

DO LEADERS AFFECT ETHICAL CONDUCT?

Giovanna d'Adda
Politecnico di Milano

Nicola Pavanini
University of Zurich

Donja Darai
University of Zurich

Roberto A. Weber
University of Zurich

Abstract

We study whether leaders influence the unethical conduct of followers. To avoid selection issues present in natural environments, we use an experiment in which we create simple laboratory firms and assign leadership roles at random. In our first experiment, firms engage in competition and unethical behavior enhances firm earnings but produces a negative externality for all firms. We vary, by treatment, two instruments through which leaders can influence follower conduct—prominent statements to the group and the allocation of monetary incentives. We find that leaders influence the ethical conduct of followers both through their statements and through the use of incentives. Moreover, leaders who are likely to have acted dishonestly in a preliminary stage of the experiment are more likely to employ mechanisms to encourage dishonesty among followers. As a result, firms randomly assigned one of these unethical leaders are more likely to engage in misreporting. A second experiment finds that the above relationships are present, though weaker, when firms do not engage in direct competition. (JEL: C92, D73, M14).

1. Introduction

Responses to unethical conduct in organizational, political, or social settings often focus on the role of leaders in facilitating or encouraging such behavior. For example, analysis of WorldCom's dramatic collapse in 2002 often emphasizes the role played by its founder, chairman, and CEO, Bernard Ebbers. Indeed, the SEC committee investigating WorldCom's collapse prominently noted that Ebbers "was the source of

The editor in charge of this paper was Nicola Gennaioli.

Acknowledgments: The authors gratefully acknowledge financial support from the Swiss National Science Foundation (Project 100018-140571). We are also grateful for comments and suggestions from participants at several conferences and seminars. We are also very thankful to Nicola Gennaioli, an associate editor and four anonymous referees for helpful comments and suggestions.

E-mail: giovanna.dadda@polimi.it (d'Adda); donja.darai@gmail.com (Darai); nicola.pavanini@econ.uzh.ch (Pavanini); roberto.weber@econ.uzh.ch (Weber)

the culture, as well as much of the pressure, that gave birth to this fraud” (Beresford et al. 2003, p. 1). More recently, former Volkswagen CEO Martin Winterkorn was accused of playing a role in the 2015 “Dieselgate” emissions manipulation scandal through a “management style . . . that fostered a climate of fear” (Cremer and Bergin 2015). Conversely, a great deal of attention is also dedicated to the idea that a leader who prioritizes ethics can create more ethical conduct within an organization (Brown and Treviño 2006).

However, despite the widespread belief that leaders play a critical role in producing unethical conduct in groups or firms that they lead, there is little direct evidence in economic or organizational research of such a relationship. This is not surprising, since there are serious challenges to cleanly identifying the influence of leaders in fomenting unethical conduct. For starters, unethical acts in the field, either by leaders or by those who follow them, are often hidden from view, which makes the study of this relationship difficult. Moreover, even if a researcher is able to document a relationship between leaders’ actions and followers’ unethical conduct, establishing causality is often impossible due to nonrandom selection of leaders: when a corrupt firm has an unethical CEO, is the leader the source of the culture, or is it the culture that led to the appointment of an unethical leader?¹

Recognizing these identification problems, we employ a novel approach to study the relationship between leadership and unethical conduct. In particular, we use laboratory experiments, which allow us to exploit the high degree of control afforded by such environments to avoid many of the problems present in more natural settings. In our experiments, we employ a task that captures key features of much real-world unethical conduct. In this task, individual “workers” in simple laboratory “firms” decide whether to lie and hurt others outside the firm for personal material gain, with the knowledge that individual lies are unlikely to be detected. We study the impact of leadership by giving some of these laboratory firms “leaders” endowed with some of the influence channels typical of organizational leaders—the ability to make public statements and control over the distribution of financial rewards. An important feature of our design is that it randomly assigns leaders to firms so that some firms have leaders more prone to personally engage in unethical conduct. We then observe how workers in these firms behave, both in comparison with firms randomly assigned more ethical leaders and with firms free of any influence by leaders. In this manner, we are able to determine whether the type of leader—and the things that leader does—affect the ethical conduct of the firms they lead.

Our study represents a novel contribution to the literature on leadership and the ethical conduct of groups. Milgram (1963) provided shocking evidence that people’s willingness to harm others can be influenced by statements from an authority figure,

1. As an example, many people attribute a large part of the corruption in FIFA, football’s international governing body, to the actions of its former leader, Sepp Blatter, who, for example, allegedly directly oversaw the payment of large bribes. However, others argue that Sepp Blatter’s corrupt behavior is a symptom, rather than a cause, of FIFA’s corruption and that the ultimate cause lies with the organization’s structure (Bialik 2015).

supporting the idea that unethical leaders can facilitate unethical conduct by followers. However, this precise relationship is often supported in the organizational literature by evidence that is not entirely compelling, due to the identification problems outlined above. For example, some of the strongest evidence comes from studies that use survey-based instruments to measure followers' perceptions of the degree to which a leader possesses "ethical" characteristics (e.g., the Ethical Leadership scale; Brown et al. 2005), and then correlate this measure with other subjective measures of whether individuals within an organization act ethically (Mayer et al. 2012).² Although these studies often find positive correlations between the two kinds of perceptions, the results must be interpreted cautiously due to many possible interpretations, such as correlated bias across subjective measures.

Given the hidden nature of unethical conduct, other studies attempt to identify a relationship with leadership by measuring observable behaviors that are potentially correlated with unethical follower conduct, such as employee exit or excess costs (Detert et al. 2007; Burks and Krupka 2012). Beekman et al. (2013) use a proxy for corrupt leader behavior—the misplacement of community resources under the leader's control—and show that this correlates with lower cooperation among community members, though not necessarily with more unethical conduct. As with the studies discussed above, these results are more suggestive of a relationship, since they ultimately only observe behavior that is imperfectly correlated with unethical conduct. Moreover, even if such a relationship exists, it is difficult to identify whether leaders are causal.³

To address concerns about causal identification, we follow other recent research that relies on laboratory experiments to study leadership (Brandts and Cooper 2007; Güth et al. 2007; Potters et al. 2007; Hamman et al. 2011; Brandts, Cooper and Weber 2015). Like us, these studies exogenously vary the abilities and characteristics of leaders to cleanly identify the effects leaders have on a particular behavior of interest.

To address the concern that unethical conduct is hard to observe, we also follow other experimental work that studies unethical conduct (Gneezy 2005; Mazar et al. 2008; Gino et al. 2009; Conrads et al. 2013, 2014; Fischbacher and Föllmi-Heusi 2013; Gibson et al. 2013). Specifically, we use a behavioral task in which workers have the opportunity to act unethically—that is, to tell a lie for profit in a manner that cannot be

2. In addition to two such correlational studies, Mayer et al. (2013) also conduct a laboratory experiment in which subjects are placed in fictitious groups and randomly exposed to experimenter-generated messages from fictitious leaders and co-workers. The experiment measures unethical conduct by whether workers report, in an exit questionnaire, that the co-workers cheated. The use of deception, the absence of real leaders and an ambiguous measure of unethical conduct make the causal connection between leadership and behavior unclear.

