance and reporting of internal auditors. Extending prior work by Hanafi and Stewart (2015) and Schneider (2003), we examine how internal auditors' objectivity reacts to economic incentives in the form of a tournament environment with the possibility of sabotage effects and an environment with team-based incentives. Given that the majority of companies use performance-based compensation schemes, our results identify potential risks of this approach. We expand the understanding of internal auditors' objectivity with regard to tournament environments and question if an MTG arrangement, as a typical example of a real-life tournament among internal auditors, strengthens or threatens their objectivity. Our findings suggest that objectivity and independence, the two main pillars of the whole internal audit profession, can be jeopardized by the potential effects of an MTG environment (as a tournament situation) as well as by a team performance-based compensation scheme. Boards and the Chief Audit Executive are well advised to take these potential threats into accountwhen setting up their internal audit function, and in general we believe that further exploration of potential compensation schemes that avoid these negative effects is a promising avenue for future research.

In addition, we have shown that moving from an objective evaluation scheme towards an environment that relies on mutual peer evaluation has the potential to reduce performance,

even in the absence of competitive or collective incentives. This finding calls for caution when employing peer evaluation and highlights the importance of improving the precision of evaluation procedures in organizations in general.

At the same time, certain limitations should be acknowledged. While we view the sample of participating professionals as a strength of our study, we are not claiming to have replicated their working environment or studied them under realistic circumstances. For instance, it has been pointed out to us that the specific peer evaluation protocol employed in our experiment corresponds only very loosely to actual firm practices and that behavior might depend critically on the chosen real effort task. Hence, as with any lab (or lab in the field) experiment, the projection of our findings onto real behavior in organizations relies on the implicit assumption that responses to incentives are sufficiently generalizable from one environment to the other. Moreover, given the logistical constraints that we faced (running sessions in a limited amount of time and with relatively small samples), we were not able to investigate a number of further dimensions relevant to our questions of interest. Notably, one interesting treatment – in addition to objective and peer evaluation – would have been one in which participants are asked to evaluate their own performance in the task under different incentives. This would allow for a more comprehensive comparison with the literature, which often measures dishonesty by means of untruthful self-reports, and it would capture one more important dimension of employee behavior.

References

Abbott, L. J., Daugherty, B., Parker, S. Peters, G. F., 2016. Internal audit quality and financial reporting quality: The joint importance of independence and competence. Journal of Accounting Research, 54(1), 3-40.

Abeler, J., Becker, A., Falk, A., 2014. Representative evidence on lying costs. Journal of Public Economics, 113, 96-104.

Abeler, J., Nosenzo, D., Raymond, C., 2016. Preferences for truth-telling. IZA Discussion Paper no. 10188.

Ahlawat, S. S., Lowe, D. J., 2004. An examination of internal auditor objectivity: In-house versus outsourcing. Auditing: A Journal of Practice & Theory, 23(2), 147-158.

Ahn, T. S., Hwang, I., Kim, M., 2010. The Impact of Performance Measure Discriminability on Ratee Incentives. The Accounting Review, 85(2), 389-417.

Araujo, F. A. d., Carbone, E., Conell-Price, L., Dunietz, M. W., Jaroszwicz, A., Landsman, R., Lamé, D., Vesterlund, L., Wang, S., Wilson, A. J., 2016. The slider task: An example of restricted inference on incentive effects. Journal of the Economic Science Association, 2(1), 1-12.

Babcock, L., Loewenstein, G., 1997. Explaining bargaining impasse: The role of self-serving biases. The Journal of Economic Perspectives, 11(1), 109-126.

Baker, G. P., Jensen, M. C., Murphy, K. J., 1988. Compensation and incentives: Practice vs. theory. The journal of Finance, 43(3), 593-616.

Balafoutas, L., Lindner, F., Sutter, M., 2012. Sabotage in tournaments: evidence from a natural experiment. Kyklos, 65(4), 425-441.

Battigalli, P., Charness, G., Dufwenberg, M., 2013. Deception: The role of guilt. Journal of Economic Behavior & Organization, 93, 227-232.

