

1-10-2019

On the Merit of Equal Pay: Performance Manipulation and Incentive Setting

Brice Corgnet

Ludivine Martin

Peguy Ndodjang

Angela Sutan

Follow this and additional works at: https://digitalcommons.chapman.edu/esi_pubs

 Part of the [Economic Theory Commons](#), and the [Other Economics Commons](#)

On the Merit of Equal Pay: Performance Manipulation and Incentive Setting

Comments

NOTICE: this is the author's version of a work that was accepted for publication in *European Economic Review*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in *European Economic Review*, volume 113, in 2019. DOI: [10.1016/j.euroecorev.2018.12.006](https://doi.org/10.1016/j.euroecorev.2018.12.006)

The Creative Commons license below applies only to this version of the article.

Creative Commons License



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License](https://creativecommons.org/licenses/by-nc-nd/4.0/).

Copyright

Elsevier

On the Merit of Equal Pay: Performance Manipulation and Incentive Setting

Brice Corgnet , Ludivine Martin , Peggy Ndodjang , Angela Sutan

PII: S0014-2921(18)30200-9
DOI: <https://doi.org/10.1016/j.euroecorev.2018.12.006>
Reference: EER 3209

To appear in: *European Economic Review*

Received date: 25 May 2018
Accepted date: 19 December 2018

Please cite this article as: Brice Corgnet , Ludivine Martin , Peggy Ndodjang , Angela Sutan , On the Merit of Equal Pay: Performance Manipulation and Incentive Setting, *European Economic Review* (2019), doi: <https://doi.org/10.1016/j.euroecorev.2018.12.006>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



On the Merit of Equal Pay:
Performance Manipulation and Incentive Setting

BRICE CORGNET, LUDIVINE MARTIN, PEGUY NDODJANG AND ANGELA SUTAN¹

Work performance is often difficult to assess thus leaving room for manipulation of commonly-used metrics. We created a laboratory workplace in which we can precisely assess both work performance along with manipulation activities. Using two independent experiments we show that, whenever pay for performance is used, manipulation is pervasive leading to both a waste of organizational resources and a weakening of incentives. By contrast, paying organizational members equally effectively deters manipulation attempts leading to higher organizational production.

(JEL C91, D23, D86, M52)

¹ Brice Corgnet, EM Lyon, GATE L-SE UMR 5824, 23 avenue guy de collongue; Ludivine Martin (Luxembourg Institute of Socio-Economic Research). Peguy Ndodjang, Université de Perpignan (France); Angela Sutan : Univ. Bourgogne Franche Comté, Burgundy School of Business-CEREN (EA 7477), LESSAC, 29 rue Sambin, 21000 Dijon, France. Brice Corgnet acknowledge the continuous support of the Economic Science Institute at Chapman University. Brice Corgnet also acknowledges the support of LABEX CORTEX (Univ Lyon). Ludivine Martin, Angela Sutan and Peguy Ndodjang acknowledge the support of the Fonds National de la Recherche, Luxembourg for the TWAIN project (contract FNR/C11/LM/1196209). Ludivine Martin also acknowledges the continuous support of LISER and of its core funding from the Ministry of Higher Education and Research of Luxembourg.

1. Introduction

The economic literature on incentives puts forward that a key aspect of successful organizations is their ability to reward each member according to their actual contribution (e.g., Holmström, 1979; 1982; Laffont and Martimort, 2002). Doing so, however, requires precise information about individual achievements which is, often, not readily accessible to managers. Managers must thus partly rely on organizational members' reports of their personal contribution and such reports are likely to be biased (Prendergast and Topel, 1993). Because presenting oneself in a positive light will typically bring monetary benefits, organizational members are likely to engage in '*window-dressing*' activities so as to appear as high contributors. These *window-dressing* activities consist in making one's contribution appear deceptively favorable. These activities which have been extensively documented in the finance and accounting literatures (see Lakonishok et al., 1991) typically result from hiding relevant pieces of information or falsifying documents (Maggi and Rodríguez-Clare, 1995; Crocker and Morgan, 1998; Crocker and Slemrod, 2007; Crocker and Gresik, 2011; Beyer et al., 2014; Roger, 2014). In the economics literature on influence costs, *window-dressing* can also take the form of blatant manipulation of managers' opinions (Milgrom, 1988; Milgrom and Roberts, 1988, 1992; Corgnet and Rodriguez-Lara, 2013; Powell, 2015). Regardless of the strategy used to distort performance measures, the organization will be negatively impacted by these performance manipulation activities as they waste time and distort incentive schemes (see Milgrom and Roberts, 1988; 1992). Performance manipulation will reduce the quality of information available to managers thus ultimately weakening the correlation between compensation and performance measures (e.g., Holmström, 1979). Performance manipulation is thus one possible explanation for the surprising limited strength of incentives in actual labor contracts (e.g., Baker et al., 1988; Chiappori and Salanié, 2000).

Because performance manipulation can only be effective when hidden from management and because such activities could be severely punished, it is not surprising that no reliable archival data can be obtained (Powell, 2015). This prevents the testing of the main conjectures of the different strands of the literature on performance manipulation which highlight the negative effect of performance manipulation on organizational performance and the necessity to weaken incentives to prevent such activities. In particular, one important insight of the influence costs literature is to show that, in the presence of performance manipulation attempts, it may be optimal for organizations to limit managers' discretion

over decisions affecting the distribution of resources. This implies that firms may adopt apparently wasteful bureaucratic rules (Bloom and Van Reenen, 2007) for efficiency concerns (Milgrom, 1988; Milgrom and Roberts, 1988, 1992; Powell, 2015). For example, firms may avoid discrete bonuses to limit influence activities (Fairburn and Malcomson, 2001). Relatedly, Milgrom and Roberts (1992) suggested that the extensive use of equal pay may be justified, despite its negative incentive effects, as a way to limit influence costs:

“One clear way to limit the competition for rents is to equalize their distribution across potential competitors, or at least limit the possible differentials. (...) The cost of the policy is that informational and incentive roles of rewards are muted by closing differentials.” (Milgrom and Roberts, 1992, page 274).

Our aim was to collect data to highlight the tradeoff between the positive incentive effect of pay for performance and its damaging effect on promoting *window-dressing*. To produce data on performance manipulation in an organizational context and thus fill the empirical gap in the literature, we created a laboratory workplace environment in which workers can, instead of completing a real-effort task, engage in a time-consuming activity that will raise their performance metric. Our aim here was to develop a computerized environment that would incorporate several features of a real-world workplace in the spirit of early real-effort experiments in the economics literature such as Dickinson (1999) and van Dijk et al., (2001).

In our setting, workers have to select a performance manipulation option (displayed on their screen) to exaggerate their actual contribution on the real-effort task. Doing so would freeze their screen for a duration of 30 seconds during which they could not engage in any activity. The frozen screen aims at representing the waste of time typically involved in performance manipulation whether it is related to the falsification of documents or to influence activities. As is described by Milgrom (1988, page 43), *“That time of course is valuable; if it were not wasted in influence activities, it could be used for directly productive activities or simply consumed as leisure.”*

These influence activities contrast with bribing activities (see e.g., Gneezy et al., 2018) as they do not involve any monetary payments thus being difficult to detect. This might explain why influence activities are likely to be pervasive within firms whereas bribing activities are not.

To assess the causal effect of performance manipulation on organizational performance, we consider two experiments in which workers can or cannot engage in *window-dressing* activities.

In our first experiment, we consider organizations composed of four workers without a manager. In that setting, *window-dressing* activities were either available or not, and workers' compensation was either an equal share (25%) of organizational production (*equal pay* treatments) or a share of organizational production which was based on workers' *observed* contribution (*observed-performance pay* treatments). *Equal pay* treatments thus provide workers with team incentives. When *window-dressing* activities were not available, *observed* contribution was equal to actual contribution in which case workers were paid as in an individual piece-rate incentive scheme. When *window-dressing* activities were available, they allowed workers to increase their *observed* contribution and claim a higher share of organizational production.

We find that *window-dressing* is much more prevalent under *observed-performance pay* than under *equal pay*. Under *observed-performance pay*, almost all workers engaged in performance manipulation spending on average 13.8% of their time with a frozen screen compared to 2.0% under *equal pay*. It follows that when *window-dressing* activities were available organizational production was significantly (28.7%) higher under *equal pay* than under *observed-performance pay* whereas the opposite was true when *window-dressing* activities were not available. This means that the beneficial effect of *equal pay* on deterring *window-dressing* activities more than offset its negative incentive effect. The participants to Experiment 1 completed a battery of tests when signing up to the lab thus allowing us to study the individual determinants of *window-dressing* activities. Our results provide insights to potential recruiters who want to avoid wasteful performance manipulation activities while relying on performance pay to motivate workers. We show that more spiteful and more cognitively reflective workers engage more often in performance manipulation and produce less in the *observed-performance pay* treatment in which *window-dressing* is available. The fact that workers who score high on the cognitive reflection scale produce less in our setting is particularly striking given that cognitive reflection is commonly associated with high performance (e.g., Corgnet et al., 2015e). In addition, cognitive reflection significantly and positively correlates with general intelligence which has been shown to be the best predictor of work performance (e.g., Schmidt, 2009) as well as diligence on the job (e.g., Corgnet et al., 2015c). For example, in the *observed-performance pay* treatment in which *window-dressing* is not available, cognitive reflection relates positively and significantly with production.

In Experiment 2, we extend Experiment 1 by setting up an organizational environment in which three workers are under the supervision of one manager. Our aim was to assess the extent to which managers can reduce performance manipulation by adjusting workers' incentive schemes. Workers' compensation

was either an equal share (20%) of organizational production (*equal pay* treatments) or a share of organizational production chosen at the discretion of the manager at the end of each period (*discrete pay* treatments). Regardless of the treatment, the manager always received a fixed share (40%) of organizational production. Experiment 2 differs from Experiment 1 as the manager could monitor in real-time the evolution of workers' performance thus being able to detect excessive *window-dressing* activities which would lead to unlikely high levels of observed performance. Thus, Experiment 2 allows us to study monitoring as a possible mechanism to tamper performance manipulation. In addition, in Experiment 2 managers could limit the incentives to engage in performance manipulation by deliberately paying every worker the same or more generally by making pay less sensitive to performance. Not surprisingly, we report less *window-dressing activities* in Experiment 2 than in Experiment 1.

Nevertheless, we replicated the main findings of the first experiment regarding the relationship between incentive schemes, *window-dressing* activities and organizational production. In particular, in the *discrete pay* treatments, organizational production was significantly lower when *window-dressing* activities were available than when they were not. We reported no significant differences in organizational production between the *equal pay* treatments. In Experiment 2, we obtained further insights on workers' *window-dressing* activities in a context in which these activities could be detected by a manager. We find, unlike Experiment 1, that workers mostly engage in minimal *window-dressing* activities exaggerating their performance metric by the minimum possible amount so as not to be detected cheating. We find that *window-dressing* activities led to an increase in workers' pay for those who exaggerated their production by the minimum possible amount. By contrast, *window-dressing* activities led to a decrease in workers' pay for those who exaggerated their production by a larger amount.

Finally, Experiment 2 also allowed us to investigate how managers set incentives when performance manipulation is available to workers. We find that managers weakened incentives when *window-dressing* activities were available by reducing the link between *observed* performance and workers' pay compared to the case in which *window-dressing* activities were not available.

The paper proceeds as follows. In the next section, we review the related literature and discuss our contribution. In Section 3 we develop our conceptual framework and present our hypotheses. In Sections 4 and 5, we present the design and results of Experiments 1 and 2. Section 6 concludes.

2. Related literature

2.1. *Theory of the firm*

An essential ingredient of any theory of the firm is to provide a convincing explanation for the empirical observation that firms heavily rely on low-powered incentive schemes (Gibbons and Roberts, 2013). By showing that equal pay can improve upon performance pay in the presence of influence activities, we provide empirical support for the influence cost theory of the firm. Our work thus belongs to a small but growing literature providing causal tests of different theories of the firm.² These works are especially crucial given the general imbalance between the *many* theories of the firm and the *few* empirical tests (e.g., Baker and Holmström, 1995; Prendergast, 2002). Our findings can also be seen as providing empirical support for the multitasking theory of the firm (Holmström and Milgrom, 1991; Dewatripont et al., 2000) which has been growingly popular among organizational economists. Our organizational setting can indeed be described as a multitasking problem where workers can either exert effort to produce a valuable product or spend time producing a valueless product. We then show that, in a context in which the performance metric only consists in the volume of production, workers will inefficiently assign their time to the production of valueless products. This activity is what we have referred to as performance manipulation or *window-dressing*. The link between the multitasking and the influence costs theories of the firm was already established by Holmström and Milgrom (1991) as well as Baker (1992). More specifically, we show that equal pay can alleviate the misallocation of time due to multitasking issues thus ultimately leading to greater organizational performance than performance pay.

2.2. *Deception literature*

Starting with the seminal work of Gneezy (2005), deception games have been used extensively to study lying behavior. The deception game, inspired from the work of Crawford and Sobel (1982), is a two-player setting in which the first player possesses private information about the monetary consequences of the action taken by a second player. The first player can send a message to induce the second player to take the action that will be most favorable to the first player. Our work extends this

² For example, the hold-up theory of the firm developed by Klein et al. (1978) was investigated by Sloof et al. (2004; 2007), the property rights model of Hart (1995) was tested by Sonnemans et al. (2001) and the monitoring in teams model of Alchian and Demsetz (1972) was tested by Grosse et al. (2011). The theory of contracts as reference points (Hart and Moore, 2008) was tested in Fehr et al. (2011).

previous literature by studying performance manipulation in an organizational context. Two crucial features characterize our organizational environment. First, unlike deception games, our setting is not a zero-sum game as performance manipulation can hurt firm profits as it creates negative externalities by wasting time that could be dedicated to organizational duties (Experiments 1 and 2). Second, we consider a hierarchical setting in which a manager can monitor and decide upon subordinates' payments (Experiment 2). Managers can detect and sanction lying behavior thus tampering its negative effect on organizational performance. They can also weaken incentives to limit the monetary gains associated with misrepresenting one's own contribution.