3. Other studies, not directly related to ethical conduct, show a relationship between leaders' behaviors and the prevalence of cooperative behavior within a group (Jack and Recalde 2015; Kosfeld and Rustagi 2015). There are rare cases in which random variation in leadership exists in the field. These studies provide valuable insights into the influence and importance of leaders for economic outcomes—e.g., leaders' gender and public good provision (Chattopadhyay and Duflo 2004) and national leaders and economic growth (Jones and Olken 2005).

detected—but where our ability to create numerous identical replications of the same situation makes inference regarding the presence of unethical conduct possible. In the task, workers privately roll a die, which provides their actual performance in a period, but have the opportunity to misreport the actual outcome, with no possibility of lies being discovered. The extent to which subjects inflate their reported die roll is the key indicator of unethical conduct in our study.

In our first experiment, firms compete with each other and the harm created by misreporting is a negative externality for all firms. When a worker acts dishonestly, she makes her firm better off but harms all other firms to a greater extent. Hence, aggregate profits are highest, in expectation, when everyone in the industry acts honestly, but any individual firm's profits always increase with misreporting. This captures many situations in which unethical conduct provides a firm an advantage over competitors, but where widespread unethical conduct is damaging to the entire industry. For example, unethical lending practices by an individual financial institution may allow it to generate more loans and short-term profits, but may yield greater long-term regulatory burdens and higher risk of financial instability. Moreover, domains involving high levels of competition, in which a win for one party produces losses for another, are where one often encounters instances of unethical conduct.⁴

At the level of a worker, this design also incorporates many key features of real-world unethical conduct. Inflating the reported die roll is personally rewarding, but wrong according to most normative moral principles—it violates deontological rules (e.g., Kant 1785) against dishonesty and, by producing more social harm than welfare, it also violates utilitarian ethics (Mill 1863).⁵ Moreover, as is often the case outside the laboratory, unethical conduct, at the individual level, is hidden from view, meaning there is no probability of sanctioning or detection of individual unethical conduct. Hence, “good” behavior is likely the result of ethical considerations, rather than fear of detection or punishment (Nagin et al. 2002).

The main part of our experimental design consists of two stages. In the first, participants perform the die-roll task once individually, with the possibility to benefit personally from misreporting, which also harms others in a subjects' experimental session. This provides a (noisy) measure of each individual's tendency to misreport. In the second stage, which is our primary focus, participants are randomly matched into four-person firms, and one person within each firm is randomly selected to be the leader, whereas the remaining three subjects are workers. The second stage of

4. Examples include competition over market share (Volkswagen's 2015 emissions scandal), in financial markets (Toshiba's 2008-2015 earnings manipulation) and in sports (organized doping in cycling and athletics).

5. In contrast with other possible designs—as when a lie benefits the decision maker without affecting other subjects (Fischbacher and Föllmi-Heusi 2013; Gibson et al. 2013)—self-interested dishonesty in our first experiment clearly produces negative net social impact. We chose this design specifically to make dishonesty inconsistent with utilitarian ethics. This design creates the possibility that highly utilitarian subjects may lie by reporting lower die rolls—though this only benefits others in the case where there is widespread dishonesty. Although we cannot rule out this possibility, a priori, the data suggest that it is not a common motive.

our experiment introduces, and varies by treatment, the principal influence channels ascribed to leaders in the leadership literature.⁶

The first influence channel available to leaders is the ability to make public statements to the workers in their firm. Organizational leadership is often associated with someone who, from a position of prominence, can articulate a broad direction for the firm and motivate employees in pursuit of that end (Kotter 1990). In his seminal book on corporate executives, Barnard (1938) notes that one of their fundamental functions “is to formulate and define the purposes, objectives, ends, of the organization” (p. 231). Modern studies of and training in corporate leadership similarly note the importance of a leader’s ability to motivate and convince followers to pursue a particular direction, often through prominent and visible speeches (Antonakis et al. 2012; Lazear 2012). We implement a simple form of this leadership function in our experiment, by providing leaders with a platform to send statements to workers in their firm between periods of the task.

Second, corporate leaders often possess discretion over financial incentives. A central theme in organizational economics is the use of monetary incentives as a means for motivating employees to act in a manner desired by their employers. Such incentives take many forms, from fixed wages, to precisely defined explicit performance contracts, to subjectively determined performance bonuses at the discretion of supervisors (Baker et al. 1994; Prendergast 1999). In principle, a leader could use any form of variable pay to incentivize ethical or unethical behavior. In our experiments, we provide leaders with the ability to distribute part of employees’ compensation in the form of a discretionary bonus.

To see how a leader might affect the behavior of followers, consider a worker with the following expected utility function, facing a choice of how much to misreport an outcome:

$$E[u] = E[\pi(r, L^I)] - c(r, L^S). \quad (1)$$

Let $r = 0$ correspond to honesty, while $r > 0$ captures the degree to which the worker engages in favorable misrepresentation of a performance measure, such as sales or company earnings. The first term on the right-hand side represents the expected monetary payoff to a worker from her action, whereas the second represents an internal psychological cost, such as guilt, from the chosen action. If $E[\pi(r, L)]$ and

6. Central to organizational leadership research is a distinction between “transactional” and “transformational” leaders (Burns 1978), with the former effecting change through sanctions and rewards and the latter doing so through persuasion and influencing follower preferences (Bass 1990). Of course, individuals who are not formal leaders—such as influential peers—might also exert a form of informal leadership influence if they make statements that are attended to by co-workers. Such influence is consistent with leadership in our experiment. Moreover, leaders also do many other things than what we allow them to do. For example, they set examples for followers through their actions (Hermalin 1998) and select or exclude organizational members (Güth et al. 2007). For simplicity, we focus on two specific leadership functions. The omission of other channels, which our experimental design can be naturally extended to include, likely underestimates the degree to which leaders influence followers in more natural settings.

$c(r, L^S)$ are both increasing in the degree of misreporting, r , then a worker is financially better off by acting unethically, but also experiences greater psychological disutility.

The terms, L^I and L^S , represent, respectively, the incentive and statement channels through which a leader can influence workers. As an example, take the case of auditors at Arthur Andersen, who were alleged to have frequently facilitated accounting violations at audited firms, including Enron and WorldCom. Corporate leaders incentivized auditors to keep clients satisfied (L^I), as a means for generating revenue from selling nonauditing services to the same clients (Toffler and Reingold 2004). Hence, auditors could personally benefit from performing lax audits. At the same time, leaders' statements justifying more lax auditing of large clients (L^S)—for instance, the head of US operations urged auditors to “empathize” with clients (Brown and Dugan 2002)—potentially led auditors to focus on the perspective of the audited firm, thereby lowering the psychological costs of overlooking auditing irregularities. Our experiments represent a very simplified form of these kinds of channels through which leaders may influence the conduct of followers.

Previewing the results of our main experiment, we find that leaders influence the ethical conduct of followers, both through their statements and through their use of incentives. Workers report higher numbers when their leader makes statements encouraging dishonest reporting and when the leader's past behavior indicates that workers who report high numbers will be rewarded. We identify these effects both from reduced-form estimates of how leaders' actions influence followers' behaviors and from structural estimates of a simple utility framework, based on the one above, in which workers' utilities depend on monetary payoffs and on a psychological cost to acting unethically, both of which can be influenced by the leader through different channels.