Battigalli, P., Dufwenberg, M., 2007. Guilt in games. American Economic Review, Papers and Proceedings, 97(2), 170-176.

Bazerman, M. H., Morgan, K., and Loewenstein, G. F., (1997). The impossibility of auditor independence. Sloan Management Review 38(4), 89-90.

Bol, J. C., 2008. Subjectivity in Compensation Contracting. Journal of Accounting Literature 27, 1-24.

Cadsby, C. B., Song, F., Tapon, F., 2007. Sorting and incentive effects of pay for performance: An experimental investigation. Academy of Management Journal, 50(2), 387-405.

Cadsby, C. B., Song, F., Tapon, F., 2010. Are you paying your employees to cheat? An experimental investigation. The BE Journal of Economic Analysis Policy, 10(1), 1-32.

Carpenter, J., Matthews, P. H., Schirm, J., 2010. Tournaments and office politics: Evidence from a real effort experiment. The American economic review, 100(1), 504-517.

Charness, G., Dufwenberg, M., 2006. Promises and partnership. Econometrica, 74, 1579-1601.

Charness, G., Kuhn, P., 2011. Lab labor: What can labor economists learn from the lab?. Handbook of labor economics, 4, 229-330.

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

Chen, K. P., 2003. Sabotage in promotion tournaments. Journal of Law, Economics, and Organization, 19(1), 119-140.

Chen, L. H., Chung, H. H., Peters, G. F., Wynn, J. P., 2017. Does Incentive-Based Compensation for Chief Internal Auditors Impact Objectivity? An External Audit Risk Perspective. Auditing: A Journal of Practice & Theory, 36(2), 21-43.

Chong, V. K., Eggleton, I. R., 2007. The impact of reliance on incentive-based compensation schemes, information asymmetry and organisational commitment on managerial performance. Management Accounting Research, 18(3), 312-342.

Chowdhury, S. M., Gürtler, O., 2015. Sabotage in contests: a survey. Public Choice, 164(1-2), 135-155.

Christ, M. H., Masli, A., Sharp, N. Y., Wood, D. A., 2015. Rotational internal audit programs and financial reporting quality: Do compensating controls help?. Accounting, Organizations and Society, 44, 37-59.

Cialdini, R. B., Reno, R. R., Kallgren, C. A., 1990. A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. Journal of personality and social psychology, 58(6), 1015.

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

Conrads, J., Irlenbusch, B., Rilke, R. M., Schielke, A., Walkowitz, G., 2014. Honesty in tournaments. Economics Letters, 123(1), 90-93.

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

DellaVigna, S., List, J., Malmendier, U., Rao, G., 2016. Estimating social preferences and gift exchange at work. NBER Working Paper no. 22043.

DeFond, M., Zhang, J., 2014. A review of archival auditing research. Journal aof Accounting and Economics, 58 (2-3), 275-326.

Dezoort, F. T., Houston, R. W., Reisch, J. T., 2000. Incentive-based compensation for internal auditors. Internal Auditor, 57(3), 42-42.

Dezoort, F. T., Houston, R. W., Peters, M. F., 2001. The Impact of Internal Auditor Compensation and Role on External Auditors' Planning Judgments and Decisions. Contemporary Accounting Research, 18(2), 257-281.

Dickins, D., O'Reilly, D., 2009. The qualifications and independence of internal auditors. Internal Auditing, 24(3), 14-21.

Eckartz, K., Kirchkamp, O., Schunk, D., 2012. How do incentives affect creativity?. CESifo Working Paper Series from CESifo Group Munich, No. 4049.

Erat, S., Gneezy, U., 2012. White lies. Management Science, 58(4), 723-733.

Eulerich, M., 2014. Die Interne Revision in Deutschland, Österreich und der Schweiz 2014. DIIR Verlag, Frankfurt 2014.

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

Felix, W., Gramling, A., Maletta, M., 2001. The contribution of internal audit as a determinant of external audit fees and factors influencing this contribution. Journal of Accounting Research, 39(3), 513-534.