Our organizational setting may affect prosocial motives (e.g., Gino and Pierce, 2009; Atanasov and Dana, 2011; Erat and Gneezy, 2012) or social identity (Shalvi and Leiser, 2013; Utikal and Fischbacher, 2013) thus impacting dishonest behavior. At the same time, an organizational setting may foster dishonest behavior as a few dishonest subjects may rapidly spread unethical behaviors in groups (e.g., Carrell et al., 2008; Gino et al., 2013). A priori, it is unclear whether our organizational context would lead to widespread dishonesty. However, we observe pervasive lying behavior in our setup confirming previous findings in deception games (see Gneezy, 2005) and in individual cheating tasks (Shalvi et al., 2011a; Fischbacher and Föllmi-Heusi, 2013). Our work not only studies dishonesty in an organizational context but it also investigates organizational mechanisms that can deter manipulation activities such as performance monitoring and incentive schemes. We find that both intensive monitoring and weak incentive schemes can effectively reduce the extent of performance manipulation activities.

Despite the large number of works studying honest behavior (e.g., Gino, 2015; Jacobsen et al., 2018), little is known on the individual determinants of lying behavior (Abeler et al., 2018; Capraro, 2017). In a meta-analysis of 50 deception games, Capraro (2017) identifies education level as a main predictor of using deceptive strategies. Abeler et al. (2018) combining data from 72 experimental studies show that the preference for being honest with a preference for being seen as honest reduce lying behavior. Cabrales et al. (2016) also show that more spiteful subjects tend to engage in more deceptive strategies. Our work complements these studies by showing that, in our organizational context, both spite and cognitive reflection (which correlates with IQ and educational levels, Frederick, 2005) predict lying behavior even after controlling for a wealth of individual characteristics including personality traits and risk attitudes.

2.3. *Cheating and sabotage in tournaments*

Related to our work is the experimental literature on cheating and sabotage in tournaments that puts forward the negative consequences of relative performance pay (e.g., Konrad, 2000; Preston and Szymanski, 2003; Harbring et al., 2007; Kräkel, 2007; Schwierien and Weichselbaumer, 2010; Harbring and Irlenbusch, 2011; Balafoutas et al., 2012b; see Chowdhury and Gürtler, 2015 for a review). Our set up differs from standard tournament settings which represent competitive environments rather than cooperative environments in which organizational members can mutually benefit from the high performance of other members. In our experiments, we consider an organizational setting in which organizational members' efforts determine the total amount of money which is available to reward workers. In contrast with tournament incentives which tend to exacerbate competition by clearly recognizing one contest winner (e.g., Sheremeta, 2016), our setup considers linear incentive schemes in which pay is proportional to performance metrics. It is worth noting that contests may share some of the organizational features of our setting as is, for example, the case of Tullock contests in which prizes depend on contestants' effort (see e.g., Chung, 1996; Chowdhury and Sheremeta, 2011; see also Chowdhury et al., 2014 for an experimental study). The organizational context we envision, unlike tournaments, may foster social motives thus leading workers to adopt a cooperative behavior because they value organizational norms (e.g., Kandel and Lazear, 1992) as well as care about the well-being of other members of the organizations (see e.g., Bandiera et al., 2010). This makes our findings regarding the pervasiveness of *window-dressing* activities especially striking.

Another important aspect of our work is to assess the effect of incentive schemes on performance manipulation. In that respect, the closest paper to ours is the work of Carpenter et al. (2010). In a real-effort experiment where workers had to prepare letters and envelopes, the authors compared workers' production in a tournament setting in which performance was assessed by a supervisor with a tournament in which performance was partly determined by peers. In their setting, performance manipulation followed from peers underreporting others' performance to increase their chance of winning the tournament prize. This led to a lower level of performance in the tournament with peer evaluations than in the tournament setting with supervisor evaluation. They compared workers' performance under tournament incentives with their performance under piece rates. The authors showed that when peers were in charge of reporting performance workers produced more under piece rates than under tournament incentives whereas the opposite was true when supervisors were in charge of performance reporting. In a recent study, Balafoutas et al. (2017) replicated these findings in a lab in the field experiment conducted with auditors at a major conference venue.

Instead of considering tournament incentives, we built on the literature in performance manipulation to assess the extent to which weakening incentives, for example using equal pay, can be beneficial to the organization by countervailing *window-dressing* activities. Our setting is thus one in which performance manipulation can arise even in the absence of relative pay. Balafoutas et al. (2017) also considered a team incentive setting in which workers could overreport others' performance to increase their partners' pay at the expense of the experimenter. Their setting differs from our experiments as workers do not have the opportunity to overreport their own performance at the expense of other workers. Importantly, in our experiments *window-dressing* activities are time consuming thus affecting negatively organizational performance and reducing the amount of money to be shared among workers. This is the case because we consider *window-dressing* activities during which workers can exaggerate their production and obtain a higher pay without engaging in sabotage. Another major difference between our setting and the works of Carpenter et al. (2010) and Balafoutas et al. (2017) is that we consider, in Experiment 2, the case in which managers can monitor workers and partly detect performance manipulation while having full discretion over workers' pay.

The paper of Charness et al. (2014) also relates to our study as it allows workers to engage in both sabotage and *window-dressing*. It differs from our study as the authors do not assess the effect of incentives. Instead, they deliberately chose a flat-wage environment to assess the willingness of workers to engage in personally costly sabotage and *window-dressing* for status-enhancing motives only.

Our main finding that equal pay outperforms performance pay is thus new to the experimental literature. Additionally, Experiment 1 allowed us to establish an individual profile of the typical performance manipulator whereas Experiment 2 showed how managers can use monitoring technologies to weaken incentives and limit *window-dressing* activities. Finally, Experiment 2 showed how workers, despite tight manager's supervision, were able to boost their pay by engaging in subtle performance manipulation. Importantly, both experiments show that the positive deterrence effect of equal pay on *window-dressing* more than offsets its negative incentive effect.

3. Conceptual framework

Our aim here is to illustrate, using a stylized model, the main implication of the literature on influence costs and performance manipulation regarding the detrimental effect of performance pay on organizational performance when *window-dressing* activities are available to workers. Our model is thus

not aimed at being an exact representation of our experimental framework. In particular, we consider a setup in which both effort and *window-dressing* are binary decisions.

We consider an organization which is composed of n workers who can either work or shirk as well as engage in performance manipulation activities. We will compare organizational performance under two contracts: one in which workers are paid based on their *observed* contribution to the organization (*observed-performance pay*) and one in which all workers are paid an equal share of organizational production (*equal pay*). We also incorporate social motives in our framework thus complementing our standard incentive theory toolkit (see e.g., Laffont and Martimort, 2002) with the insights of behavioral contract theory (e.g., Rotemberg, 1994; Grund and Sliwka, 2007; Bartling and von Siemens, 2010; Dur and Sol, 2010; Bierbrauer et al., 2017).

3.1. Workers

Worker $i \in \{1, \dots, n\}$ can either exert effort on the work task ($e_i = 1$) or shirk ($e_i = 0$). Working on the task implies a cost $c > 0$. Exerting effort will generate a positive output for the organization ($y_i = y$) with probability ρ and no output ($y_i = 0$) with probability $1 - \rho$. Exerting no effort always leads to no output. We assume that exerting high effort is socially efficient because the expected value created by workers when exerting effort is greater than its cost ($\rho y_i \geq c$). Instead of working, workers can engage in time-consuming *window-dressing* activities ($d_i = 1$) at a private cost χ or not do so ($d_i = 0$). Workers will thus choose whether to exert effort and whether to engage in *window-dressing* activities by maximizing their expected utility subject to the constraint that $e_i + d_i \leq 1$. This constraint follows from the fact that workers can either exert effort in which case *window-dressing* activities are useless ($e_i = 1, d_i = 0$), shirk while engaging in *window-dressing* activities ($e_i = 0, d_i = 1$) or not ($e_i = 0, d_i = 0$).

For *window-dressing* to be appealing, we assume it is less costly than completing the work task so that $\chi < c$. In addition to the private cost χ , *window-dressing* activities are socially costly because they prevent workers from completing the work task and generate output for the organization (see Milgrom, 1988). Nonetheless, workers may gain from *window-dressing* activities as they will positively affect the *observed* measure of output of their effort (\hat{y}_i). We assume that when worker i engages in *window-dressing*, the observed output of shirking will be equal to $\hat{y}_i = y$ with probability $\rho_d \leq \rho$. It follows that $\hat{y}_i := (e_i \rho + d_i \rho_d) y$.

In addition to observing \hat{y}_i , the manager knows the total output of the organization. However, because the manager cannot break the budget and is thus assumed to redistribute all the organizational output to workers, he or she cannot use the total output of the organization to prevent *window-dressing* activities by not paying workers if proof of *window-dressing* activities is obtained.

We assume workers to be risk neutral, abstracting away from risk-sharing issues, and focusing our attention on the effect of performance manipulation on contract design and organizational performance (see Milgrom and Roberts, 1988; Gibbons, 2005; Powell, 2015).

In addition to profit maximization, we consider workers' social motives which may include altruism (caring about others' payoffs), inequality aversion (preferring equal distribution of payoffs) or spite (valuing others' payoffs negatively) (e.g., Fehr and Schmidt, 2006; Balafoutas et al., 2012a).

As organizations represent a social context, we expect social motives to play an important role. For example, Rotemberg (1994) and Andreoni (1995) show that altruism may alleviate shirking in organizations. Because altruistic workers care about other organizational members' well-being, they may refrain from engaging in *window-dressing* activities as they ultimately affect others' pay negatively. By contrast, spiteful workers who are willing to earn more than others even at a cost for themselves or for the organization are likely to engage in *window-dressing* activities to boost their earnings. To study the impact of social preferences on *window-dressing* activities, we assume workers are endowed with the following utility function (à la Fehr and Schmidt, 1999) subject to $e_i + d_i \leq 1$:

$$(1) \quad U_i(e_i, d_i; c, \gamma_i^\alpha, \gamma_i^\beta) := w(\hat{y}_i) - ce_i - d_i\chi - \frac{\gamma_i^\alpha}{n-1} \sum_{j \neq i} \max\{w(\hat{y}_j) - w(\hat{y}_i); 0\} - \frac{\gamma_i^\beta}{n-1} \sum_{j \neq i} \max\{w(\hat{y}_i) - w(\hat{y}_j); 0\}$$

where $w(\hat{y}_i)$ stands for the pay of worker i given observed performance \hat{y}_i . The parameters γ_i^α and γ_i^β capture envy and shame when positive. Altruistic individuals are characterized by negative envy and shame ($\gamma_i^\alpha < 0, \gamma_i^\beta > 0$) whereas spiteful individuals are such that $\gamma_i^\alpha > 0$ and $\gamma_i^\beta < 0$ (see e.g., Corgnet et al., 2015a). Selfish workers are such that $\gamma_i^\alpha = \gamma_i^\beta = 0$.

3.2. Payment schemes

We now turn to the payment schemes offered to workers ($w(\hat{y}_i)$). We consider two types of contracts: *equal pay* and *observed-performance pay*. *Equal pay* contracts reward each worker an equal share of organizational production. In that setting, all workers are thus paid the same amount regardless of

observed performance so that there is no material incentive or social motive to engage in *window-dressing*. Under *observed-performance pay* contracts are based on the *observed* contribution of workers who are rewarded as follows:

$$(2) \quad w(\hat{y}_i) := \frac{\hat{y}_i}{\sum_j \hat{y}_j} Y$$

where Y denotes organizational production ($\sum_j y_j$) and $(\hat{y}_i / \sum_j \hat{y}_j)$ is the *observed* contribution of worker i . Regardless of the payment scheme, organizational production (Y) is observable by all workers.

In Appendix A, we formally derive two hypotheses.

Hypothesis 1 (Performance manipulation and organizational production)

- A) Workers engage in window-dressing activities under observed-performance pay but not under equal pay.
- B) Under observed-performance pay, the positive link between workers' pay and performance is reduced by the presence of window-dressing activities.
- C) Under observed-performance pay, organizational production is lower in the presence of window-dressing activities. Under equal pay, organizational production is not affected by the presence of window-dressing activities.

Hypothesis 1C stresses the negative consequences of window-dressing on organizational production in line with the influence costs literature (e.g., Milgrom and Roberts, 1988; Gibbons, 2005). It implies that equal pay will be an effective solution to thwart *window-dressing*. An empirical question is to assess whether the negative incentive effect of equal pay can be more than offset by its beneficial deterrence effect on *window-dressing*. That is, weaker incentives schemes may be optimal when *window-dressing* activities are available to workers.

Because *window-dressing* takes away resources from the firm and negatively affects other organizational members, altruistic workers are less likely to engage in such activities whereas spiteful workers are more likely to do so (see Appendix A). Because *window-dressing* activities prevent workers from completing the effortful task, the opportunity costs of such activities are the highest for high-skill workers. We derive the following hypothesis to capture these implications of our model.

Hypothesis 2 (Profiling the manipulators)

- A) More spiteful and less altruistic workers engage in more window-dressing activities.
- B) More skilled workers engage in less window-dressing activities.

Hypothesis 2A is similar to the conjecture derived in Grund and Sliwka (2007) in the context of tournaments in which they show that spiteful workers are more likely to engage in sabotage activities whereas compassionate workers are less likely to do so.

In Experiment 1, we test these hypotheses in a stylized work environment that captures the essential features of our model. In Experiment 2, we will consider an environment that extends our stylized model allowing for managers to monitor workers and set their pay.

4. Experiment 1

4.1. *Experimental design and procedures*

The virtual workplace.— We use an environment in which participants can undertake a real-effort task while having access to Internet (real-leisure alternative) at any point in time during the experiment (Corgnet et al., 2015d; 2015e). The introduction of Internet as the real-leisure alternative is motivated by the widespread use of Internet at work (e.g., Malachowski, 2005; Ugrin and Pearson, 2013). An appealing feature of Internet as an alternative to the work task is the wide range of activities that can be completed online.

We consider a particularly long and laborious summation task (e.g., Dohmen and Falk, 2011; Eriksson et al., 2009; Niederle and Vesterlund, 2007) so as to ensure that completing the *work task* required a substantial amount of effort. Participants were asked to sum up matrices of 36 numbers comprised between 0 and 3 for two periods of ten minutes each. The numbers in each table were generated randomly. Participants were not allowed to use a pen, scratch paper or calculator. This rule amplified the level of effort participants had to exert in order to complete tables correctly.