We also provide evidence that leaders who are, themselves, more dishonest in the first stage of the experiment are more likely to use incentives and statements to encourage dishonesty. As a consequence, being randomly assigned a less ethical leader ends up making a firm's workers act less ethically. Thus, in combination, we both show that unethical leaders beget more unethical firms and we identify a causal channel, from a leader's type, to the leaders' actions, to the influence that those actions have on followers.

Our main experiment features an element of rivalry—due to the competition between teams for shares of a prize—that may interact with leadership to foster dishonesty. To test the importance of such rivalry for driving the influence of leaders, we conduct a second experiment in which we change the nature of the externality, to one in which the harm for misreporting negatively impacts a passive third party, rather than other firms. Although this experiment yields similar qualitative results to those in the first experiment, they are smaller in magnitude and statistically weaker. This suggests that the effects of an unethical leader on unethical conduct may be stronger in environments with greater inter-firm competition.

The rest of the paper is structured as follows. In Section 2, we describe the design for our main experiment and we present the results for this study in Section 3. Section 4

presents our second experiment. Finally, we discuss our findings and conclude in Section 5.

2. Experiment 1: Design

Each experimental session consists of 20 participants interacting through computers. The experiment comprises three stages, with the first two stages constituting the main part (see Figure A.1 in the [Online Appendix](#) for an overview). In the first stage, subjects engage in an individual reporting task, in which they can act unethically by inflating their die-roll score and thereby obtain higher individual earnings at the expense of social welfare. The second stage modifies the task into one in which subjects repeatedly compete in “firms”, with varying degrees and forms of leadership, for ten periods. Our primary focus is on how leadership influences misreporting in the second stage. In the third stage, we elicit measures of perceived social norms of conduct and also several individual-level characteristics. Subjects receive the instructions for each stage separately, at the beginning of the stage, but are informed about the overall structure of the experiment from the outset.⁷

2.1. First Stage

In Stage 1, all 20 participants, $i \in \{1, \dots, 20\}$, compete for a prize of varying size, V . We introduce the possibility of cheating in the competition—which is undetectable and profitable for an individual, but socially inefficient—as a way of studying unethical behavior.

At the beginning of the stage, each subject privately rolls a fair six-sided die, the outcome of which, $p_i \in \{1, 2, \dots, 6\}$, can be thought of as a subject’s actual realized type or performance in that period. Subjects are instructed that it is this value, p_i , which they are supposed to report as their performance for the purposes of the competition. However, since only a subject observes the die roll, the performance report, r_i , can be any integer from 1 to 6.

Each subject receives a share of a prize, s_i , and this share increases in that subject’s own reported performance, r_i , relative to the performance reported by other subjects, r_{-i} . The size of the total prize available, $V(r_i, r_{-i})$, is a function of the average performance reported by all subjects. Specifically, the total prize obtains its maximum possible value as long as the average of all the r_i is equal to or below the expected mean of 20 fair die rolls, or 3.5. However, if the mean reported performance exceeds 3.5 then the size of the prize decreases linearly with the average reported performance. Thus, misreporting negatively impacts the size of the prize, and the prize is lowest when all subjects report the maximum possible performance of 6. As we note earlier,

7. We provide experimental instructions in the [Online Appendix](#).

this design feature captures the property that widespread unethical conduct can harm a society or industry.

More precisely, a subject's profit, π_i , is the subject's share of the prize, determined by the ratio of own performance to total performance, multiplied by the total size of the prize:

$$\pi_i = s_i(r_i, r_{-i}) V(r_i, r_{-i}) = \frac{r_i}{\sum_{j=1}^{20} r_j} V(r_i, r_{-i}) \quad (2)$$

with

$$V(r_i, r_{-i}) = \begin{cases} a & \text{if } \frac{\sum_{j=1}^{20} r_j}{20} \leq \mu, \\ a - b \left(\frac{\sum_{j=1}^{20} r_j}{20} - \mu \right) & \text{if } \frac{\sum_{j=1}^{20} r_j}{20} > \mu, \end{cases} \quad (3)$$

where $\mu = 3.5$, $a = 1250$, and $b = 300$. The parameters are chosen such that, under the assumption of self-interest and no psychological cost to lying, the unique Nash equilibrium is for all players to report the highest possible outcome of the die roll, $r_i = 6$ (see [Online Appendix D](#)). This yields a total prize of $V = 500$, which is considerably lower than the maximal possible total prize of $V = 1250$.

After rolling the die privately, each subject enters his or her performance report on the computer. Although we cannot detect lying at the individual level, the distribution of reported performance scores allows us to detect, statistically, the degree of misreporting (cf. Houser et al. 2012; Fischbacher and Föllmi-Heusi 2013; Gino et al. 2013). The task is performed only once in Stage 1 and afterward all subjects are informed about the average reported performance across all subjects, the total size of the prize, and their own payoff for Stage 1.

2.2. Second Stage

In Stage 2 we use the same task, repeated for ten periods, but this time in the context of competing "firms." The 20 subjects in a session are randomly matched into five four-person firms—consisting each of three workers and one leader (referred to as the "supervisor"). Workers individually and privately each roll a die and report performance, as in Stage 1. The function of the leaders varies by condition.

Similarly to Stage 1, firms compete for shares of a prize, V . Each firm, $f \in \{1, 2, \dots, 5\}$, obtains a share, s_f , based on the average reported performance by the three workers in that firm, $r_f = (\sum_{i=1}^3 r_{if})/3$, relative to the average reported performance in other firms. As in Stage 1, V is highest when the average of all firms' reported performance levels is no greater than 3.5, but decreases for higher average reported performance across the industry. The profit obtained by each firm is then:

$$\pi_f = s_f(r_f, r_{-f}) V(r_f, r_{-f}) = \frac{r_f}{\sum_{g=1}^5 r_g} V(r_f, r_{-f}) \quad (4)$$

TABLE 1. Overview of treatment conditions in Experiment 1.

Treatment	No incentive power	Incentive power
No public statements	Inactive Leader	Leader Incentives Only
Public statements	Leader Statements Only	Leader

Note: For each condition, we conducted four sessions with 20 subjects each.

with

$$V(r_f, r_{-f}) = \begin{cases} a & \text{if } \frac{\sum_{g=1}^5 r_g}{5} \leq \mu, \\ a - b \left(\frac{\sum_{g=1}^5 r_g}{5} - \mu \right) & \text{if } \frac{\sum_{g=1}^5 r_g}{5} > \mu, \end{cases} \quad (5)$$

where $\mu = 3.5$, $a = 1250$, and $b = 300$.

Each subject in firm f receives an individual share x_{if} of the firm's profit π_f . The leader always receives a share of one-fourth of the total firm profit, $x_{Lf} = 0.25$. The shares received by workers can vary across experimental conditions. In some of our conditions, each of the four firm members receives an equal share of the firm's profits, $x_{if} = 0.25$, $i \in \{L, 1, 2, 3\}$. If this is the case, each individual worker benefits from reporting the highest possible performance level, or 6, regardless of what other subjects report (see [Online Appendix D](#)). In other experimental conditions, described below, the firm supervisor can determine the share of the firm's profits received by each worker. In these cases, the supervisor's reward strategy can either strengthen or weaken the incentives to report performance of 6.