Fischbacher, U., Föllmi-Heusi, F., 2013. Lies in disguise—an experimental study on cheating. Journal of the European Economic Association, 11(3), 525-547.

Gill, D., Prowse, V., Vlassopoulos, M., 2013. Cheating in the workplace: An experimental study of the impact of bonuses and productivity. Journal of Economic Behavior & Organization, 96, 120-134.

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

Gino, F., Krupka, E., Weber, R., 2013. License to cheat: Voluntary regulation and ethical behavior. Management Science, 59(10), 2187-2203.

Gneezy, U., 2005. Deception: The role of consequences. The American Economic Review, 95(1), 384-394.

Gneezy, U., Rockenbach, B., Serra-Garcia, M., 2013. Measuring lying aversion. Journal of Economic Behavior & Organization, 93, 293-300.

Hanafi, H. M.,Stewart, J., 2015. The effect of incentive-based compensation on internal auditors' perceptions of objectivity. International Journal of Auditing, 19(1), 37-52.

Harbring, C., Irlenbusch, B., 2005. Incentives in tournaments with endogenous prize selection. Journal of Institutional and Theoretical Economics JITE, 161(4), 636-663.

Harbring, C., Irlenbusch, B., 2011. Sabotage in tournaments: Evidence from a laboratory experiment. Management Science, 57(4), 611-627.

Hoos, F., Messier, W. F., Smith, J., Tandy, P., 2014. The effects of serving two masters and using the internal audit function as a management training ground on internal auditors' objectivity. Unpublished paper, HEC Paris, and University of Nevada, Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2358149.

IIA (Institute of Internal Auditors Research Foundation), 2016. International Professional Practices Framework (IPPF). Institute of Internal Auditors Research Foundation.

Irlenbusch, B., Villeval, M. C., 2015. Behavioral ethics: how psychology influenced economics and how economics might inform psychology?. Current Opinion in Psychology, 6, 87-92.

Kajackaite, A., Gneezy, U., 2015. Lying costs and incentives. UC San Diego Discussion Paper.

Keizer, K., Lindenberg, S., Steg, L., 2008. The spreading of disorder. Science, 322(5908), 1681-1685.

King, R. R., 2002. An experimental investigation of self serving biases in an auditing trust game: The effect of group affiliation. Accounting Review 77, 265-285.

Koch, C. W., Weber, M., Wüstemann, J., 2012. Can auditors be independent? Experimental evidence on the effects of client type. European Accounting Review, 21(4), 797–823.

Kochetova-Kozloski, N., Messier Jr, W. F., 2011. Strategic analysis and auditor risk judgments. Auditing: A Journal of Practice Theory, 30(4), 149-171.

Kroher, M., Wolbring, T., 2015. Social control, social learning, and cheating: Evidence from lab and online experiments on dishonesty. Social Science Research, 53, 311-324.

Kunz, J., 2015. Objectivity and subjectivity in performance evaluation and autonomous motivation: An exploratory study. Management Accounting Research 27, 27-46.

Lazear, E. P., 1989. Pay equality and industrial politics. Journal of political economy, 97(3), 561-580.

Lazear, E. P., 2000a. The power of Incentives. The American Economic Review, 90(2), 410-414.

Lazear, E. P., 2000b. Performance Pay and Productivity. The American Economic Review, 90(5), 1346-1361.

Messier Jr, W. F., Reynolds, J. K., Simon, C. A., Wood, D. A., 2011. The effect of using the internal audit function as a management training ground on the external auditor's reliance decision. The Accounting Review, 86(6), 2131-2154.

Mutchler, J. F., 2003. Independence and objectivity: A framework for internal auditors', in Baily. A. D., Gramling, A. A., Ramamoorti, S. (eds), Research Opportunities in Internal Auditing, Altamonte Springs, FL: The Institute of Internal Auditors Research Foundation, 231–68.

Rosaz, J., & Villeval, M.C., 2012. Lies and biased evaluation: A real-effort experiment. Journal of Economic Behavior & Organization, 84(2), 537-549.