Each table completed correctly generated 40 cents of individual production while a penalty of 20 cents was subtracted from individual production for each incorrect answer. Penalties did not apply when individual production was equal to zero so that individual production could not be negative. Individuals could thus not sabotage the organization's output. Also, individuals could not suffer losses. Penalties were introduced so that participants could not generate monetary rewards by simply guessing the right answer.

At any point during the experiment, participants could switch from the work task to the leisure activity (browsing the Internet). Each activity was undertaken separately, in a different screen, so that participants could not complete tables while being on the Internet. Participants were informed that their use of the Internet was strictly confidential. Participants were free to consult their email or visit any web page. The Internet browser was embedded in the software so that the experimenter could keep record of the exact amount of time participants spent on each activity. At the beginning of each of the two periods, participants were randomly assigned to groups of four.

Payment schemes.—At the end of each period, participants were rewarded a share of the production of the organization defined as the sum of the individual production of the four organizational members. In the *equal pay* treatments, each of the four participants was rewarded an equal share of 25% of organizational production. In the *observed-performance pay* treatments, the share of organizational production assigned to workers was calculated as $\hat{y}_i / \sum_j \hat{y}_j$ where worker i 's observed performance (\hat{y}_i) is the sum of one's actual performance and the increase in observed performance obtained from *window-dressing* activities. Regardless of the payment scheme, organizational production (Y) was observed by all workers.

Window-dressing activities.—In the *window-dressing* treatments, participants had access to an additional activity which was referred to as *boost* in the experiment. This activity allowed workers to increase their observed level of performance by 40 cents every sixty seconds. This was calibrated so that an average-ability worker could achieve a higher level of observed performance by simply engaging in *window-dressing* activities. If workers only dedicated their time to *window-dressing* they could achieve a level of observed performance of 400 cents whereas average performance in the *observed-performance pay* treatment in which *window-dressing* was not available was 340 cents. After clicking on a confirmation button, the screen of the workers was frozen for 30 seconds. This aimed at representing the cost associated with time-consuming *window-dressing* activities that detract workers from the work task (Milgrom, 1988). This 30-second freeze represented 5% of the time available in a given period and about half the time an average-ability worker needs to complete one table correctly in the work task.³ Participants could easily keep track of their *window-dressing* activities in a given period because the total amount by which they exaggerated their production was recorded in the history panel at the bottom

³ This estimate was computed using data from Corbett et al. (2015b) who used the same task for a different experiment on goal setting.

of their screen. Workers could also keep track of their actual level of production during a given period. At the end of each period, participants were informed of their individual pay as well as organizational production.

In the *no window-dressing* treatments, workers could not exaggerate their production as the *boost* option was disabled. We conducted four treatments as part of a 2×2 factorial design (see Table 1). The complete set of instructions for Experiment 1 is available online (see Appendix D).

TABLE 1—FACTORIAL DESIGN

Number of participants (organizations) [sessions]		Payment Scheme	
		Observed-performance Pay	Equal Pay
<i>Window-dressing</i> Activities	Absent	56 (14) [7]	56 (14) [7]
	Present	56 (14) [7]	56 (14) [7]

Procedures.—Our participant pool consisted of students from a major US University. In total, 224 participants completed the experiments, divided in 28 sessions. We conducted seven sessions for each treatment. In each session, we had a total of 8 participants divided up in two organizations of four workers.

The experiment was computerized and all of the interactions were anonymous. The instructions were displayed on participants' computer screens. Participants had exactly 20 minutes to read the instructions (a 20-minute timer was shown on the laboratory screen).

All of the subjects who were recruited for this experiment already participated in a prior one-hour survey as part of the laboratory policy to collect individual information, including gender, summation skills and prosociality, about subjects who are registered in the pool (see Corgnet et al., 2018a). On average, participants to our experiment undertook this survey 6 months prior to the current experiment. In the “summation skills” task, participants were asked to sum as many five one-digit numbers as they could during two minutes. Each correct answer was rewarded 10 cents and the average earnings on this task were \$1.5. The number of correct answers is what we refer to as “summation skills”. Prosociality was measured using the elicitation task of Bartling et al. (2009) which was later extended by Corgnet et al. (2015a).

Participants were paid their earnings in cash. Individual earnings at the end of the experiment were computed as the sum of the earnings in the 2 periods. Participants earned on average \$13.4, including a show-up fee of \$7. In Experiment 1, the show-up fee corresponds to about 50% of total earnings which is undoubtedly high. However, the policy of the lab where the experiment was conducted requires using

the same show-up fee of \$7 for all experiments regardless of length. In Experiment 2, the show-up fee represents slightly less than 30% of total earnings. Experimental sessions lasted on average an hour.

4.2. Results

We use regression analyses to assess the statistical significance of our results. All reported p-values thus refer to regression results if not stated otherwise. We use panel regressions with random effects when assessing individual effects (Hypotheses 1A, 1B and 2) and conduct Ordinary Least Square (OLS) regressions when assessing organizational production (Hypothesis 1C). In each regression, we control for summation skills and gender. We thus consider a total of 112 (56 subjects \times 2 periods) observations for each of the treatments when studying individual effects. For organizational production, we consider a total of 56 independent observations. Our results are qualitatively unaffected by using non-parametric tests instead of regression analyses. Our methodological choice is to rely exclusively on regression results as they allow us to control for differences in ability levels across treatments. We proceed by testing each of our hypotheses.

Hypotheses 1A & 1B: Performance manipulation.—In line with Hypothesis 1A, we report that, under *observed-performance pay*, almost all participants (92.9%) engaged (at least once) in *window-dressing* activities in a given period compared to 33.3% in the case of *equal pay* (see Table 2). Under *observed-performance pay*, workers used *window-dressing* activities 5.5 times on average in a period compared to 0.8 times for *equal pay*. This implies that workers' screens were on average frozen for 13.8% of their available time under *observed-performance pay* compared to only 2.0% for *equal pay*. Thus, participants exaggerated their production on average by 170.6% under *observed-performance pay* compared to 15.5% under *equal pay*.

TABLE 2—SUMMARY OF WINDOW-DRESSING ACTIVITIES FOR BOTH THE OBSERVED-PERFORMANCE PAY AND THE EQUAL PAY TREATMENTS

Window-dressing Activities	Payment Scheme		p-value ^a
	Observed-performance Pay	Equal Pay	
Proportion of workers who engaged in <i>window-dressing</i> activities in a given period	92.9% (25.9%)	33.3% (47.2%)	<0.001
% of available time dedicated to <i>window-dressing</i> activities	13.8% (8.5%)	2.0% (4.2%)	<0.001
% of exaggeration of performance	170.6% (219.1%)	15.5% (34.9%)	<0.001

Notes: Standard deviations (in parentheses).

⁴ These p-values are computed using a panel regression with random effects and robust standard errors clustered at the organization level. The reported p-values correspond to the coefficient of the *discrete pay dummy* (which takes value one for *observed-performance pay* treatments and value zero for *equal pay* treatments). We use a gender dummy and summation skills as controls. Similar results are obtained using standard parametric and non-parametric tests.

In line with Hypothesis 1A, we show that *window-dressing* activities were pervasive and substantially more pronounced under *observed-performance pay* than under *equal pay*. Under *equal pay*, *window-dressing* activities are not totally absent, however. Participants may have felt curious to try, at least once, the *window-dressing* option which was available on their screen. This interpretation is consistent with the fact that, under *equal pay*, more than half (51.3%) of those who engaged in *window-dressing* activities only did it once compared to only 8.7% under *observed-performance pay*. Also, we note that the number of times participants engaged in *window-dressing* decreased from 0.91 in the first period to 0.69 in the second period under *equal pay* (Sign rank test, p-value = 0.008) whereas it increased from 4.8 to 6.1 under *observed-performance pay* (Sign rank test, p-value = 0.007).⁴

In line with Hypothesis 1B, we show that the relationship between a worker's pay and a worker's contribution is significantly weaker in the presence of *window-dressing* activities than in their absence. Indeed, in the absence of *window-dressing*, a \$1 increase in one's performance increases pay by exactly \$1 compared to only \$0.58 in the presence of *window-dressing* (see Table B.1 in Appendix B, p-value < 0.001 for testing the equality of coefficients, using a Wald test). The relationship between performance and pay in the presence of *window-dressing* is, however, significantly higher under *observed-performance pay* than under *equal pay* (p-value < 0.001).⁵ This means that, despite the presence of *window-dressing* activities, observed performance still incorporated some valuable information about individual contribution.

Hypothesis 1C: Organizational production.—Under *observed-performance pay*, workers dedicated 13.8% of their time to *window-dressing* thus leaving only 86.2% of their time available for completing the task compared to 98% under *equal pay*. In the treatments in which *window-dressing* was not available, workers could potentially dedicate all their time to the completion of the task. Because workers spent a very limited share of their time to browse the Internet regardless of the treatment, almost all the time workers did not spend on *window-dressing* activities was dedicated to the work task.⁶ We thus expect, in line with Hypothesis 1C, that under *observed-performance pay* organizational production

⁴ Similar results are obtained when using panel regression analyses with robust standard errors clustered at the organizational level.

⁵ This is the p-value corresponding to the Wald test that the *observed performance* coefficient is equal to 0.25. Indeed, by construction, in *equal pay* treatments a \$1 increase in one's production will increase pay by \$0.25.

⁶ Workers spent on average 2.5% of their time on the Internet across all treatments. The overall time spent online by each worker did not vary across treatments as the six pairwise comparisons across treatments produced p-values greater than 0.10 for the corresponding Wilcoxon Rank Sum Tests.

will be lower when *window-dressing* activities are available than when they are not. We expect no differences in production under *equal pay* whether *window-dressing* activities are available or not. Our findings support this claim as is illustrated in Figure 1 below. Under *observed-performance pay*, organizational production is 26.3% lower when *window-dressing* activities are available than when they are not (p-value < 0.001). About half of this effect (13.8%) can be attributed to the time wasted in *window-dressing* activities which could have been used to undertake the task. The other half of the effect can be attributed to the weakening of incentives in the *window-dressing* treatment (Hypothesis 1B).

Under *equal pay*, organizational production is not significantly different whether *window-dressing* activities are available or not (p-value = 0.361).

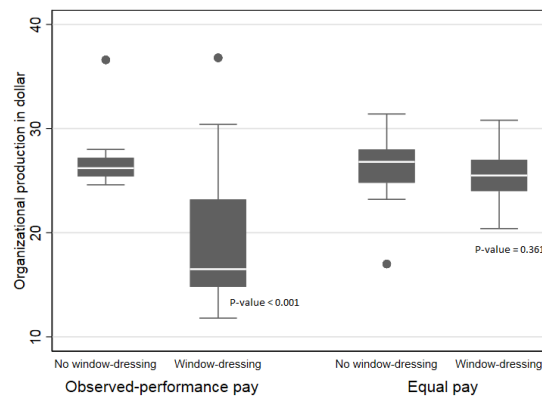


FIGURE 1. Box-plots for organizational production across treatments^A

Notes: ^a P-values are computed using OLS regressions with robust standard errors. The reported p-values correspond to the coefficient of the window-dressing dummy (which takes value one for treatments in which window-dressing is available and value zero otherwise). We control for gender and summation skills. These variables are defined as organizational levels averages.

We also put forward that, in the presence of *window-dressing* activities, organizational production is 28.7% higher under *equal pay* than under *observed-performance pay* (p-value = 0.003). By contrast, in the absence of *window-dressing* activities, organizational production is slightly, though not significantly, higher under *observed-performance pay* than under *equal pay* (p-value = 0.322). However, organizational production under *observed-performance pay* and *no window-dressing* activities is significantly greater when compared to *both equal pay* treatments (p-value = 0.076).

These results echo an important finding in the influence costs literature according to which limiting managerial discretion by relying on equal pay can be an effective response to influence activities. Thus,

weaker incentive schemes such as *equal pay* are likely to be popular in organizations as they lead to higher production levels than performance pay when performance manipulation is possible.

Hypothesis 2: Individual profiling of performance manipulators.—We test Hypothesis 2 according to which *window-dressing* activities and workers' production are affected by workers' social motives and skills. To assess social motives, we use the elicitation task of Bartling et al. (2009) (and extended by Corgnet et al., 2015a). In particular, we define a prosocial index which measures the extent to which a person values others' payoffs positively. This index takes values from 0 to 6, where the lowest value corresponds to spitefulness (a person who values negatively others' payoffs) and the highest possible value corresponds to altruism (see Table C.3 in online Appendix C). We obtain similar results when using envy and negative envy to assess spitefulness and altruism, respectively. To isolate the respective effects of spitefulness and altruism, we define a spiteful dummy (altruism dummy) which takes value one whenever a person's score is in the bottom quartile (top quartile) of the prosocial index.⁷ Because of the highly negative and significant correlation between spitefulness and altruism ($\rho = -0.514$, p-value < 0.001), we conduct separate regressions for these two social motives to avoid collinearity issues. In line with Hypothesis 2, we report that workers' summation skills negatively affect *window-dressing* (see Table 3). In addition, we show that prosociality reduces *window-dressing* activities whereas it affects production positively. Also in line with Hypothesis 2, spitefulness relates positively to *window-dressing* whereas altruism has the opposite effect. Spitefulness also significantly reduces production whereas altruism has a positive effect on production (although this effect is not statistically significant, p-value = 0.365).

A plausible mechanism by which spite could have fostered *window-dressing* activities hinges upon workers' negative reciprocity toward those organizational members who are believed to engage in *window-dressing*. The positive and significant coefficient associated to the variable period in Table 3 suggests that workers may have started with an optimistic belief regarding the extent of others' *window-dressing* activities which was then revised upward thus leading to an increase in the magnitude of *window-dressing* activities over time.