Roles of workers and supervisors are allocated randomly within each firm, and remain fixed during all 10 periods of Stage 2. After each period, all subjects are informed about their own profit, their firm's profit, the average reported performance level of each of the five firms, the overall average reported performance level, and the individual reported performance levels and profits of all members of their own firm.

2.3. Leader Conditions

We study four treatment conditions, in a 2×2 design (see Table 1), which vary the instruments available to the leader in the second stage of the experiment. In particular, we vary the ability of leaders to determine incentives, through performance bonuses, or to articulate a direction for the firm, through prominent statements to workers.

Public Statements. In the two conditions with public statements, leaders have 90 seconds at the beginning of the period—before workers roll their dice and report their performance—to send a typed message to the firm's three workers. Leaders send statements to their workers, but workers cannot reply to the leader or send messages

to one another. In the two conditions without public statements, leaders cannot send messages to workers.

Incentive Power. In the two conditions in which leaders have no incentive power, all four subjects in each firm receive the same share, 25%, of the firm's profit in a period. Thus, workers' payoffs are independent of any actions of the leader. In the conditions with leader incentive power, leaders distribute financial rewards among the workers. Specifically, leaders observe the reported performance level of each of the three workers, and the resulting firm profit, and then decide how to allocate 45% of the group profit among the three workers.⁸ The leader has to allocate the entire 45% "bonus pool" among the three workers and cannot keep any part of it. More precisely, of the profits received by the firm, π_f , the leader receives a fixed share of one-fourth, $x_{Lf} = 0.25$. The three workers each receive a guaranteed portion, 10%, plus a share of the remaining 45% of the firm profits, allocated by the leader. A worker's share of the firm's payoff is thus, $x_{if} = 0.10 + y_{if}$, where $y_{if} \in [0, 0.45]$ is the discretionary bonus allocated to that worker and $\sum_{i=1}^3 y_{if} = 0.45$. A profit-maximizing leader will always prefer workers to report higher numbers.⁹ At the end of the period, workers in all conditions are informed about the reported performance, allocated rewards, and total earnings of all workers in their firm. Importantly, this means they can infer how the leader rewards reported performance.

The *Leader* condition provides leaders with both channels through which they can influence the conduct of firm workers, that is, the leader can send statements before every period and allocate rewards at the end of each period. The *Inactive Leader* condition serves as a control—there is still a person in the role of "supervisor" and this person receives a 25% share of the firm's profits, but this person cannot do any of the things that leaders do. The two other conditions, *Leader Statements Only* and *Leader Incentives Only*, vary only in the presence or absence of these two instruments. By eliminating one instrument and keeping the other, we can study the relative importance and impact of these distinct potential influence channels. Moreover, within each condition with an active leader (i.e., excluding the *Inactive Leader* condition), we

8. We choose this incentive mechanism for a few reasons. First, subjective allocation of rewards is commonplace in firms (e.g., discretionary monetary bonuses or fixed pay increases in contexts including financial firms and academic departments, the allocation of nonfinancial rewards such as desirable office space). Second, since we are interested in unethical conduct, which is often unenforceable in pay contracts, a discretionary bonus seems appropriate. Third, we wanted to provide leaders with flexibility in their ability to allocate rewards; this is limited if we provide one specific kind of incentive contract (e.g., a piece-rate or target-based scheme) and complicates the experiment if we introduce too many schemes. Note also, in combination with the ability to make statements, our firm leaders can, in principle, specify a large variety of ex ante performance contracts.

9. Consider the following illustrative examples: assume that the average reported performance by other firms is 3, then if firm workers all report 6 a leader receives a 63% higher payoff than if they report 3. Even if all other firms report 6, the leader still receives a 32% higher payoff if his workers report 6 than if they report 3.

can test how different characteristics of and strategies employed by leaders affect the degree of misreporting.

2.4. Third Stage

Following Stage 2, we elicit subjects' perceptions of social norms regarding the appropriateness of inflating performance in the first and second stages of the experiment. We follow the method of Krupka and Weber (2013), in which subjects are provided with a description of a possibly unethical act, and then rate the "social appropriateness" of this behavior, with an incentive to match the ratings provided by others.¹⁰ Finally, we collect several psychological measures of personality traits (Protected Values toward acting dishonestly, following Gibson et al. 2013; Big Five 15-item version; Machiavellianism MACH IV) as well as sociodemographic measures.

2.5. Procedural Details

In total, 320 subjects participated in the experiment—80 subjects (20 firms) in each condition. [Online Appendix Table B.1](#) provides a summary of subject demographic characteristics. The experiment was computerized using the software z-tree (Fischbacher 2007) and subject recruitment used ORSEE (Greiner 2015). All sessions were conducted in January and February 2013 at the Birmingham Experimental Economics Laboratory of the University of Birmingham. Subjects' payoffs were denominated in experimental currency units (ECU), which were converted into GBP at the rate of 40 ECU = 1 GBP. Each session lasted about 2 hours, and subjects earned, on average, £19.94, including a show-up fee of £2.50.

3. Results

In presenting our results, we first discuss subjects' behavior in Stage 1, to get a sense of the prevalence of dishonesty. Then, we turn to our main research question—whether leaders' characteristics and actions influence workers' performance reports in Stage 2.

3.1. Individual Behavior in Stage 1

Table 2 shows the average reported performance in Stage 1, by condition. Recall that Stage 1 is identical across all conditions; instructions for Stage 2 are only distributed after Stage 1. Therefore, unsurprisingly, individual behavior in the first stage does not differ significantly between the four conditions (two-sided Kruskal–Wallis test

10. For instance, we ask subjects to rate how socially appropriate it is for a subject who rolled a 1 in the first stage of the experiment to report a higher number. We incentivize their answer by giving them an extra £0.5 if their answer matches the answer of a randomly chosen other subject in their session.

TABLE 2. Average reported performance in Stage 1.

Inactive Leader	Leader	Leader Incentives Only	Leader Statements Only	Total
4.48 (0.17)	4.49 (0.17)	4.59 (0.17)	4.44 (0.15)	4.50 (0.08)
$N = 80$	$N = 80$	$N = 80$	$N = 80$	$N = 320$

Note: Standard errors in parentheses.

$p = 0.85$).¹¹ In all conditions, we find evidence of misreporting. Overall, the mean reported score is approximately 4.5, which is higher than the empirically expected mean score, of 3.5. The distribution of reported performance levels is also highly right-skewed: 26% of observations are 3 or lower, whereas 77% are 4 and above, with the median at 5 and mode at 6. Table B.2 in the [Online Appendix](#) shows that, in every condition, the frequency of scores of 1 (6) is significantly lower (higher) than the expected frequency of 1/6. Note, however, that many reported scores are still well below the individual payoff-maximizing report of 6.