Rose, A. M., Rose, J. M., Norman, C. S., 2013. Is the objectivity of internal audit compromised when the internal audit function is a management training ground?. Accounting Finance, 53(4), 1001-1019.

Ross, S. A., 1973. The Economic Theory of Agency: The Principal's Problem. The American Economic Review, 63(2), 134-139.

Schneider, A., 2003. An examination of whether incentive compensation and stock ownership affect internal auditor objectivity. Journal of Managerial Issues, 15(4), 486-497.

Schweitzer, M. E., Ordóñez, L., Douma, B., 2004. Goal setting as a motivator of unethical behavior. Academy of Management Journal, 47(3), 422-432.

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

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

Skinner, D. J., Srinivasan, S., 2012. Audit quality and auditor reputation: Evidence from Japan. The Accounting Review, 87(5), 1737-1765.

Sutter, M., 2009. Individual behavior and group membership: Comment. The American Economic Review, 99(5), 2247-2257.

Takahashi, H., Shen, J., Ogawa, K., 2016. An experimental examination of compensation schemes and level of effort in differentiated tasks. Journal of Behavioral and Experimental Economics, 61, 12-19.

Watts, R. L., Zimmerman, J. L., 1990. Positive accounting theory: A ten year perspective. The Accounting Review, 65(1), 131–56.

Wiltermuth, S. S., 2011. Cheating more when the spoils are split. Organizational Behavior and Human Decision Processes, 115(2), 157-168.

Tables and Figures

Table 1: Background information on participants

	Both treatments	N	**Objective**	N	**Peer**	*p* value
Age	42.40 (8.67)	45	42.09 (9.32)	48	42.69 (8.11)	0.79
% Female	34.74	45	33.33	50	36.00	0.83
Experience as IA in years	11.09 (6.46)	45	10 (6.71)	50	12.08 (6.12)	0.07
% variable wage component	76.60	45	68.89	49	83.67	0.14
% variable component depends on personal performance	62.11	45	55.56	50	68.00	0.29
% variable component depends on department performance	12.63	45	11.11	50	14.00	0.76
% variable component depends on company performance	60.00	45	48.89	50	70.00	0.06
% working in public sector	11.58	43	13.33	50	10.00	0.75
% working in finance sector	37.89	43	28.89	50	46.00	0.10
% Master or PhD degree	18.95	42	17.78	50	20.00	0.80
% having at least one person working for them	36.84	45	42.22	47	32.22	0.40
% who are in MTG	30.77	45	28.89	46	32.61	0.82
risk attitude	5.21 (2.20)	45	5.18 (2.06)	49	5.24 (2.34)	0.76
competitive perception	2.90 (1.23)	45	3.04 (1.21)	49	2.78 (1.25)	0.26

Notes: Table presents mean values (with standard deviations in parentheses) or percentages. p values testing equality across treatments, based on Fisher's exact tests, except for variables age, experience, risk attitudes and competitive perception (Mann-Whitney U tests). The number of observations varies slightly across questions due to the fact that some participants did not fill out the complete questionnaire.

Table 2: Mean actual and reported performance, by stage and treatment (standard deviations in parentheses)

Panel (a): Treatment *Peer*

	Individual	**Competitive**	**Collective**
Actual points	5.96	5.87	6.16
	(1.63)	(1.89)	(2.03)
Reported points	6.07	4.57	7.17
	(2.08)	(3.99)	(2.17)
Difference (reported-actual)	0.11	-1.3	1.01
	(1.23)	(3.63)	(2.10)

Panel (b): Treatment *Objective*

	Individual	**Competitive**	**Collective**
Actual points	6.73	6.66	7.05
	(2.26)	(2.63)	(1.74)
Reported points	6.75	6.52	6.98
	(2.14)	(2.74)	(1.84)
Difference (reported-actual)	0.02	-0.13	-0.06
	(1.08)	(1.17)	(0.70)

Notes: N=50 for treatment *Peer* and N=45 for treatment *Objective*. However, since two subjects left the experiment before completing their reports for the collective stage, entries for this stage in treatment *Objective* are based on N=43.