TABLE 3—PANEL REGRESSION WITH RANDOM EFFECTS FOR WINDOW-DRESSING (IN SECONDS) AND WORKERS' PRODUCTION (IN CENTS) IN THE OBSERVED-PERFORMANCE PAY TREATMENT WITH WINDOW-DRESSING

Dependent	Window-dressing	Production
-----------	-----------------	------------

⁷ Our findings are robust to alternative definitions of our spiteful and altruism dummies such as considering the top and bottom 10% values of the prosocial index as thresholds.

variable						
Intercept	185.53*** (69.18)	112.29*** (68.67)	162.65*** (67.08)	-59.41 (63.04)	25.15 (50.85)	-33.93 (59.09)
Prosocial index	-17.19*** (6.48)	-	-	15.88** (2.95)	-	-
Spiteful dummy	-	56.14* (30.64)	-	-	-76.87*** (25.57)	-
Altruism dummy	-	-	-39.42** (20.48)	-	-	22.693 (25.08)
Period	50.71*** (18.32)	50.71*** (18.32)	50.71*** (18.32)	19.29 (14.92)	19.29 (14.92)	19.29 (14.92)
Summation Skills	-3.59* (2.07)	-3.98** (1.67)	-3.69* (2.14)	15.21*** (2.96)	15.73*** (2.34)	15.27*** (3.09)
Gender Dummy (1 if female)	-48.43 (36.07)	-55.88 (35.346)	-37.95 (37.02)	13.06 (30.11)	24.54 (28.29)	2.81 (31.53)
Observations (organizations)	n = 112 (14)	n = 112 (14)	n = 112 (14)	n = 112 (14)	n = 112 (14)	n = 112 (14)
R ²	0.115	0.113	0.096	0.420	0.447	0.401

Notes: Estimation output using robust standard errors clustered at the organization level (in parentheses).

*** Significant at the 1 percent level; ** Significant at the 5 percent level; * Significant at the 10 percent level.

The one-hour survey completed by the workers participating in Experiment 1 included more individual measures than the ones reported in Table 3 such as personality traits, fluid intelligence, risk attitudes or cognitive reflection (see online Appendix C for a full description). We assess the potential explanatory power of these measures on *window-dressing* and production in Table B.2 in Appendix B. We first find that our previous findings regarding the importance of skills and prosocial concerns is robust to including additional individual controls.⁸ The only variable that significantly predicts *window-dressing* and production in addition to prosociality and skills is cognitive reflection. This result is statistically significant at the 1% level despite our analysis controlling for a standard cognitive skills measure such as the Raven test score. This suggests that it is cognitive reflection and not standard cognitive skills that can explain workers' behavior in our environment. Cognitive reflection is a good proxy for effortful and deliberative thinking (Toplack et al., 2011; 2014), often referred to as System 2 (Kahneman, 2011). Our results show that in addition to social motives and ability, a deliberative mindset may lead one to engage in more *window-dressing* activities thus contributing less to organizational performance. This result is consistent with Capraro (2016) who shows a negative correlation between lying in deception games and cognitive reflection scores. Our findings are consistent with the Truth-Default-Theory which posits that honesty comes naturally while dishonesty requires deliberation (see Verschuere and Shalvi, 2013). Relatedly, Gino and Ariely (2012) found that creative thinkers which

⁸ Even though we only present the results using the *prosocial* index as independent variable, our results are robust to considering spiteful and altruism dummies as independent variables (see Table B.3 in Appendix B) as well as controlling for additional individual characteristics (see Table B.4 in Appendix B).

have been found to exhibit higher cognitive reflection scores (see Corgnet et al., 2015a) are more inclined to engage in more dishonest behaviors.

Our results may seem to contrast with recent findings using time pressure manipulations showing that dishonest behaviors are intuitive rather than deliberative (e.g., Mead et al., 2009; Gino et al., 2011; Shalvi et al., 2012). However, these works make use of individual cheating tasks such as the ‘die-under-cup task’ in which participants privately roll a die and report, truthfully or not, the outcome which will then determine their pay. We believe our findings show that, in a more complex organizational environment than individual cheating tasks, lying may require deliberation. It may also be that trait-level deliberation (as measured by cognitive reflection scores) leads to different results than state-level deliberation induced by time pressure manipulations (Corgnet et al., 2016).

The fact that reflective workers produce less in our setting is particularly striking given that cognitive reflection has been commonly associated with high performance (e.g., Burks et al., 2015; Corgnet et al., 2015c). Our findings also challenge the idea that more reflective people are generally willing to engage in behaviors that benefit the group (e.g., Corgnet et al., 2015a; Ponti and Rodriguez-Lara, 2015). For example, in the *observed-performance pay* treatment in which *window-dressing* is not available, cognitive reflection relates positively and significantly with production (see Table B.3 in Appendix B). To our knowledge, our results are the first to show that the presence of reflective workers may be detrimental to organizational production.

Our analysis of the individual determinants of performance manipulation can be used as guidance for recruiters who want to limit wasteful *window-dressing* activities inside their organizations.

In the short run, the effectiveness of selecting specific workers to thwart performance manipulation can only be limited. In line with our findings, organizations could then rely on equal pay to deter performance manipulation. We investigate this possibility in more detail in our second experiment by assessing the extent to which managers can reduce performance manipulation by supervising workers and adjusting their pay accordingly. To that end, we extend Experiment 1, by considering a workplace environment that goes beyond our stylized model and include relevant features of real-world organizations such as the presence of a manager who will be able to monitor workers’ performance as well as decide upon their compensation.

5. Experiment 2

In this experiment, we consider a workplace environment which differs from Experiment 1 as each organization of four members will now consist of three employees and one manager. Our aim is to test the robustness of our findings in a more complex environment that endows managers full discretion over pay (in the *discrete pay* treatments) while allowing them to monitor employees' performance in real time. We briefly describe our workplace environment below. In the online Appendix E, we extend the model presented in Section 3 to the case in which a manager can set contracts to agents. We derive a similar conjecture to Hypothesis 1 which is stated as Hypothesis 1E.

5.1. *Experimental design and procedures*

The virtual workplace.—We use the same virtual workplace platform as in Experiment 1. The main difference is that organizations of four members now include one manager and three employees. At the beginning of the experiment, participants were randomly assigned to one of these two roles. Unlike *Experiment 1*, participants kept the same role and the same partners for the whole duration of the experiment which consisted of five periods of 10 minutes. This was done so that managers could adjust their pay policy over time based on their observation of employees' behavior. A longer experiment was also considered to assess the persistence of the effects uncovered in Experiment 1. The spirit of the design of Experiment 2 was to extend Experiment 1 so as to test of the effect of payment schemes on the magnitude of *window-dressing* activities in a work environment which more closely mimics actual workplaces.

The manager.—In addition to working on the task and browsing the Internet, the manager could observe the value of individual production (in cents) of each of the three employees at any time during a period by accessing a separate window with a monitoring screen. At the end of each period, the manager received a monitoring summary which indicated the observed individual production of each of the employees the last time the principal accessed the monitoring screen. As a result, managers could easily obtain the maximum amount of information regarding employees' period production by accessing the monitoring screen in the very last seconds of each period. At the end of each period, the manager as well as each employee also received feedback regarding their own pay and the production of the organization.

Payment schemes.—At the end of each of the five periods, employees and managers were rewarded a share of the production of the organization. Regardless of the treatment, the manager was always rewarded 40% of organizational production. In the *equal pay* treatments, each of the three employees was rewarded an equal share of 20% of organizational production. In the *discrete pay* treatments, the manager chose how to allocate the remaining 60% of organizational production to the three employees. Regardless of the payment scheme, organizational production (Y) was observed by all workers.

Window-dressing activities.—As in Experiment 1, employees had access to an additional activity which was referred to as *boost* in the *window-dressing* treatments. Unlike Experiment 1, employees were given full discretion on the frequency at which and on the amount by which they wanted to exaggerate their observed performance.

At the end of each period, the manager knew the *actual* production of the organization. Thus, in cases in which employees engaged in *window-dressing* activities, the *actual* production of the organization differed from the sum of the *observed* individual productions in the manager's monitoring summary. In addition, individual *window-dressing* activities could have been detected by the manager if an employee chose a *boost* amount which was not a multiple of 30. This is the case because each correct table generated 60 cents while an incorrect answer implied a 30-cent penalty. The value of a correct table and the penalty associated to an incorrect one differs from Experiment 1 simply because of the need to calibrate participants' earnings so that they would on average be equal to the average hourly pay offered in the lab. Thus, choosing a *boost* amount which was not a multiple of 30 would have indicated to the manager that the employee had engaged in *window-dressing* activities. An employee who decides to engage in excessively large or frequent *window-dressing* activities may also be detected by the manager monitoring the worker's performance.

In the *no window-dressing* treatments, performance manipulation was not possible as the *boost* option was disabled. As in Experiment 1, we conducted four treatments as part of a 2×2 factorial design (see Table B.5). The complete set of instructions for Experiment 1 is available online (see Appendix D).

Procedures.—Procedures were similar to Experiment 1. Upon arrival at the lab, "summation skills" were assessed using the same two-minute task as for the one-hour survey used in Experiment 1. Participants earned on average \$25.1, including a show-up fee of \$7. Experimental sessions lasted on average two hours and a half. Unlike Experiment 1, the subjects participating in this experiment were not recruited

from the pool of subjects who undertook the one-hour survey that collected information about individual characteristics.⁹

5.2. Results

In this extended setting, we aim at testing Hypothesis 1A regarding the prevalence of *window-dressing* activities when managers can pay employees at their own discretion. We also want to assess the weakening of incentives when *window-dressing* is available (Hypothesis 1B). The weakening of incentives in the context in which managers have full discretion over workers' pay will, unlike Experiment 1, crucially depend on managers' pay decisions. We also test Hypothesis 1C regarding the negative effect of *window-dressing* activities on organizational production when managers have full discretion over pay.

Hypothesis 1A & 1B: Performance manipulation.—In line with Experiment 1, *window-dressing* activities are pervasive and significantly more pronounced under *discrete pay* than under *equal pay* (see Table 4). As is the case for Experiment 1, workers spent a very limited share of their time browsing the Internet regardless of the treatment. Workers spent on average 2.1% of their time on the Internet across all treatments.¹⁰

TABLE 4—SUMMARY OF WINDOW-DRESSING ACTIVITIES
FOR BOTH THE DISCRETE PAY AND THE EQUAL PAY TREATMENTS

Window-dressing activities	Payment Scheme		p-value ^a
	Discrete Pay	Equal Pay	
Proportion of workers who engaged in <i>window-dressing</i> activities in a given period	55.0% (49.8%)	35.8% (48.0%)	<0.001
% of available time dedicated to <i>window-dressing</i> activities	4.6% (8.2%)	3.1% (7.9%)	0.252
Median <i>window-dressing</i> amount	148.0¢ (175.4¢)	110.8¢ (171.4¢)	0.253
% of exaggeration of performance	45.7% (83.1%)	38.7% (62.9%)	0.652

Notes: Standard deviations (in parentheses).

^a Consistently with the rest of the results section, these p-values are computed using a panel regression with random effects and robust standard errors clustered at the organization levels. The reported p-values correspond to the coefficient of the *discrete pay dummy* (which takes value one for *discrete pay* treatments and value zero for *equal pay* treatments). We use a gender dummy and summation skills as controls. Similar results are obtained using standard parametric and non-parametric tests.

⁹ Because of funding and timing constraints, not all subjects in the pool of subjects where the experiment was conducted completed the one-hour survey.

¹⁰ The overall time spent online by each worker did not vary across treatments as the six pairwise comparisons across treatments produced p-values greater than 0.10 for the corresponding Wilcoxon Rank Sum Tests.

Under *equal pay*, *window-dressing* activities occupied employees for 3.1% of the time which is similar to the time dedicated by employees to *window-dressing* in Experiment 1 (2.0%). However, the percentage of time spent on *window-dressing* activities under *discrete pay* is substantially lower (4.6%) than the time dedicated to these activities in the *observed-performance pay* treatment in Experiment 1 (4.6% vs 13.8%). Also, only about half of the employees engaged in *window-dressing* activities in a given period under *discrete pay* compared to 92.9% in Experiment 1 in the *observed-performance pay* treatment. This suggests that the benefits of *window-dressing* activities were more limited in the presence of a supervising manager. This is likely due to the fact that *window-dressing* activities may be detected in Experiment 2 if they are undertaken too intensively.

To investigate that possibility in more detail, we classify *window-dressing* activities as either *detectable* or *undetectable*. Because each correct table generated 60 cents while an incorrect answer implied a 30-cent penalty, an employee's production level was necessarily a multiple of 30. An employee who decided to exaggerate his or her production by an amount which was not a multiple of 30 was thus engaging in *detectable window-dressing* activities. An employee who exaggerated his or her production by an excessively large amount could also be seen as engaging in *detectable window-dressing* activities.

Window-dressing activities which are very unlikely to be detected by the manager are such that employees exaggerated their period production by the smallest possible amounts (i.e., one or two correct tables).¹¹ These amounts were actually the ones which were most likely to be chosen by workers. As is shown in Figure 2, the large majority (68.2%) of *window-dressing* activities involved exaggerating one's own production by two tables or less. Employees who exaggerated their production by two tables or less increased their *observed* production by 26.6% on average. By contrast, employees who exaggerated their production by at least three tables increased their *observed* production by 131.3% on average.

¹¹ We refer to the smallest integer number of tables. In our setting, workers could also exaggerate their production by less than the monetary value of one table.