3.2. Do Unethical Leaders Produce Unethical Groups?

We next consider behavior in Stage 2, when subjects performed the task collectively in firms, sometimes with active leaders. We begin by analyzing whether unethical leaders influence misreporting among their workers. For this, we construct a measure of a leader's propensity to act honestly. Specifically, we use a binary variable, *dishonest leader*, which indicates whether a leader's reported Stage 1 performance equals 6. Since the frequency of scores of 6 in Stage 1, equal to 37.5%, is more than twice as high as the expected frequency of 1/6 (see [Online Appendix Table B.2](#)), we expect this variable to be correlated with misreporting. That is, conditional on observing a reported performance of 6 in Stage 1, it is more likely that an individual misreported than otherwise. According to this classification, 19 of the 60 leaders in the three conditions with active leaders are coded as dishonest.¹²

11. Mean reported performance differs little for subjects subsequently assigned to be leaders (4.43) and workers (4.52). For comparison, the mean reported die-roll outcomes of the "externality treatment" in Fischbacher and Föllmi-Heusi (2013) is 4.18, which is lower than what we find (t -test, $p = 0.10$). We also investigate whether individual characteristics, measured at the end of the experiment, predict Stage 1 responses. Regression analysis shows that reported performance in the first stage is significantly higher if the subject is male, older, studies economics, or scores lower on the Big Five conscientiousness dimension (Table B.3 in the [Online Appendix](#)).

12. The results are similar if we, instead, define the indicator variable as those leaders who reported performance of 5 or 6 in Stage 1 or who reported a Stage 1 performance higher than the average performance reported in their session. Of course, these are all noisy measures of dishonesty—for example, they misclassify some honest leaders who actually obtained scores of 6 in Stage 1. Our goal is to use a measure that captures a tendency toward high scores, which likely captures a greater proportion of those who misreport. To the extent there is noise in our classification, the results will likely underestimate the impact of unethical leaders.

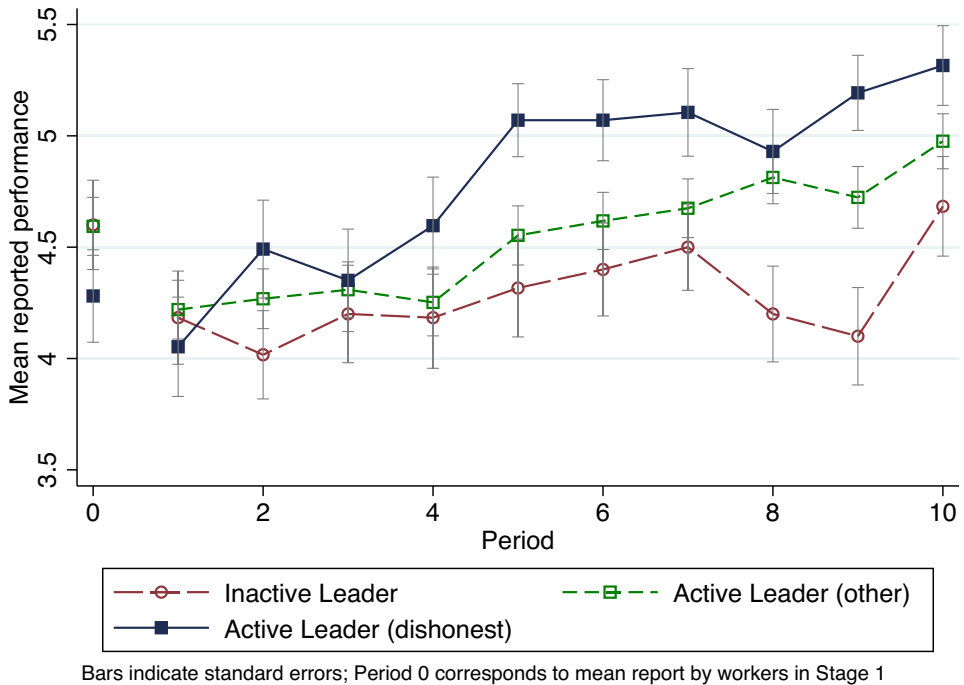


FIGURE 1. Average reported performance over time by leader type.

Figure 1 shows average reported group performance over time, based on the type of leader assigned to a firm. For example, the line marked by circles indicates average reports in firms with inactive leaders. The distance of these mean scores from 3.5 provides us with an estimate of the degree of misreporting. We divide the observations from the active leader conditions, which we pool here, depending on whether the leader is classified as dishonest or not. The graph also includes, as “Period 0”, the average Stage 1 performance of leaders and workers in the respective conditions. In all cases, there is a general tendency for the mean report to be greater than 3.5.¹³

In the Inactive Leader condition, average reported performance is fairly stable over time, with only a slight upward trend, and the average reported performance is generally similar to that from Stage 1. Thus, although there is some evidence of misreporting in the absence of active leaders, its incidence does not change much over time—with the exception, perhaps, of the last period—and remains relatively modest in magnitude.

13. Since, in Stage 2, we have multiple observations per worker, we can also explore the tendency to misreport at the individual level. Across all conditions the mean of the 10 reported die rolls at least marginally statistically significantly differs from the expected mean of 3.5 for 54.17% of all workers (i.e., comparisons using *t*-tests yield significance levels of $p < 0.1$). Of those, only 4.62% report a lower mean than 3.5, which is roughly what one would expect by chance, and the lowest mean is 2.1. This suggests that the tendency to decrease performance reports to benefit others is not widespread in our experiment.

In contrast, reported performance is higher and increases more strongly over time in conditions with active leaders. This increase in reported performance is stronger for groups with leaders classified as dishonest. Considering the average reported performance by a worker across all periods, the mean report is higher for those led by an active leader classified as dishonest (4.82), than for those led by other active leaders (4.54) or without an active leader (4.28), and both comparisons are statistically significant ($p = 0.032$ and $p = 0.003$, respectively, two-sided Mann–Whitney tests).¹⁴ Moreover, the distributions also differ substantially between workers led by dishonest leaders and other workers (see [Online Appendix](#) Figure A.2). For example, the frequency of reports of 6 is 34% with inactive leaders, 36% with active leaders not classified as dishonest, and 53% with dishonest active leaders.

Table 3 tests the relationship between leaders' honesty and followers' behavior. Model 1 regresses the reported performance of workers in Stage 2 on whether there is an active leader (i.e., whether the condition is Leader, Leader Incentives Only, or Leader Statements Only) and on whether the active leader is also a dishonest leader. That is, the coefficient for "active leader" measures the effect of having an active leader who is not dishonest, whereas "dishonest leader" measures the *additional* effect of having a dishonest leader. We also control for time trends and for a worker's propensity to misreport, measured by the worker's reported die roll in Stage 1. Given the high frequency of reports at the upper boundary of 6, we use a tobit specification in all analyses of workers' reported performance. However, the results are similar using either linear or ordered probit specifications.¹⁵

Consistent with Figure 1, model 1 shows that simply having an active leader produces slightly higher reported performance than an inactive leader, but this effect is not statistically significant. However, having a dishonest leader has a large and statistically significant additional effect on reported performance.¹⁶ Model 2 introduces interaction effects to allow for different time trends across the conditions. The indicator variables for active and dishonest leaders are now both statistically insignificant, but there is a positive time trend in all conditions and, consistent with Figure 1, the time trend for active leaders is significantly more positive than that for inactive leaders. The time trend for dishonest leaders is large in magnitude, indicating that mean reports increase by 0.118 more per period than those for groups led by active leaders who

14. As a placebo, we would not expect dishonest leaders to affect reported performance in the Inactive Leader condition. Indeed, average reported performance is very similar for dishonest (4.25) and other (4.29) leaders.