Table 3: Share of (un)truthful reports, by stage and treatment

Panel (a): Treatment *Peer* (number of reports, out of 50)

	Individual	**Competitive**	**Team**
untruthful	7	13	15
truthful	43	37	35

Panel (b): Treatment *Objective* (number of reports, out of 45)[§]

	Individual	**Competitive**	**Team**
untruthful	12	7	5
truthful	33	38	38

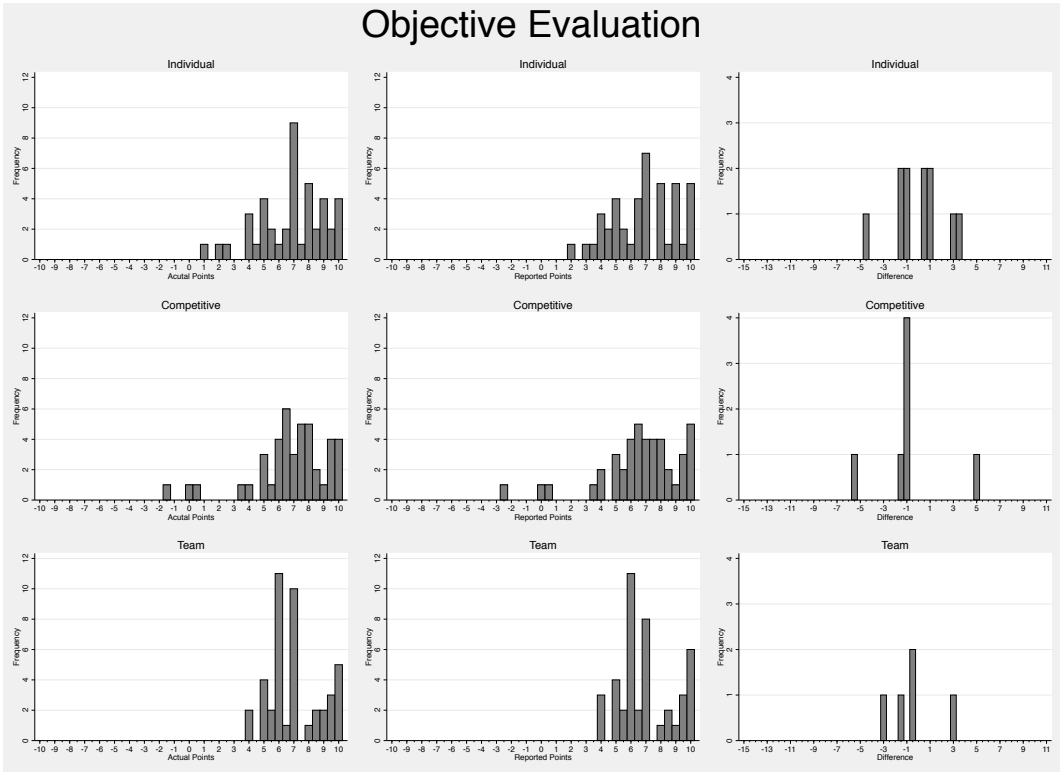
[§] Since two subjects left the experiment before completing their reports for the collective stage, entries for this stage in treatment *Objective* are for N=43.