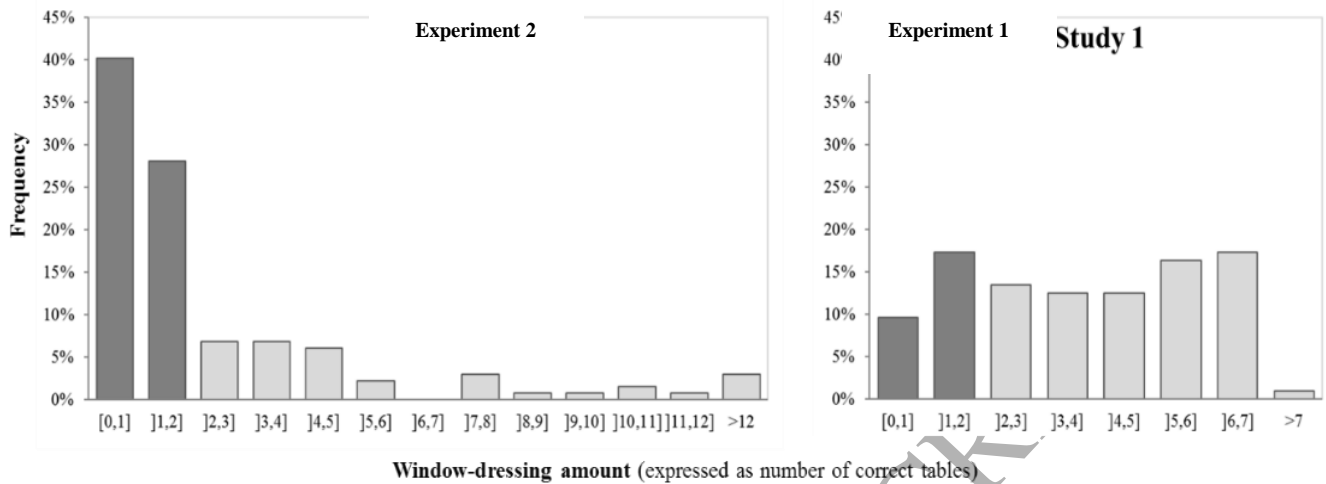


FIGURE 2. Histogram for the amount (expressed as the number of correct tables) by which a worker decided to exaggerate his or her observed production in a given period. experiment 2 is represented on the left panel and experiment 1 on the right panel.

The fact that most employees decided to exaggerate their production by the smallest possible amount is in line with the literature on cheating in individual tasks according to which people are willing to avoid major lies (e.g., Shalvi et al., 2011b). In our case, however, the main motive is to avoid detection of *window-dressing* activities by the manager. Indeed, when a manager is not present and cheating cannot be detected, then only a minority of *window-dressing* activities (26.9%) consisted in exaggerating one's own production by two tables or less (see right panel for Experiment 1 in Figure 2).

Next, we attempt to understand employees' caution to engage in *window-dressing* activities in Experiment 2 by analyzing the relationship between *window-dressing* activities and employees' pay. In Table 5, we regress an employee's pay on an independent variable that aims at capturing *undetectable window-dressing* activities ("*Two tables or less*" *window-dressing dummy*). This dummy variable takes value one whenever an employee exaggerates his or her production in a given period by a total amount either equal to 30¢, 60¢, 90¢ or 120¢ (that is, the monetary value of at most two correct tables, see column [1] in Table 5). In columns [2] and [3], we also include in our regression analysis a dummy variable that takes value one whenever an employee exaggerated his or her production by three tables or more ("*Three tables or more*" *window-dressing dummy*). Finally, we control for the actual contribution of the employee as it is supposedly an important determinant of pay even when *window-dressing* activities are available. We also control for employees' ability on the task using the "summation skills" measure and gender.

We report a positive and significant effect of *undetectable window-dressing* activities on employees' pay (see “*Two tables or less*” *window-dressing dummy* in columns [1] and [2] of Table 5). By contrast, we do not observe such positive effect of *window-dressing* activities when considering only those *window-dressing* activities by which employees exaggerated their production by three tables or more (see “*Three tables or more*” *window-dressing dummy* in columns [2] and [3] in Table 5). Our regression results thus suggest that an employee could increase his or her *discrete pay* by engaging in *window-dressing* activities as long as such activities were *undetectable*. Interestingly, summation skills appear to have no effect on employee's pay, regardless of the specification of the regression. That is, in the presence of *window-dressing* activities, ability on the task does not seem to be an important driver of employees' performance and of their subsequent pay. In Table B.6 in Appendix B, we provide additional analyses regarding the effect of *window-dressing* activities on pay. These findings show the robustness of the positive effect of *undetectable window-dressing* activities on pay. The effectiveness of *window-dressing* activities in boosting pay echoes the work of Robin et al. (2014). In their experimental setting, workers' pay was determined by managers after observing workers' performance and workers' opinions on non-work-related issues. The authors show that workers were able to positively influence their pay by expressing opinions which were similar to their bosses.

TABLE 5—PANEL REGRESSIONS ASSESSING THE EFFECT OF UNDETECTABLE WINDOW-DRESSING ACTIVITIES ON DISCRETE PAY FOR THE TREATMENT WITH WINDOW-DRESSING ACTIVITIES AND DISCRETE PAY

Dependent variable	Discrete pay (in %)		
	[1]	[2]	[3]
Intercept	21.883*** (2.724)	21.867*** (2.707)	22.987*** (2.862)
Period Trend	0.016 (0.042)	0.032 (0.059)	0.037 (0.051)
Actual contribution (%)	0.338*** (0.094)	0.337*** (0.095)	0.340*** (0.094)
“Two tables or less” <i>window-dressing dummy</i>	3.882** (1.700)	3.779** (1.768)	-
“Three tables or more” <i>window-dressing dummy</i>	-	-0.719 (2.346)	-1.618 (2.172)
Summation Skills	-0.070 (0.179)	-0.064 (0.167)	-0.042 (0.185)
Gender Dummy (1 if female)	-1.190 (2.139)	-1.244 (2.210)	-1.364 (2.128)
Observations (organizations)	n = 240 (16)	n = 240 (16)	n = 240 (16)
R ²	0.254	0.253	0.232

Notes: Estimation output using robust standard errors clustered at the organization level (in parentheses).

*** Significant at the 1 percent level.; ** Significant at the 5 percent level; * Significant at the 10 percent level.

In sum, the fact that *window-dressing* activities are less prevalent in Experiment 2 compared to Experiment 1 is likely to be due to the presence of the manager who can detect and then punish excessive manipulation attempts. It is interesting to note that monitoring, even in the absence of possible punishments associated to lower pay, might inflict moral costs on workers thus deterring them from engaging in manipulation attempts (see Gneezy et al., 2018; Abeler et al., 2018). Another reason why *window-dressing activities* may be limited can also be linked to the manager's ability to weaken incentives thus rendering performance manipulation ineffective. To test the weakening of incentives in the presence of *window-dressing* activities (Hypothesis 1B), we proceed, as for Experiment 1, by assessing the extent to which performance explains employees' pay in the two *discrete pay* treatments. We find that when *window-dressing* activities are available, a \$1 increase in one's performance increases one's pay by \$0.39 (when *window-dressing* consists of two tables or less) compared to \$0.58 in the absence of *window-dressing*. This difference is statistically significant (see Table B.7 in the Appendix B, p-value = 0.028). As is the case for Experiment 1, the relationship between performance and pay in the presence of *window-dressing* is significantly higher under *discrete pay* than under *equal pay* whether *window-dressing* is present or not (p-values < 0.001).¹² This means managers rely on employees' observed performance to determine their pay even when *window-dressing* pervasively distorts the informational content of the performance metric. Managers thus seem reluctant to rely exclusively on equal pay to deter performance manipulation. This could be because managers intrinsically value their decision rights over the pay of their subordinates (e.g., McClelland, 1975; Bartling et al., 2014) or because they believe employees value payment schemes that are at least partly based on individual performance metrics (e.g., Ogonnaya et al., 2017).

At the same time, the results in Table 5 show that managers discounted observed performance whenever employees exaggerated their performance by a large amount (see negative coefficient for the "Three tables or more" *window-dressing dummy*). Finally, managers were more likely to allocate organizational production equally across employees in the presence of *window-dressing* (22.5% of the

¹² These p-values correspond to the Wald tests that the coefficient associated to performance (see Table B.7) is equal to 0.25. Indeed, by construction, in *equal pay* treatments a \$1 increase in one's production will increase pay by \$0.25.

cases) than in its absence (18.7% of the cases) although this difference is not statistically significant (p-value = 0.355).¹³

Hypothesis 1C: Organizational production.—We find support for Hypothesis 1C by showing that, under *discrete pay*, the average production of the organization (per period) was on average 14.6% lower in the *window-dressing* treatment (\$16.25) compared to the *no window-dressing* treatment (\$18.62) (see Figure 3 below).

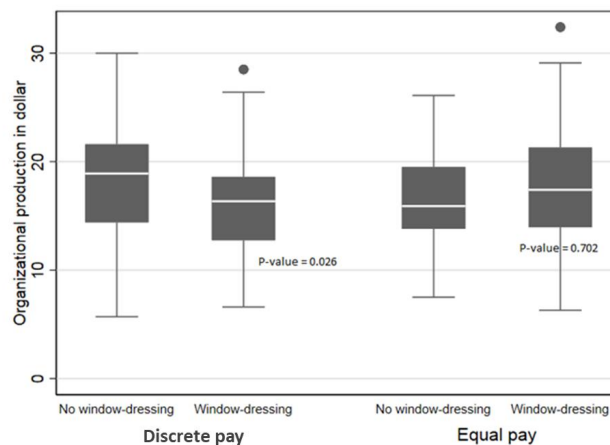


FIGURE 3. Box-plots for organizational production (per period, in cents) across treatments^a

Notes: ^a P-values are computed using OLS regressions with robust standard errors. The reported p-values correspond to the coefficient of the *window-dressing dummy* (which takes value one for treatments in which window-dressing is available and value zero otherwise). We use controls for gender and summation skills as controls. These variables are defined as organizational levels averages.

As for Experiment 1 and in line with Hypothesis 1C, we find no significant differences between *equal pay* treatments in terms of organizational production. In line with Experiment 1, we also find that organizational production was 14.0% greater under *discrete pay* than under *equal pay* in the absence of *window-dressing* activities (p-value = 0.076). By contrast, organizational production was 8.6% lower under *discrete pay* than under *equal pay* in the presence of *window-dressing* (although the difference is not statistically significant, p-value = 0.222).

Our findings thus support the recommendation derived from the influence costs literature to give away managerial discretion by setting an equal pay to all workers.

¹³ This is the p-value associated to the coefficient for the dummy of the *discrete pay* treatment with *window-dressing* in a probit panel regression with random effects and robust standard errors in which the dependent variable takes value 1 if a manager allocates organizational production equally across workers in a given period and value zero otherwise.

6. Conclusion

We provided the first causal test of the negative effect of performance manipulation activities in an organizational context. We showed that weakening incentives and limiting managerial discretion may be optimal to deter performance manipulation. In particular, *equal pay* seemed to substantially reduce *window-dressing* activities thus leading to higher organizational production than pay for performance. This was the case because the positive effect of *equal pay* on deterring *window-dressing* activities more than offset its negative incentive effect. This echoes the theoretical literature on tournaments showing that sabotage can be alleviated by reducing the gap between winning and losing prizes thus reducing the strength of incentives (see e.g., Lazear, 1989). This prediction has been supported in a series of laboratory experiments (Harbring and Irlenbush, 2004; 2005; 2011; Vandergrift and Yavas, 2010). The experimental literature in tournaments also suggests that giving workers the possibility to retaliate when observing that others engage in *window-dressing* activities might deter such manipulation attempts (Harbring et al., 2007; Vandergrift and Yavas, 2010).

In addition, our second experiment showed that performance manipulation could be reduced by putting a manager in charge of both supervising and compensating workers. This was the case because, thanks to their supervision technology, managers seemed to be able to detect and thus penalize excessive performance manipulation. Furthermore, managers were able, thanks to their discretion over workers' pay, to lessen the link between performance and pay thus limiting workers' incentives to engage in performance manipulation.

Even when managers were endowed with a supervision technology and full discretion over workers' pay, organizational production was still negatively affected by the availability of *window-dressing* activities. This was the case because curbing *window-dressing* activities led managers to weaken incentives.

Managers may, however, not be willing to rely on equal pay to deter performance manipulation. This would be the case if managers intrinsically valued their decision rights (e.g., McClelland, 1975; Bartling et al., 2014). This implies that, in the absence of inflexible bureaucratic rules, managers may voluntarily leave the door open to manipulative attempts. In that case, it might be helpful to physically separate employees and managers to limit influence activities. This separation might motivate the existence of rigid hierarchical structures (see Lazear, 1989). This suggests future avenues of research assessing the causal impact of multi-layer hierarchies and different structures of communication on the extent of *window-dressing* activities within firms.

An appealing alternative strategy to eradicate performance manipulation could be based on recruiting workers who are less inclined to engage in influence activities. We show that such workers cannot be identified using standard recruiting techniques based on personality traits and fluid intelligence. Instead, we identify performance manipulators as being spiteful and reflective.

Because current recruiting methods emphasize fluid intelligence and because cognitive reflection correlates significantly and positively with fluid intelligence, firms may be inadvertently recruiting potential manipulators in large numbers. This selection effect suggests *window-dressing* activities are likely to be pervasive in actual firms. Thus, extending our analysis to profiling performance manipulators within actual firms would be a crucial step in the research agenda.

Appendix A: Performance manipulation model

Our model can be represented as a game with a pure strategy space represented as follows, $S_i := \{e_i, d_i\}$ where $e_i \in \{0,1\}$, $d_i \in \{0,1\}$ and $e_i + d_i \leq 1$.

In Table A.1, we derive the optimal effort and *window-dressing* decisions for $\rho = 1$ in terms of workers' cost levels across social motives and payment schemes. We consider efficient equilibria in which all workers exert effort.

TABLE A.1—EFFICIENT EQUILIBRIUM CONDITIONS ACROSS WORKERS' SOCIAL MOTIVES WHEN $\rho = 1$

Payment scheme	Equal pay	Observed-performance pay
<i>Social motive</i>		
Selfish	$e_i^* = 1 \Leftrightarrow c \leq \frac{y}{n}$	$e_i^* = 1 \Leftrightarrow c \leq y - \max\{\frac{(n-1)\rho_a y}{n} - \chi; 0\}$
Altruistic / Spiteful	$e_i^* = 1 \Leftrightarrow c \leq \frac{y}{n}$	$e_i^* = 1 \Leftrightarrow c \leq y - \max\{\frac{(n-1)\rho_a y}{n} - \chi - (1 - \rho_a)\gamma_i^\alpha y; 0\}$

We now consider *window-dressing* equilibria. In our setting, these equilibria can only occur when some workers exert effort whereas others do not. For this to happen, we consider that the population of workers is divided into two types of workers. Low-ability workers are such that: $\rho := \rho_L < \rho_a$ and high-ability workers are such that: $\rho := \rho_H = 1$ and $n_H < n$ denotes the number of high-ability workers in the population. We derive the *window-dressing* equilibrium conditions below.