15. In [Online Appendix C](#), we report linear and ordered probit models for all tobit specifications employed in the paper. Note that only one observation is right censored in model 4—in only one group did all workers report 6 in all periods—suggesting that the tobit specification is unnecessary; comparison with the linear specification in Appendix Table C.3a reveals almost identical results.

16. Recall that the coefficient for *dishonest leader* identifies the marginal effect of having an active dishonest leader over merely having an active leader. The sum of coefficients for *dishonest leader* and *active leader* provides the effect of a dishonest leader, relative to an inactive leader. This statistic is positive and statistically significant in models 1 ($\chi^2(1) = 12.69, p < 0.001$), 3 ($F_{1,237} = 8.29, p = 0.004$), and 4 ($F_{1,77} = 8.94, p = 0.004$).

TABLE 3. Leader honesty and reported performance.

Dependent variable	Reported Stage 2 performance			
	(1)	(2)	(3)	(4)
Unit of analysis	Worker in a period		Worker mean across periods	Group mean across periods
Active leader	0.280 (0.235)	-0.176 (0.282)	0.235 (0.164)	0.263* (0.153)
Dishonest (active) leader	0.887*** (0.292)	0.275 (0.426)	0.384** (0.187)	0.319* (0.168)
Period	0.151*** (0.024)	0.063** (0.026)		
Active leader × period		0.084** (0.038)		
Dishonest leader × period		0.118 (0.075)		
(Mean) Stage 1 report by worker(s)	0.366*** (0.058)	0.366*** (0.059)	0.193*** (0.040)	0.099 (0.060)
Constant	2.343*** (0.327)	2.814*** (0.316)	3.430*** (0.224)	3.822*** (0.318)
Number of observations	2400	2400	240	80
Obs. left censored (≤ 1)	142	142	0	0
Obs. right censored (≥ 6)	952	952	16	1
Number of workers, firms	240, 80	240, 80	240, 80	-, 80
Log likelihood	-3766.09	-3757.09	-315.57	-69.40
Wald/F	$\chi^2(4) = 84.9$ ($p < 0.001$)	$\chi^2(6) = 83.6$ ($p < 0.001$)	$F_{3,237} = 10.3$ ($p < 0.001$)	$F_{3,77} = 3.63$ ($p = 0.017$)

Notes: Active leader equals 1 in all conditions with an active leader; active (dishonest) leader equals 1 when the group has an active leader and this leader reported a 6 in Stage 1. All models report tobit regressions. Models 1 and 2 are panel specifications that use a worker's performance report in a period as an observation and include worker random effects, with bootstrapped standard errors clustered at the firm level. Model 3 uses the average report by a worker across all periods, with standard errors clustered at the firm level. Model 4 uses the average report within a firm across all periods, with robust standard errors. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

are not dishonest. Although this coefficient is not statistically significant, tests of the restriction that $\beta_{dishonest\ leader} + t\beta_{dishonest\ leader \times period}$ equals zero are rejected at $p < 0.05$ for all $t \geq 3$ ($\chi^2(1) \geq 4.36$). Hence model 2 provides statistical support for the persistent differences that arise for dishonest leaders after three periods in Figure 1. Finally, models 3 and 4 test the effects of the different leader types when aggregating the data, respectively, at the level of the individual or group. In both cases, dishonest leaders produce greater misreporting than either inactive leaders or than other active leaders and the coefficients are at least marginally statistically significant.

3.3. How Do Leaders Affect Workers' Honesty?

We next investigate more precisely how the strategies employed by leaders affect workers' behavior. We begin with the ability to distribute financial incentives and then analyze the role of statements.

In conditions with leader incentive power, leaders can choose how to distribute rewards among workers, after observing their reported performance. Leaders can either reward high or low reported performance or they can distribute the bonus independently of workers' performance reports. If workers use their leader's past behavior to form expectations of how rewards will be distributed, this can affect the expected profitability of inflated performance reports. Therefore, we construct a measure, for each firm in period t , of the cumulative correlation between workers' performance shares and their bonus shares through period $t - 1$. Specifically, let $s_{itf} = r_{itf} / \sum_{j=1}^3 r_{jif}$ represent worker i 's share of the firm's total performance reports in a period, and let $b_{itf} = y_{itf} / \sum_{j=1}^3 y_{jif}$ be the share of the total bonus received by worker i . Then, for a leader f in period t , ρ_{ft} is the Pearson correlation for all s_{itf} and b_{itf} up through period t . This measures leaders' use of incentives to encourage dishonesty up to a particular period: $\rho_{ft} > 0$ indicates that a leader provided a greater share of the bonus to workers reporting high relative performance, $\rho_{ft} < 0$ indicates that the leader provided a greater share for low relative performance and $\rho_{ft} = 0$ indicates that the leader assigned bonus shares independently of relative performance.¹⁷

Table 4 studies how the leader's prior incentive use affects workers' performance reports. The dependent variable is a worker's reported performance in a period and the key independent variable is the cumulative correlation between performance shares and bonus shares awarded by the leader, through the previous period (ρ_{ft-1}). As before, we employ a tobit specification due to the large number of observations at the upper boundary of 6.¹⁸ All models control for a worker's performance report in Stage 1. Model 1 shows that performance reports are strongly positively impacted by how the leader distributed bonuses in the past. Model 2 controls for the worker's personal experience with performance and bonus shares in the previous period, by including the worker's lagged performance report, lagged share of bonus received, and their interaction. Workers who reported larger numbers in the previous period are likely to report higher numbers again, but their share of the bonus received and its interaction with their own bonus add little explanatory power. Models 3, 4, and 5 test whether the performance-bonus share correlation has differential effects when, respectively,

17. The variable is defined as long as both s_{itf} and b_{itf} exhibit some variation in periods $1, \dots, t$ (i.e., $s_{itf} \neq 1/3$ for some i, t and $b_{itf} \neq 1/3$ for some i, t). Since this is not always satisfied, we set $\rho_{ft} = 0$ for undefined cases, reflecting the fact that workers at that point have no information on how the leader distributes bonuses. This is sensible if the lack of variation occurs only because $b_{itf} = 1/3$ for all periods (while $s_{itf} \neq 1/3$ in at least one period), since this indicates the leader allocated bonuses equally even though performance shares varied. The cases in which $s_{itf} = 1/3$, for all i, t , are harder to interpret, since this means workers have no information on how the leader rewards high relative performance reports. If we drop all such observations from our analysis (5% of cases), the results are unchanged.

18. Table C.4a and b in [Online Appendix C](#) reports analogous linear and ordered probit specifications. The only substantive difference is that *previous period performance report* has a marginally statistically significant coefficient in the tobit and ordered probit versions of model 2, which is highly significant in the linear model.

TABLE 4. Reported performance and leader's incentive use.