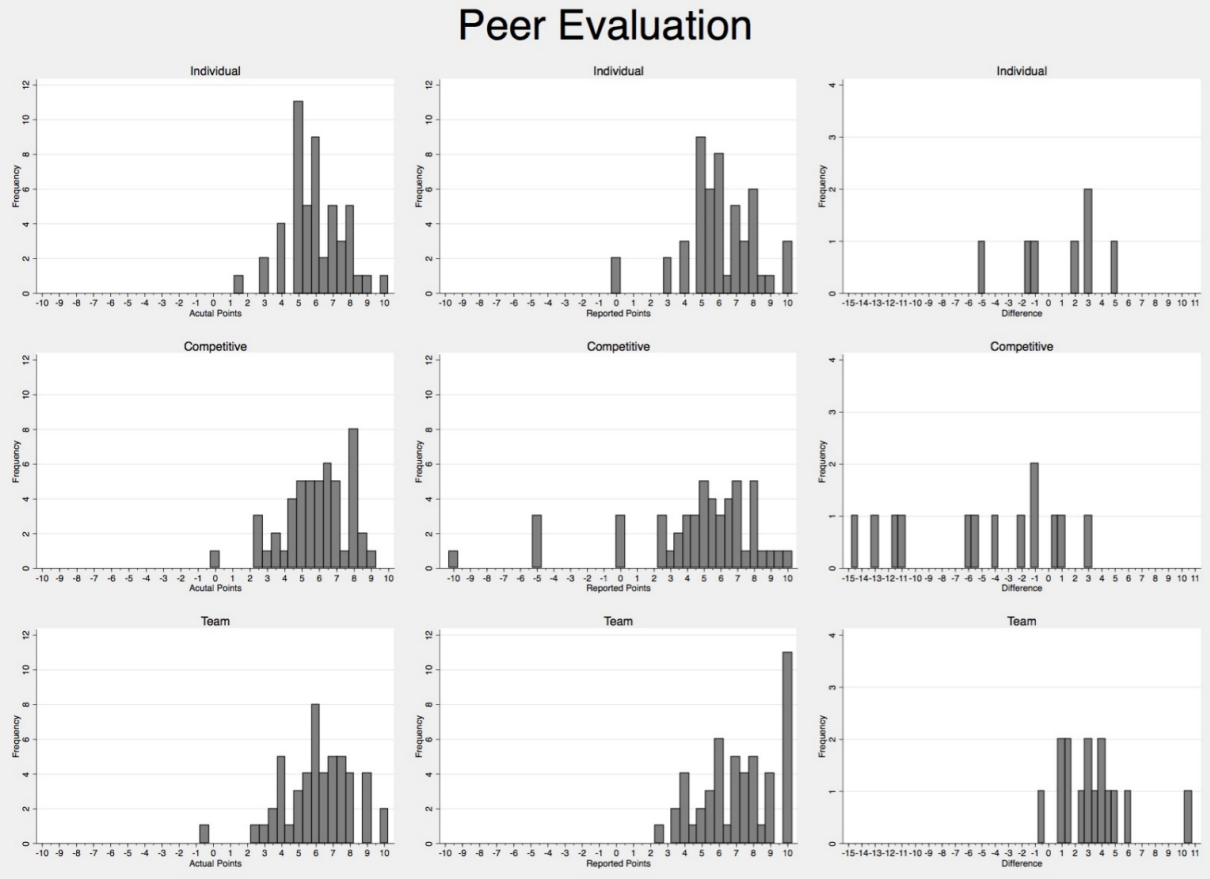
Table 4: Regression analysis

Dep. variable	(1) reported – actual points	(2) reported – actual points	(3) reported – actual points	(4) reported – actual points	(5) actual points	(6) actual points
Treatment	Both	Both	*Objective*	*Peer*	Both	Both
other perf.	-0.135*	-0.135*	-0.081	-0.129		
	(0.073)	(0.074)	(0.068)	(0.182)		
own perf.	0.094	0.163*	0.015	0.398*		
	(0.063)	(0.082)	(0.060)	(0.205)		
comp. stage	-0.159	-0.256	-0.224	-1.662***	-0.078	0.141
	(0.206)	(0.245)	(0.233)	(0.615)	(0.391)	(0.392)
team stage	-0.304	-0.303	-0.254	1.004**	0.225	0.269
	(0.213)	(0.223)	(0.226)	(0.391)	(0.266)	(0.306)
peer reporting	0.056	0.038			-0.773*	-0.881*
	(0.221)	(0.280)			(0.409)	(0.454)
comp. x peer	-1.255**	-1.401**			-0.012	-0.116
	(0.537)	(0.657)			(0.483)	(0.516)
team x peer	1.212***	1.367***			-0.025	-0.019
	(0.363)	(0.431)			(0.381)	(0.437)
female		0.434*	0.209	0.289		-0.069
		(0.241)	(0.334)	(0.383)		(0.415)
MTG		-0.014	-0.473	0.928		-0.529
		(0.377)	(0.411)	(0.839)		(0.417)
risk measure		0.087	0.067	0.057		0.015
		(0.087)	(0.078)	(0.135)		(0.102)
comp. measure		0.248	0.006	0.669*		-0.258
		(0.168)	(0.131)	(0.354)		(0.161)
variable wage		-0.031	0.218	0.239		0.094
		(0.245)	(0.322)	(0.398)		(0.494)
wage level		0.036	0.209	-0.368*		0.278**
		(0.115)	(0.129)	(0.217)		(0.138)
constant	0.295	-1.504	-0.494	-3.284	6.733***	6.662***
	(0.619)	(1.135)	(0.906)	(2.658)	(0.337)	(0.917)
N	283	235	115	120	283	235