TABLE A.2—WINDOW-DRESSING EQUILIBRIA CONDITIONS ACROSS WORKERS' ABILITY TYPES

Payment scheme	Equal pay	Observed-performance pay
<i>Social motive</i>		
Selfish	$d_i^* = 0$ always	$d_i^* = 1 \Leftrightarrow$ $c_L \geq \chi + \max\{\rho_L y \frac{n_H + 1}{n}; c_L\} - \rho_d y \frac{n_H}{n}$ $c_H \leq \rho_d y \frac{n_H}{n} + (1 - \rho_d)y - \max\{\rho_d y \frac{n_H - 1}{n} - \chi; 0\}$
Altruistic / Spiteful	$d_i^* = 0$ always	$d_i^* = 1 \Leftrightarrow$ $c_L \geq \chi + \max\left\{\rho_L y \frac{n_H + 1}{n} - (1 - \rho_L)\gamma_i^\alpha y n_H - \rho_L(1 - \rho_d)\gamma_i^\beta y(n_L - 1); c_L\right\}$ $- \rho_d y \frac{n_H}{n} + (1 - \rho_d)\gamma_i^\alpha y n_H$ $c_H \leq \rho_d y \frac{n_H}{n} + (1 - \rho_d)y - (1 - \rho_d)\gamma_i^\beta y n_L$ $- \max\{\rho_d y \frac{n_H - 1}{n} - \chi - (1 - \rho_d)\gamma_i^\alpha y(n_H - 1); 0\}$

To derive Hypothesis 1A, we note that *window-dressing* equilibria may arise under *observed-performance pay* whereas they do not under *equal-pay* (see Table A.2). Under *observed-performance pay*, it follows that in the presence of *window-dressing* some workers (see low-ability workers in Table A.2) could be rewarded although they have not produced output ($y_i = 0$). This does not occur in the absence of *window-dressing* activities in which case workers are always paid according to their actual contribution to the work task. Hypothesis 1B thus follows.

Hypothesis 1C follows directly from the fact that ensuring workers' effort in equilibrium under *equal pay* is unaffected by the presence of *window-dressing* activities ($c \leq \frac{y}{n}$) whereas under *observed-performance pay* the range of cost of effort values that implements the efficient equilibrium is smaller in the absence of *window-dressing* activities ($c \leq y$) than in its presence. Indeed, in the presence of *window-dressing* activities the upper bound on the worker's cost of effort ensuring the existence of an efficient equilibrium is less than y (see right column in Table A.2).

Hypothesis 2A follows from the fact that under *observed-performance pay* the conditions for low-ability workers to engage in *window-dressing* in equilibrium are less likely for large values of γ_i^α (see Table A.2). This is the case because an increase in γ_i^α increases (decreases) the upper (lower) bound for

the cost of effort of high-ability (low-ability) workers that ensures a *window-dressing* equilibrium exists. The intuition is that envious [altruistic] workers ($\gamma_i^\alpha > 0$) [$\gamma_i^\alpha < 0$] are less [more] likely to deviate from the *window-dressing* equilibrium because it would most likely lead them to earn lower payoff than their counterparts.

Note that for $\rho_d \neq 1$, shame (γ_i^β) also affects the conditions for the existence of a *window-dressing* equilibrium. We do not derive any conjecture on shame because it is found to matter in only specific circumstances. In particular, a shameful low-ability worker might be less likely to deviate from the *window-dressing* equilibrium by exerting effort in the case in which this effort generates output (with probability ρ_L) whereas other low-ability workers' *window-dressing* attempts fail (with probability $1 - \rho_d$).

In line with Hypothesis 2B, it is also the case that *window-dressing* activities can only arise in equilibrium when low-ability workers engage in them while high-ability workers decide to work on the task. By contrast, it cannot be the case in equilibrium that high-ability workers engage in *window-dressing* while low-ability workers do not. The intuition is that deviating from the *window-dressing* equilibrium is less costly for high-ability workers than for low-ability workers because they are more likely to secure a payment when exerting effort in the absence of *window-dressing*.

Appendix B: Additional tables and analyses

EXPERIMENT 1

TABLE B.1—LINEAR PANEL REGRESSION WITH RANDOM EFFECTS FOR WORKERS' PAY (IN CENTS)

Dependent variable	Workers' pay (per period) in cents
Treatment	Observed-performance pay with window-dressing
Intercept	79.14*** (19.76)
Period Trend	8.05 (6.09)
Performance	0.58*** (0.05)
Gender Dummy (1 if female)	8.46 (12.36)
Observations (organizations)	n = 112 (28)
R ²	0.689

Notes: Estimation output using robust standard errors clustered at the organization level (in parentheses).

*** Significant at the 1 percent level.; ** Significant at the 5 percent level; * Significant at the 10 percent level.

TABLE B.2—PANEL REGRESSION WITH RANDOM EFFECTS FOR WINDOW-DRESSING (IN SECONDS) AND WORKERS' PRODUCTION (IN CENTS) AS A FUNCTION OF A LARGE SET OF INDIVIDUAL CONTROLS

Dependent variable	Window-dressing	Production
Intercept	393.84 (407.74)	-315.02 (428.187)
Period	50.71*** (19.44)	19.29 (15.83)
Individual characteristics		
Demographics		
Gender Dummy (1 if female)	-46.95 (48.29)	1.64 (41.23)
Age (in years)	16.01* (8.82)	-6.43 (9.42)
Work experience	5.41 (17.33)	-7.62 (14.39)
Volunteering experience	15.57 (13.11)	-17.02 (12.01)
Cognitive skills		
Summation Skills	-8.66*** (2.01)	19.87*** (2.70)
Raven test score	-1.40 (8.18)	1.27 (7.93)
Cognitive reflection test score	25.44*** (7.08)	-15.48*** (5.95)
Personality & social motives		
Big five personality traits	ns	a
Intrinsic motivation score	-42.27 (37.36)	94.27 (58.84)
Prosocial index	-15.71** (6.19)	13.65* (8.17)
Risk attitudes		
Risk aversion ^b	-0.88 (7.93)	-2.92 (8.17)
Observations (organizations)	n = 112 (28)	n = 112 (28)
R ²	0.317	0.546

Notes: Estimation output using robust standard errors clustered at the organization level (in parentheses).

*** Significant at the 1 percent level; ** Significant at the 5 percent level; * Significant at the 10 percent level.

^a Conscientiousness explains production significantly (p-value = 0.021) whereas the other four traits are not related to production.

^b Results are not affected by restricting our analysis to only the 46 (out of 56) participants who did not switch back and forth between the safer and the riskier option in the Holt and Laury (2002) test.

TABLE B.3—LINEAR PANEL REGRESSION WITH RANDOM EFFECTS FOR WORKERS' PRODUCTION (IN CENTS) AND WINDOW-DRESSING (IN SECONDS) AND A LARGE SET OF INDIVIDUAL CONTROLS

Dependent variable	Window-dressing		Production	
Intercept	258.75 (402.55)	249.88 (393.44)	-316.17 (411.23)	-323.17 (421.87)
Period Trend	50.71*** (19.44)	50.71*** (19.44)	19.29 (15.83)	19.29 (15.83)
Individual characteristics				
Demographics				
Gender Dummy (1 if female)	-49.80 (49.65)	-40.96 (47.39)	8.22 (40.11)	3.03 (42.41)
Age (in years)	15.21 (9.48)	16.39 (8.84)	-4.69 (9.85)	-7.13 (9.81)
Work experience	6.04 (17.81)	5.74 (17.63)	-8.89 (15.11)	-7.42 (14.07)
Volunteering experience	14.37 (13.78)	17.21 (12.88)	-13.90 (12.91)	-18.93 (11.700)
Cognitive skills				
Summation Skills	-8.46*** (2.04)	-8.80*** (1.88)	19.95*** (2.63)	19.6*** (2.67)
Raven test score	-2.94 (7.95)	-2.94 (7.95)	3.21 (7.52)	1.19 (7.88)
Cognitive reflection test score	24.58*** (7.18)	24.08*** (7.55)	-17.09*** (5.65)	-12.49** (5.93)
Personality & social motives				
Big five personality traits				
Intrinsic motivation score	ns -41.28 (38.61)	ns -46.35 (38.91)	a 88.86 (55.99)	b 99.38 (61.09)
Spiteful dummy	45.94* (29.09)	-	-64.44** (30.26)	-
Altruism dummy	-	-43.72** (17.88)	-	17.90 (36.55)
Risk attitudes				
Risk aversion ^c	-1.77 (7.48)	-1.62 (7.82)	-1.63 (7.58)	-2.65 (7.61)
Observations (organizations)	n = 112 (28)	n = 112 (28)	n = 112 (28)	n = 112 (28)
R ²	0.312	0.312	0.560	0.535

Notes: Estimation output using robust standard errors clustered at the organization level (in parentheses).

*** Significant at the 1 percent level.; ** Significant at the 5 percent level; * Significant at the 10 percent level.

^a Conscientiousness explains production significantly (p-value = 0.051) whereas the other four traits are not related to production.

^b Conscientiousness explains production significantly (p-value = 0.031) whereas the other four traits are not related to production.

^c Results are not affected by restricting our analysis to only the 46 (out of 56) participants who did not switch back and forth between the safer and the riskier option in the Holt and Laury (2002) test.

TABLE B.4—LINEAR PANEL REGRESSION WITH RANDOM EFFECTS FOR WORKERS' PRODUCTION (IN CENTS) AND A LARGE SET OF INDIVIDUAL CONTROLS

Treatment	Observed-performance pay (no window-dressing)	Equal-pay (both treatments)
Intercept	-132.87 (308.19)	-28.27 (211.68)
Period Trend	46.98*** (14.83)	71.75*** (9.77)
Individual characteristics		
Demographics		
Gender Dummy (1 if female)	-8.18 (33.34)	-51.00 (20.20)
Age	5.18 (9.52)	-4.28 (4.61)
Work experience	6.72 (7.04)	3.60 (6.10)
Volunteering experience	5.56 (7.51)	-6.35 (5.91)
Cognitive skills		
Summation Skills	12.49*** (3.25)	14.88*** (2.07)
Raven test score	-2.57 (4.79)	2.64 (4.23)
Cognitive reflection test score	18.10*** (7.12)	7.01 (5.76)
Personality & social motives		
Big five personality traits	ns	ns
Intrinsic motivation score	92.63*** (35.15)	67.11** (30.84)
Prosocial index	-10.75 (7.53)	-8.22 (6.39)
Risk attitudes		
Risk aversion ^a	7.63 (5.54)	0.57 (4.53)
Observations (organizations)	n = 112 (28)	n = 112 (28)
R ²	0.425	0.505

Notes: Estimation output using robust standard errors clustered at the organization level (in parentheses).

*** Significant at the 1 percent level; ** Significant at the 5 percent level; * Significant at the 10 percent level.

^a Results are not affected by restricting our analysis to only the 46 (out of 56) participants who did not switch back and forth between the safer and the riskier option in the Holt and Laury (2002) test.

EXPERIMENT 2

TABLE B.5—FACTORIAL DESIGN

	Number of participants (organizations) [sessions]	Payment Scheme	
		Discrete Pay	Equal Pay
<i>Window-dressing</i>	Absent	60 (15) [5]	60 (15) [5]
Activities	Present	64 (16) [5]	64 (16) [5]

In Table B.6, we provide additional robustness checks regarding the positive effect of *detectable window-dressing* activities on agent pay. We define as *detectable window-dressing* activities those for which:

- [1] Agents boosted their production by an amount which was not a multiple of 30.
- [2] Agents boosted their production by an *excessive* amount.

In particular, we define as *excessive* those *window-dressing* activities that led agents' production (as *observed* by the principal) to be in the top 35% (25%) [20%] {10%} of agents' production levels in the *discrete pay* treatment without influence: 540¢ (600¢) [660¢] {720¢}. These thresholds imply that the respective proportions of *undetectable window-dressing* activities are 22.2%, 26.2%, 30.0% and 33.3%.

Our regression results (see Table B.6) are consistent with Table 5 showing that agents can increase their *discrete pay* by engaging in *undetectable window-dressing* activities.

TABLE B.6—LINEAR PANEL REGRESSIONS ASSESSING THE EFFECT OF UNDETECTABLE WINDOW-DRESSING ACTIVITIES ON DISCRETE PAY FOR THE TREATMENT WITH WINDOW-DRESSING ACTIVITIES AND DISCRETE PAY

Dependent variable	Workers' pay (per period) in cents			
	Undetectable window-dressing activities thresholds			
	Top 35% 540¢	Top 25% 600¢	Top 20% 660¢	Top 10% 720¢
Intercept	21.296*** (2.999)	21.225*** (3.015)	21.704*** (2.938)	21.738** (2.921)
Period Trend	-0.027 (0.049)	-0.009 (0.059)	-0.001 (0.029)	-0.013 (0.035)
Actual contribution (%)	0.351*** (0.094)	0.351*** (0.094)	0.344*** (0.094)	0.345*** (0.095)
Undetectable Window-Dressing Dummy	4.294** (2.195)	4.299** (2.118)	1.953 [◊] (1.370)	2.030 [◊] (1.345)
Summation Skills	-0.044 (0.185)	-0.049 (0.178)	-0.050 (0.185)	-0.054 (0.181)
Gender Dummy (1 if female)	-1.114 (2.039)	-1.119 (2.108)	-1.169 (2.176)	-1.167 (2.160)
Observations (organizations)	n = 240 (16)	n = 240 (16)	n = 240 (16)	n = 240 (16)
R ²	0.255	0.259	0.242	0.243

Notes: Estimation output using robust standard errors clustered at the organization level (in parentheses).

*** Significant at the 1 percent level.; ** Significant at the 5 percent level; * Significant at the 10 percent level; [◊] Significant at the 10 percent level (one-sided test).