Dependent variable	Reported Stage 2 performance				
	(1)	(2)	(3)	(4)	(5)
Bonus/perf. shares corr. up to previous period (ρ_{ft-1})	1.394*** (0.162)	1.181*** (0.174)	1.389*** (0.223)	1.324*** (0.214)	1.123*** (0.223)
Communication condition $\times \rho_{ft-1}$			0.017 (0.314)		
Dishonest leader $\times \rho_{ft-1}$				0.204 (0.356)	
Dishonest worker $\times \rho_{ft-1}$					0.624 (0.456)
Stage 1 report by worker	0.380*** (0.080)	0.319*** (0.077)	0.380*** (0.082)	0.386*** (0.082)	0.369*** (0.079)
Previous period performance report		0.236* (0.141)			
Previous period bonus share		-0.712 (1.476)			
Previous period performance report \times bonus share		0.184 (0.403)			
Constant	3.458*** (0.394)	2.646*** (0.565)	3.459*** (0.399)	3.413*** (0.415)	3.553*** (0.392)
Number of observations	1080	1080	1080	1080	1080
Obs. left censored (≤ 1)	51	51	51	51	51
Obs. right censored (≥ 6)	444	444	444	444	444
Number of workers, firms	120, 40	120, 40	120, 40	120, 40	120, 40
Log Likelihood	-1667.79	-1654.36	-1667.79	-1667.65	-1666.31
Wald	$\chi^2(2) = 83.7$ ($p < 0.001$)	$\chi^2(5) = 118.6$ ($p < 0.001$)	$\chi^2(3) = 89.7$ ($p < 0.001$)	$\chi^2(3) = 86.9$ ($p < 0.001$)	$\chi^2(3) = 82.0$ ($p < 0.001$)

Notes: Data from periods 2 through 10 of both conditions in which leaders allocate bonuses (Leader, Leader Incentives Only). Bonus/performance shares correlation up to previous period (ρ_{ft-1}) is the overall correlation between performance shares and bonus shares demonstrated by the leader up through the prior period. All models report panel tobit regressions that use a worker's performance report in a period as an observation and include worker random effects, with bootstrapped standard errors clustered at the firm level. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

leaders can send messages, a firm's leader is classified as dishonest and a worker is similarly classified as dishonest (i.e., the worker reported a score of 6 in Stage 1). None of these interactions yields any significant differences. Most importantly, across all regressions, we see that how the leader distributed bonuses in the past strongly affects reported performance.

We next explore the effect of leaders' statements on workers' reported performance. To this end, we conducted content analysis, relying on classifications of the statements by three independent coders. We initially developed eight categories that distinguish between messages encouraging *high* and *low* reported performance. Specifically, the categories *request high* and *request low* identify specific requests for high or low performance numbers, *praise high* and *praise low* identify praise for having reported high or low numbers, *dishonest* and *honest* identify direct appeals to either dishonesty or honesty, and *bonus high* and *bonus low* identify instances (in the condition with leader incentives) in which the leader states that bonuses will be awarded for either high or low performance reports. Coders observed the statement made by a particular leader

in a period, and then identified which categories applied to that particular statement. Where there was disagreement, we use the median coder's rating.¹⁹

Using these category codings, we construct a measure, $\omega_{jt} \in [-1, 1]$, of the tendency of a leader to make statements exerting upward (request high, praise high, dishonest, bonus high) versus downward (request low, praise low, honest, bonus low) pressure on workers' reported performance in a period. Specifically, we sum the number of high categories satisfied by a leader's statements in a period, subtract the number of low categories satisfied in that period, and divide this sum by the total number of either positive or negative categories possible in that firm's condition.²⁰

Table 5 reports random-effects tobit regressions of workers' reported performance in a period on the degree to which the leader's statements at the start of the period encourage high performance reports (ω_{jt}). As before, we employ a tobit specification and control for the worker's reported score in Stage 1.²¹ Model 1 shows that statements encouraging the reporting of high numbers lead workers to do so. In model 2, we additionally control for a worker's tendency to misreport in the previous period, which does not change the results. Model 3 tests whether the degree to which statements influence followers differs between the two conditions in which leaders made statements: Leader and Leader Statements Only. The variable, *Incentive condition* $\times \omega_{jt}$, identifies whether statements by leaders who could also set incentives (in the Leader condition) had a differential effect. The results indicate that statements by such leaders were more effective, though the effect even for leaders without this ability remains positive and statistically significant.²² Model 4 studies whether statements

19. Table B.4 in the [Online Appendix](#) reports all message categories, their definitions, and frequencies. The first eight categories are the ones that we believed, ex ante, would be the most relevant, so we focus our analysis on these. For completeness, we also included additional categories, not used in the current analysis, that identify messages referring to other groups' performance, to the size of the prize or to the group's share of the prize, and residual categories for messages containing apologies, jokes, general encouragement, or miscellaneous messages. These additional categories generally have little statistically significant effect on reported performance (see Table B.5 in the [Online Appendix](#)), with two exceptions. First, references to other groups have a marginally significantly positive effect on performance reports. Second, humor has a larger and statistically stronger positive relationship, though it is reasonable that the causality might run the other way (high performance reports elicit humorous comments from the leader). Since we did not have strong priors regarding these message categories' relationships with misreporting, we exclude them from our primary analysis.

20. That is, $\omega_{jt} = (\sum_{j=1}^K k_{jft}^{high} - \sum_{j=1}^K k_{jft}^{low}) / K$, where $k_{jft}^{high} \in \{request\ high_{jt}, praise\ high_{jt}, dishonest_{jt}, bonus\ high_{jt}\}$, $k_{jft}^{low} \in \{request\ low_{jt}, praise\ low_{jt}, honest_{jt}, bonus\ low_{jt}\}$ and $K = 4$ in the Leader condition and $K = 3$ in the Leader Statements Only condition. Hence, a score of $\omega_{jt} = 1$ ($\omega_{jt} = -1$) indicates that the leader used all of the forms of content to encourage high (low) numbers and none to encourage low (high) numbers.

21. In [Online Appendix C](#), Tables C.5a and b report linear and ordered probit specifications of the same models. The results are similar in both cases. [Online Appendix](#) Table B.6 analyzes the relationship between individual categories and workers' reported performance. The signs of all relationships are what one would expect and all categories except for *Praise high* and *Praise low* demonstrate statistically significant relationships. The strongest statistical relationships appear to be for *Request high*, *Request low*, *Honest*, and *Reward low*.

22. Based on this finding, we also studied whether leaders with incentive power employ different communication strategies than those without. The frequencies of *request high*, *request low*, *praise high*,

TABLE 5. Reported performance and leader's statement use.