Notes: Ordinary least squares regressions. Standard errors in parentheses, clustered by participant. All regression specifications include participant random effects. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure 1: Distribution of actual points (first column), reported points (second column) and (non-zero) differences between reported and actual points (third column), by treatment and stage





Notes: The first row in each graph corresponds to the individual stage, the second row to the competitive stage and the third row to the team stage. The first column in each row shows actual points, the second column shows reported points, and the third column shows the difference, (reported points – actual points). For ease of exposition, in the third column (differences) we omit values of zero.

Appendix

Figure A.1: Example of the real-effort task

	Zahl 1	Zahl 2	Zahl 3	Ergebnis	Falsch ?
1.	56	61	15	132	
2.	79	20	94	192	
3.	16	77	28	121	
4.	55	60	42	157	
5.	65	55	10	130	
6.	20	67	36	113	
7.	89	99	65	253	
8.	35	77	98	120	
9.	43	20	59	122	
10.	82	51	33	164	
11.	39	47	48	134	
12.	23	30	47	100	
13.	39	11	79	130	
14.	15	39	33	87	
15.	52	28	97	177	
16.	41	60	19	120	
17.	62	12	46	140	
18.	20	37	88	145	
19.	40	50	84	174	
20.	45	55	12	112	
21.	82	90	24	195	
22.	79	45	74	189	
23.	68	63	40	171	
24.	66	24	96	186	
25.	33	77	24	134	
26.	17	64	70	151	
27.	53	35	88	166	
28.	34	15	16	65	
29.	51	48	46	144	
30.	85	11	36	132	

Gesamtpunkteanzahl

Table A.1: Additional regressions for relative misreporting

	(1)	(2)
Dep. variable [§]	Relative misreporting	Relative misreporting
other perf.	-0.941	-0.973
	(0.833)	(0.969)
own perf.	0.828	1.447
	(0.837)	(1.011)
comp. stage	-3.898	-5.041
	(4.894)	(5.556)
team stage	-4.300	-4.594
	(4.968)	(5.654)
peer reporting	0.338	-0.525
	(5.533)	(6.422)
comp. x peer	-6.196	-6.022
	(6.745)	(7.801)
team x peer	22.262***	25.567***
	(6.751)	(7.885)
female		2.892
		(5.012)
MTG		-7.353
		(5.645)
risk measure		0.393
		(1.182)
comp. measure		0.014
		(2.316)
variable wage		4.614
		(5.83)
wage level		1.364
		(1.730)
constant	5.595	-5.168
	(8.751)	(16.643)
N	283	235
comp. stage + comp. x peer	-10.093 **	-11.062 **
team stage + team x peer	17.962 ***	20.973 ***

Notes: Tobit regressions. Standard errors in parentheses. All regression specifications include participant random effects. *** $p < 0.01$. The last two rows report significance based on χ^2 tests on the restriction that the respective joint coefficient is equal to zero.

[§] Dependent variable is relative misreporting in %, defined as the difference between reported and actual points divided by the maximum possible difference in each observation and multiplied by 100.