TABLE B.7—LINEAR PANEL REGRESSION WITH RANDOM EFFECTS FOR WORKERS' PAY (IN CENTS)

Treatment	No window-dressing	Window-dressing	No window-dressing & Window-dressing
Intercept	104.00*** (33.88)	142.23*** (19.41)	94.44*** (33.19)
Period Trend	7.28*** (2.60)	13.44*** (3.15)	10.506*** (2.05)
Performance	0.58*** (0.07)	0.39*** (0.05)	0.57*** (0.06)
Window dressing dummy × Performance	-	-	-0.17** (0.07)
Window dressing dummy	-	-	57.86 (38.93)
Gender Dummy (1 if female)	-17.64 (19.62)	5.60 (23.72)	-5.14 (15.53)
Observations (organizations)	n = 225 (45)	n = 240 (48)	n = 465 (93)
R ²	0.668	0.376	0.549

Notes: Estimation output using robust standard errors clustered at the organization level (in parentheses).

*** Significant at the 1 percent level.; ** Significant at the 5 percent level; * Significant at the 10 percent level.

References

- Abeler, J., Nosenzo, D., Raymond, C., 2018. Preferences for truth-telling. *Econometrica* forthcoming. 2016-13.
- Alchian, A.A., Demsetz, H., 1972. Production, information costs, and economic organization. *The American Economic Review* 62, 777–795. <https://www.jstor.org/stable/1815199>.
- Andreoni, J., 1995. Cooperation in public goods experiments: Kindness or confusion? *The American Economic Review* 85, 891–904. <https://www.jstor.org/stable/2118238>.
- Atanasov, P., Dana, J.D., 2011. Leveling the playing field: Dishonesty in the face of threat. *Journal of Economic Psychology* 32, 809–817. <http://dx.doi.org/10.1016/j.joep.2011.07.006>.
- Baker, G.P., 1992. Incentive contracts and performance measurement. *Journal of Political Economy* 100, 598–614. <http://dx.doi.org/10.1086/261831>.
- Baker, G.P., Holmström, B., 1995. Internal labor markets: Too many theories, too few facts. *The American Economic Review Papers and Proceedings* 85, 255–259. <https://www.jstor.org/stable/2117928>.
- Baker, G.P., Jensen, M.C., Murphy, K.J., 1988. Compensation and incentives: Practice vs. theory. *Journal of Finance* 43, 593–616. <http://dx.doi.org/10.1111/j.1540-6261.1988.tb04593.x>.
- Balafoutas, L., Czermak, S., Eulerich, M., Fornwagner, H., 2017. Incentives for dishonesty: An experimental study with internal auditors. Faculty of Economics and Statistics, University of Innsbruck, Working paper 2017-06.
- Balafoutas, L., Kerschbamer, R., Sutter, M., 2012a. Distributional preferences and competitive behavior. *Journal of Economic Behavior & Organization* 83, 125–135. <http://dx.doi.org/10.1016/j.jebo.2011.06.018>.
- Balafoutas, L., Lindner, F., Sutter, M., 2012b. Sabotage in tournaments: Evidence from a natural experiment. *Kyklos* 65, 425–441. <http://dx.doi.org/10.1111/kykl.12000>.
- Bandiera, O., Barankay, I., Rasul, I., 2010. Social incentives in the workplace. *Review of Economic Studies* 77, 417–458. <http://dx.doi.org/10.1111/j.1467-937X.2009.00574.x>.
- Bartling, B., Fehr, E., Herz, H., 2014. The intrinsic value of decision rights. *Econometrica* 82, 2005–2039. <http://dx.doi.org/10.3982/ECTA11573>.
- Bartling, B., Fehr, E., Maréchal, M., Schunk, D., 2009. Egalitarianism and competitiveness. *The American Economic Review* 99, 93–98. <http://dx.doi.org/10.1257/aer.99.2.93>.

- Bartling, B., Von Siemens, F.A., 2010. The intensity of incentives in firms and markets: Moral hazard with envious agents. *Labour Economics* 17, 598–607. <http://dx.doi.org/10.1016/j.labeco.2009.10.002>.
- Becker, G. S., 1968. Crime and Punishment: An Economic Approach. *Journal of Political Economy* 76, 169–217. <http://doi.org/10.1086/259394>.
- Bénabou, R., Tirole, J., 2003. Intrinsic and extrinsic motivation. *Review of Economics Studies* 70, 489–520. <https://www.jstor.org/stable/3648598>.
- Besley, T., Ghatak, M., 2005. Competition and incentives with motivated agents. *The American Economic Review* 95, 616–636. <http://dx.doi.org/10.1257/0002828054201413>.
- Beyer, A., Guttman, I., Marinovic, I., 2014. Optimal contracts with performance manipulation. *Journal of Accounting Research* 52, 817–847. <http://dx.doi.org/10.1111/1475-679X.12058>.
- Bierbrauer, F., Ockenfels, A., Pollak, A., Rückert, D., 2017. Robust mechanism design and social preferences. *Journal of Public Economics* 149, 59–80. <http://dx.doi.org/10.1016/j.jpubeco.2017.03.003>.
- Bloom, N., Van Reenen, J., 2007. Measuring and explaining management practices across firms and countries. *The Quarterly Journal of Economics* 122, 1351–1408. <http://dx.doi.org/10.1162/qjec.2007.122.4.1351>.
- Burks, S.V., Connor, L., Kivi, P., Wiener, A., Anderson, J.E., Lorenz, G., DeYoung, C.G., Rustichini, A., 2015. Cognitive skills, personality, and economic preferences in collegiate success. *Journal of Economic Behavior & Organization* 115, 30–44. <https://doi.org/10.1016/j.jebo.2015.01.007>.
- Cabrales, A., Feri, F., Gottardi, P., Meléndez, M., 2016. Can there be a market for cheap-talk information? some experimental evidence. European University Institute, Working Paper EUI ECO 2016/07.
- Capraro, V., 2017. Who lies? a meta-analysis of the effect of sex, age, and education on honesty. Cornell University Library, Physics and Society Working Paper arXiv:1703.03739.
- Carpenter, J., Gong, E., 2016. Motivating agents: How much does the mission matter? *Journal of Labor Economics* 34, 211–236. <http://dx.doi.org/10.1086/682345>.
- Carpenter, J., Matthews, P., Schirm, J., 2010. Tournaments and office politics: Evidence from a real effort experiment. *The American Economic Review* 100, 504–517. <http://dx.doi.org/10.1257/aer.100.1.504>.
- Carrell, S.E., Malmstrom, F.V., West, J.E., 2008. Peer effects in academic cheating. *Journal of Human Resources* 43, 173–207. <http://dx.doi.org/10.3368/jhr.43.1.173>.
- Charness, G., Masclet, D., Villeval, M.-C., 2014. The dark side of competition for status. *Management Science* 60, 38–55. <http://dx.doi.org/10.1287/mnsc.2013.1747>.

- Chiappori, P.A., Salanié, B., 2000. Testing Contract Theory: A Survey of Some Recent Work. In: Dewatripont, M., Hansen, L.P. and Turnovski, S.J. (Eds.), *Advances in Economic Theory. Eighth World Congress of the Econometric Society*. Cambridge University Press, Cambridge.
- Chowdhury, S., Gürtler, O., 2015. Sabotage in contests: A survey. *Public Choice* 164, 135–155. <http://dx.doi.org/10.1007/s11127-015-0264-9>.
- Chowdhury, S., Sheremeta, R., 2011. A generalized Tullock contest. *Public Choice* 147, 413–420. <http://dx.doi.org/10.1007/s11127-010-9636-3>.
- Chowdhury, S., Sheremeta, R., Turocy, T., 2014. Overbidding and overspreading in rent-seeking experiments: Cost structure and prize allocation rules. *Games and Economic Behavior* 87, 224–238. <http://dx.doi.org/10.1016/j.geb.2014.05.004>.
- Chung, T.Y., 1996. Rent-seeking contest when the prize increases with aggregate efforts. *Public Choice* 87, 55–66. <http://dx.doi.org/10.1007/BF00151729>.
- Corgnet, B., DeSantis, M., Porter, D., 2018a. What makes a good trader? on the role of intuition and reflection on trader performance. *The Journal of Finance* 73 (3), 1113–1137. <http://dx.doi.org/10.1111/jofi.12619>.
- Corgnet, B., Espin, A.M., Hernan-Gonzalez, R., 2015a. The cognitive basis of social behavior: Cognitive reflection overrides antisocial but not always prosocial motives. *Frontiers in Behavioral Neuroscience* 9. <http://dx.doi.org/10.3389/fnbeh.2015.00287>.
- Corgnet, B., Espin, A.M., Hernan-Gonzalez, R., 2016. Creativity and cognitive skills among millennials: Thinking too much and creating too little. *Frontiers in Psychology* 7. <http://dx.doi.org/10.3389/fpsyg.2016.01626>.
- Corgnet, B., Gómez-Miñambres, J., Hernan-Gonzalez, R., 2015b. Goal setting and monetary incentives: When large stakes are not enough. *Management Science* 61, 2926–2944. <https://doi.org/10.1287/mnsc.2014.2068>.
- Corgnet, B., Gómez-Miñambres, J., Hernan-Gonzalez, R., 2018b. Goal setting in the principal-agent model: Weak incentives for strong performance. *Games and Economics Behavior* 109, 311–326.
- Corgnet, B., Hernan-Gonzalez, R., Mateo, R., 2015c. Cognitive reflection and the diligent worker: an experimental study of Millennials. *PloS one* 10, e0141243. <http://dx.doi.org/10.1371/journal.pone.0141243>.
- Corgnet, B., Hernan-Gonzalez, R., Schniter, E., 2015e. Why real leisure really matters: Incentive effects on real effort in the laboratory. *Experimental Economics* 18, 284–301. <http://dx.doi.org/10.1007/s10683-014-9401-4>.

- Corgnet, B., Rodriguez-Lara, I., 2013. Are you a good employee or simply a good guy? Influence costs and contract design. *Journal of Economic Behavior & Organization* 92, 259–272. <http://dx.doi.org/10.1016/j.jebo.2013.06.003>.
- Crawford, V., Sobel, J., 1982. Strategic information transmission. *Econometrica*, 1431–1451.
- Crocker, K.J., Gresik, T., 2011. Optimal compensation with earnings manipulation: Managerial ownership and retention. Smeal College of Business The Pennsylvania State University.
- Crocker, K.J., Morgan, J., 1998. Is honesty the best policy? Curtailing insurance fraud through optimal incentive contracts. *Journal of Political Economy* 106, 355–375. <http://dx.doi.org/10.1086/250012>.
- Crocker, K.J., Slemrod, J., 2007. The economics of earnings manipulation and managerial compensation. *RAND Journal of Economics* 38, 698–713. <http://dx.doi.org/10.1111/j.0741-6261.2007.00107.x>.
- Dewatripont, M., Jewitt, I., Tirole, J., 2000. Multitask agency problems: Focus and task clustering. *European Economic Review* 44, 869–877. [http://dx.doi.org/10.1016/S0014-2921\(00\)00059-3](http://dx.doi.org/10.1016/S0014-2921(00)00059-3).
- Dickinson, D., 1999. An experimental examination of labor supply and work intensities. *Journal of Labor Economics* 17, 638–670. <http://dx.doi.org/10.1086/209934>.
- Dohmen, T., Falk, A., 2011. Performance pay and multi-dimensional sorting: Productivity, preferences and gender. *The American Economic Review* 101, 556–590. <http://dx.doi.org/10.1257/aer.101.2.556>.
- Dur, R., Sol, J., 2010. Social interaction, co-worker altruism, and incentives. *Games and Economic Behavior* 69, 293–301. <http://dx.doi.org/10.1016/j.geb.2009.10.013>.
- Erat, S., Gneezy, U., 2012. White lies. *Management Science* 58, 723–733. <http://dx.doi.org/10.1287/mnsc.1110.1449>.
- Eriksson, T., Poulsen, A., Villeval, M.-C., 2009. Feedback and incentives: Experimental evidence. *Labour Economics* 16, 679–688. <http://dx.doi.org/10.1016/j.labeco.2009.08.006>.
- Fairburn, J.A., Malcomson, J.M., 2001. Performance, promotion, and the Peter principle. *The Review of Economic Studies* 68, 45–66. <http://dx.doi.org/10.1111/1467-937X.00159>.
- Fehr, E., Hart, O., Zehnder, C., 2011. Contracts as reference points-Experimental evidence. *The American Economic Review* 101, 493–525. <http://dx.doi.org/10.1257/aer.101.2.493>.
- Fehr, E., Schmidt, K.M., 1999. A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics* 114, 817–868. <https://doi.org/10.1162/003355399556151>. Fehr, E., Schmidt, K.M., 2006. The Economics of Fairness, Reciprocity and Altruism - Experimental Evidence and New Theories. In: Kolm, S.-C., Mercier Ythier, J. (Eds.), *Handbook of the Economics of Giving, Altruism and Reciprocity*, vol. 1. North-Holland, Amsterdam, pp. 615–691.

- Fischbacher, U., Föllmi-Heusi, F., 2013. Lies in disguise-An experimental study on cheating. *Journal of the European Economic Association* 11, 525–547. <http://dx.doi.org/10.1111/jeea.12014>.
- Frederick, S., 2005. Cognitive reflection and decision making. *Journal of Economic Perspectives* 19, 25–42. <http://dx.doi.org/10.1257/089533005775196732>.
- Gibbons, R., 2005. Four formal(izable) theories of the firm? *Journal of Economic Behavior & Organization* 58, 200–245. <https://doi.org/10.1016/j.jebo.2004.09.010>.
- Gibbons, R., Roberts, J., 2013. Economics Theories of Incentives in Organizations. In: Gibbons, R., Roberts, J. (Eds.), *Handbook of Organizational Economics*. Princeton University Press, Princeton, Oxford, pp. 56–99.
- Gino, F., 2015. Understanding ordinary unethical behavior: Why people who value morality act immorally. *Current Opinion in Behavioral Sciences* 3, 107–111. <http://dx.doi.org/10.1016/j.cobeha.2015.03.001>.
- Gino, F., Ariely, D., 2012. The dark side of creativity: Original thinkers can be more dishonest. *Journal of Personality and Social Psychology* 102, 445–459. <http://dx.doi.org/10.1037/a0026406>.
- Gino, F., Krupka, E.L., Weber, R.A., 2013. License to cheat: Voluntary regulation and ethical behavior. *Management Science* 59, 2187–2203. <https://doi.org/10.1287/mnsc.1120.1699>.
- Gino, F., Pierce, L., 2009. Dishonesty in the name of equity. *Psychological Science* 20, 1153–1160. <https://doi.org/10.1111/j.1467-9280.2009.02421.x>.
- Gino, F., Schweitzer, M.E., Mead, N.L., Ariely, D., 2011. Unable to resist temptation: How self-control depletion promotes unethical behavior. *Organizational Behavior and Human Decision Processes* 115, 191–203. <http://dx.doi.org/10.1016/j.obhdp.2011.03.001>.
- Gneezy, U., 2005. Deception: The role of consequences. *The American Economic Review* 95, 384–394. <http://dx.doi.org/10.1257/0002828053828662>.
- Gneezy, U., Saccardo, S., van Veldhuizen, R., 2018. Bribery: Behavioral drivers of distorted decisions. *Journal of the European Economic Association* forthcoming. Available at SSRN. <http://dx.doi.org/10.2139/ssrn.2803623>.
- Goerg, S.J., Kube, S., 2012. Goals (th)at work- goals, monetary incentives, and workers' performance. Max Planck Institute for Research on Collective Goods Bonn, Working Paper 2012/19.
- Grosse, S., Rockenbach, B., Putterman, L., 2011. Monitoring in teams: Using laboratory experiments to study a theory of the firm. *Journal of the European Economic Association* 9, 785–816. <http://dx.doi.org/10.1111/j.1542-4774.2011.01026.x>.

- Grund, C., Sliwka, D., 2007. Envy and compassion in tournaments. *Journal of Economics and Management Strategy* 14, 187–207. <http://dx.doi.org/10.1111/j.1430-9134.2005.00039.x>.
- Harbring, C., Irlenbusch, B., 2004. Sabotage and the endogenous design of tournaments. Discussion paper, University of Bonn.
- Harbring, C., Irlenbusch, B., 2005. Incentives in tournaments with endogenous prize selection. *Journal of Institutional and Theoretical Economics* 161, 636–663.
- Harbring, C., Irlenbusch, B., 2008. How many winners are good to have? On tournaments with sabotage. *Journal of Economic Behavior and Organization* 65, 682–702.
- Harbring, C., Irlenbusch, B., 2011. Sabotage in tournaments: Evidence from a laboratory experiment. *Management Science* 57, 611–627. <http://dx.doi.org/10.1287/mnsc.1100.1296>.
- Harbring, C., Irlenbusch, B., Kraekel, M., Selten, R., 2007. Sabotage in corporate contests - An experimental analysis. *International Journal of the Economics of Business* 14, 367–392. <http://dx.doi.org/10.1080/13571510701597445>.
- Hart, O., 1995. *Firms, Contracts, and Financial Structure*. Oxford: Oxford University Press.
- Hart, O., Moore, J., 2008. Contracts as reference points. *The Quarterly Journal of Economics* 123, 1–48. <http://dx.doi.org/10.1162/qjec.2008.123.1.1>.
- Holmström, B., 1979. Moral hazard and observability. *Bell Journal of Economics* 10, 74–91. <http://dx.doi.org/10.2307/3003320>.
- Holmström, B., 1982. Moral hazard in teams. *Bell Journal of Economics* 13, 324–340.
- Holmström, B., Milgrom, P., 1991. Multitask principal-agent analyses: Incentive contracts, asset ownership and job design. *Journal of Law, Economics and Organization* VII, 24–52. <https://www.jstor.org/stable/764957>.
- Holt, C.A., Laury, S.K., 2002. Risk aversion and incentive effects. *The American Economic Review* 92, 1644–1655. <http://dx.doi.org/10.1257/000282802762024700>.
- Jacobsen, C., Fosgaard, T.R., Pascual-Ezama, D., 2018. Why do we lie? a practical guide to the dishonesty literature. *Journal of Economic Surveys* 32, 357–387. <http://dx.doi.org/10.1111/joes.12204>.
- John, O.P., Donahue, E.M., Kentle, R.L., 1991. *The big five inventory-versions 4a and 54*. Berkeley: University of California, Institute of Personality and Social Research.
- John, O.P., Naumann, L.P., Soto, C.J., 2008. Paradigm shift to the integrative big five trait taxonomy: History, measurement, and conceptual issues, In: John, O.P., Robins, R.W. and Pervin, L.A. (Eds.), *Handbook of personality: Theory and research*, vol. 3. Guilford Press, New York, pp. 114–158.

- Kahneman, D., 2011. *Thinking, fast and slow*. Farrar, Straus and Giroux, Macmillan, New York.
- Kandel, E., Lazear, E., 1992. Peer pressure and partnerships. *Journal of Political Economy* 100, 801–817. <http://dx.doi.org/10.1086/261840>.
- Klein, B., Crawford, R.G., Alchian, A.A., 1978. Vertical integration, appropriable rents, and the competitive contracting process. *Journal of Law and Economics* 21, 297–326. <http://dx.doi.org/10.1086/466922>.
- Konrad, K., 2000. Sabotage in rent-seeking contests. *Journal of Law, Economics and Organization* 16, 155–165. <http://dx.doi.org/10.1093/jleo/16.1.155>.
- Kosfeld, M., Neckermann, S., 2011. Getting more work for nothing? Symbolic awards and worker performance. *American Economic Journal: Microeconomics* 3, 86–99. <http://dx.doi.org/10.1257/mic.3.3.86>.
- Kräkel, M., 2007. Doping and cheating in contest-like situations. *European Journal of Political Economy* 23, 988–1006. <http://dx.doi.org/10.1016/j.ejpoleco.2006.11.003>.
- Laffont, J.J., Martimort, D., 2002. *The Theory of Incentives: The Principal-Agent Model*. Princeton University Press, Princeton, Oxford.
- Lakonishok, J., Shleifer, A., Thaler, R., Vishny, R., 1991. Window dressing by pension fund managers. *The American Economic Review Papers and Proceedings* 81, 227–231.
- Lazear, E., 1989. Pay equality and industrial politics. *Journal of Political Economy* 97, 561–580.
- Levin, J., 2003. Relational incentive contracts. *The American Economic Review* 93, 835–857. <http://dx.doi.org/10.1257/000282803322157115>.
- Maggi, G., Rodríguez-Clare, A., 1995. Costly distortion of information in agency problems. *Rand Journal of Economics* 26, 675–689. <https://www.jstor.org/stable/2556012>.
- Malachowski, D., 2005. Wasted time at work costing companies billions. *San Francisco Chronicle* July 11.
- Manso, G., 2011. Motivating innovation. *Journal of Finance* 66, 1823–1869. <http://dx.doi.org/10.1111/j.1540-6261.2011.01688.x>.
- McClelland, D., 1975. *Power: The Inner Experience*. Halsted Press, Irvington, New York.
- Mead, N.L., Baumeister, R.F., Gino, F., Schweitzer, M.E., Ariely, D., 2009. Too tired to tell the truth: Self-control resource depletion and dishonesty. *Journal of Experimental Social Psychology* 45, 594–597. <http://dx.doi.org/10.1016/j.jesp.2009.02.004>.

- Milgrom, P., 1988. Employment contracts, influence activities and efficient organization design. *Journal of Political Economy* 96, 42–60. <http://dx.doi.org/10.1086/261523>.
- Milgrom, P., Roberts, J., 1988. An economic approach to influence activities in organization. *American Journal of Sociology* 94, 154–179. <http://dx.doi.org/10.1086/228945>.
- Milgrom, P., Roberts, J., 1992. *The Economics, Organization and Management*. Prentice Hall, Englewood Cliffs.
- Niederle, M., Vesterlund, L., 2007. Do women shy away from competition? Do men compete too much? *The Quarterly Journal of Economics* 3, 1067–1101. <http://dx.doi.org/10.1162/qjec.122.3.1067>.
- Ogbonnaya, C., Daniels, K., Nielsen, K., 2017. Does contingent pay encourage positive employee attitudes and intensify work? *Human Resource Management Journal* 27, 94–112. <http://dx.doi.org/10.1111/1748-8583.12130>.
- Ponti, G., Rodriguez-Lara, I., 2015. Social preferences and cognitive reflection: evidence from dictator game experiment. *Frontiers in Behavioral Neuroscience* 9. <http://dx.doi.org/10.3389/fnbeh.2015.00146>.
- Powell, M., 2015. An influence-cost model of organizational practices and firm boundaries. *Journal of Law, Economics, & Organization* 31, 104–142. <https://doi.org/10.1093/jleo/ewv005>.
- Prendergast, C., 2002. The tenuous trade-off between risk and incentives. *Journal of Political Economy* 110, 1071–1102. <http://dx.doi.org/10.1086/341874>.
- Prendergast, C., Topel, R., 1993. Discretion and bias in performance evaluation. *European Economic Review* 37, 355–365. [http://dx.doi.org/10.1016/0014-2921\(93\)90024-5](http://dx.doi.org/10.1016/0014-2921(93)90024-5).
- Preston, I., Szymanski, S., 2003. Cheating in contests. *Oxford Review of Economic Policy* 19, 612–624. <http://dx.doi.org/10.1093/oxrep/19.4.612>.
- Robin, S., Rusinowska, A., Villeval, M.-C., 2014. Ingratiation: Experimental evidence. *European Economic Review* 66, 16–38. <http://dx.doi.org/10.1016/j.euroecorev.2013.11.005>.
- Roger, G., 2014. Optimal contract under moral hazard with soft information. *American Economic Journal: Microeconomics* 5, 55–80. <http://dx.doi.org/10.1257/mic.5.4.55>.
- Rotemberg, J.J., 1994. Humans relations in the workplace. *Journal of Political Economy* 102, 684–717. <http://dx.doi.org/10.1086/261951>.
- Schmidt, F., 2009. Select on Intelligence. In: Locke, E.A. (Ed.), *Handbook of Principles of Organizational Behavior*, John Wiley and Sons, United Kingdom, pp. 3–18.
- Schwieren, C., Weichselbaumer, D., 2010. Does competition enhance performance or cheating? A laboratory experiment. *Journal of Economic Psychology* 31, 241–253. <http://dx.doi.org/10.1016/j.joep.2009.02.005>.

- Shalvi, S., Dana, J., Handgraaf, M.J., De Dreu, C.K., 2011a. Justified ethicality: Observing desired counterfactuals modifies ethical perceptions and behavior. *Organizational Behavior and Human Decision Processes* 115, 181–190. <https://doi.org/10.1016/j.obhdp.2011.02.001>.
- Shalvi, S., Eldar, O., Bereby-Meyer, Y., 2012. Honesty requires time (and lack of justifications). *Psychological Science* 23, 1264–1270. <http://dx.doi.org/10.1177/0956797612443835>.
- Shalvi, S., Handgraaf, M.J., Dreu, C.K.D., 2011b. Ethical manoeuvring: Why people avoid both major and minor lies. *British Journal of Management* 22, 16–27. <http://dx.doi.org/10.1111/j.1467-8551.2010.00709.x>.
- Shalvi, S., Leiser, D., 2013. Moral firmness. *Journal of Economic Behavior & Organization* 93, 400–407. <http://dx.doi.org/10.1016/j.jebo.2013.03.014>.
- Sheremeta, R.M., 2016. The pros and cons of workplace tournaments. *IZA World of Labor* 302, 1–10. <http://dx.doi.org/10.15185/izawol.302>.
- Sliwka, D., 2007. Trust as a signal of a social norm and the hidden costs of incentive schemes. *The American Economic Review* 97, 999–1012. <http://dx.doi.org/10.1257/aer.97.3.999>.
- Sloof, R., Oosterbeek, H., Sonnemans, J., 2007. Does making specific investments unobservable boost investment incentives? *Journal of Economics & Management Strategy* 16, 911–942. <http://dx.doi.org/10.1111/j.1530-9134.2007.00162.x>.
- Sloof, R., Sonnemans, J., Oosterbeek, H., 2004. Specific investments, holdup, and the outside option principle. *European Economic Review* 48, 1399–1410. [http://dx.doi.org/10.1016/S0014-2921\(03\)00017-5](http://dx.doi.org/10.1016/S0014-2921(03)00017-5).
- Sonnemans, J., Oosterbeek, H., Sloof, R., 2001. On the relation between asset ownership and specific investments. *Economic Journal* 111, 791–820. <http://dx.doi.org/10.1111/1468-0297.00660>.
- Toplack, M.E., West, R.F., Stanovich, K.E., 2011. The cognitive reflection test as a predictor of performance on heuristics and biases tasks. *Memory & Cognition* 39, 1275–1289. <http://dx.doi.org/10.3758/s13421-011-0104-1>.
- Toplack, M.E., West, R.F., Stanovich, K.E., 2014. Assessing miserly information processing: An expansion of the cognitive reflection test. *Thinking & Reasoning* 20, 147–168. <http://dx.doi.org/10.1080/13546783.2013.844729>.
- Ugrin, J.C., Pearson, J.M., 2013. The effects of sanctions and stigmas on cyberloafing. *Computers in Human Behavior* 29, 812–820. <http://dx.doi.org/10.1016/j.chb.2012.11.005>.
- Utikal, V., Fischbacher, U., 2013. Disadvantageous lies in individual decisions. *Journal of Economic Behavior & Organization* 85, 108–111. <http://dx.doi.org/10.1016/j.jebo.2012.11.011>.

- Vandegrift, D., Yavas, A., 2010. An experimental test of sabotage in tournaments. *Journal of Institutional and Theoretical Economics* 166, 259–285.
- Van den Steen, E., 2010. Interpersonal authority in a theory of the firm. *The American Economic Review* 100, 466–490. <http://dx.doi.org/10.1257/aer.100.1.466>.
- Van Dijk, F., Sonnemans, J., van Winden, F., 2001. Incentive systems in a real effort experiment. *European Economic Review* 45, 187–214. [http://dx.doi.org/10.1016/S0014-2921\(00\)00056-8](http://dx.doi.org/10.1016/S0014-2921(00)00056-8).
- Verschuere, B., Shalvi, S., 2013. The truth comes naturally! Does it? *Journal of Language and Social Psychology* 33, 417–423. <http://dx.doi.org/10.1177/0261927X14535394>.
- Wu, G., Heath, C., Larrick, R., 2008. A prospect theory model of goal behavior. University of Chicago.