Dependent variable	Reported Stage 2 performance				
	(1)	(2)	(3)	(4)	(5)
Leader's encouragement of high performance reports (ω_{ft})	2.116*** (0.496)	2.058*** (0.483)	1.450*** (0.494)	1.310*** (0.408)	2.031*** (0.509)
Incentive condition $\times \omega_{ft}$			1.668* (0.868)		
Dishonest leader $\times \omega_{ft}$				2.374** (0.925)	
Dishonest worker $\times \omega_{ft}$					0.243 (0.947)
Stage 1 report by worker	0.410*** (0.079)	0.333*** (0.081)	0.403*** (0.076)	0.435*** (0.077)	0.396*** (0.094)
Previous period performance report		0.255*** (0.095)			
Constant	3.375*** (0.379)	2.629*** (0.472)	3.378*** (0.353)	3.204*** (0.368)	3.438*** (0.425)
Number of observations	1200	1080	1200	1200	1200
Obs. left censored (≤ 1)	59	50	59	59	59
Obs. right censored (≥ 6)	565	533	565	565	565
Number of workers, firms	120, 40	120, 40	120, 40	120, 40	120, 40
Log likelihood	-1756.76	-1531.14	-1750.33	-1744.61	-1756.65
Wald	$\chi^2(2) = 44.0$ ($p < 0.001$)	$\chi^2(3) = 51.9$ ($p < 0.001$)	$\chi^2(3) = 54.6$ ($p < 0.001$)	$\chi^2(3) = 60.1$ ($p < 0.001$)	$\chi^2(3) = 45.0$ ($p < 0.001$)

Notes: Data from both conditions in which leaders send statements (Leader, Leader Statements Only). Leader's encouragement of high performance reports indicates the relative use of statement categories encouraging high (1) versus low (-1) performance reports. Models 1, 3, and 4 use all periods of data; model 2 uses periods 2 through 10. All models report panel tobit regressions that use a worker's performance report in a period as an observation and include worker random effects, with bootstrapped standard errors clustered at the firm level. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

from leaders classified as dishonest were differentially effective. This is the case: statements from dishonest leaders have a greater impact than statements from other leaders. Although we are cautious, given the post hoc nature of these two findings, they suggest that statements requesting unethical conduct can have complementary effectiveness with the characteristics of and tools available to a leader. Finally, model 5 tests whether statements directed toward workers who reported 6 in Stage 1 were more effective, finding no relationship.

Figure 2 provides a graphical representation of the above relationships, using the firm as the unit of observation. The vertical axis in each graph indicates the impact a leader has on the reports by workers. Specifically, it shows the average performance reports by all three workers in a firm across all 10 periods of Stage 2 minus the three workers' reported scores in Stage 1. The horizontal axes show the leaders' tendencies to employ the above two instruments—incentives and statements—to encourage high performance reports over the course of the experiment. In Figure 2(a), this corresponds to the correlation between the performance shares and the bonus shares awarded by a

praise low, dishonest, and honest do not differ much by leader type (never by more than 9.5%) and the differences are never statistically significant.

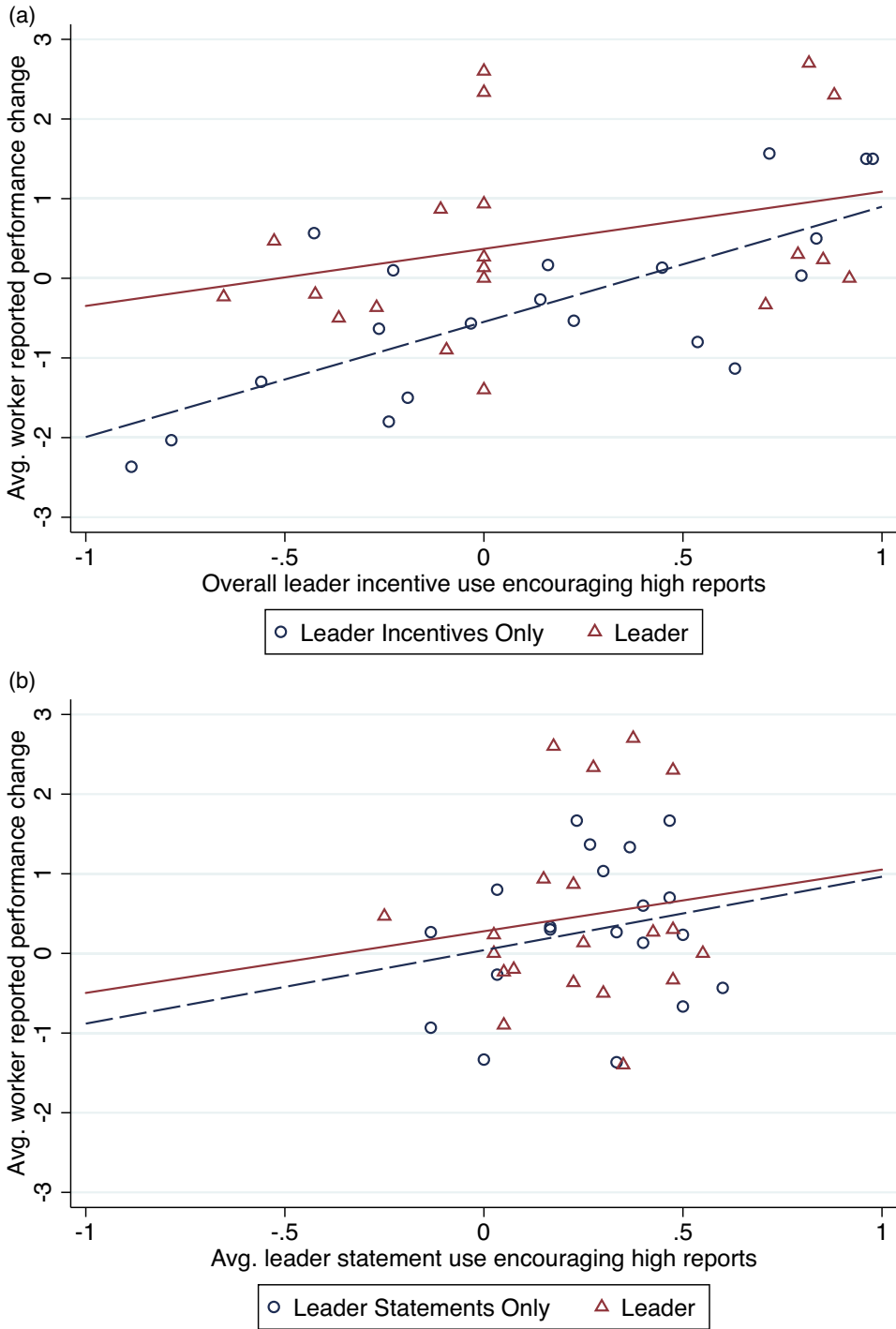


FIGURE 2. (a) Effect of leader's incentive use on workers' performance reports. (b) Effect of leader's statement use on workers' performance reports.

leader across all 10 periods of Stage 2 (ρ_{f10}); in Figure 2(b), it is the leader's average tendency to use statements eliciting high performance reports ($\bar{\omega}_f = \sum_{t=1}^{10} \omega_{ft}/10$). Both graphs provide visual evidence of the statistical relationships in Tables 4 and 5, aggregated at the level of the firm.²³

An additional interesting observation arises when comparing the two graphs. In Figure 2(a), there is considerable dispersion in how leaders award bonuses across Stage 2, from a minimum of -0.885 to a maximum of 0.997 . The mean and median values are both close to zero ($\rho_{f10}^{mean} = 0.134$; $\rho_{f10}^{median} = 0$), with the former not statistically different from zero ($t_{39} = 1.566$, $p = 0.125$). Moreover, 16 of 40 leaders (40%) have negative values of ρ_{f10} . Figure 2(b), however, reveals a different pattern. First, there is considerably less dispersion ($\bar{\omega}_f^{min} = -0.25$; $\bar{\omega}_f^{max} = 0.6$), which might be due to the coarse method in which this variable is constructed from a few noisy categories. However, there is also a clear positive tendency in leaders' statement use scores ($\bar{\omega}_f^{median} = 0.271$; $\bar{\omega}_f^{mean} = 0.25$, $t_{39} = 7.66$, $p < 0.001$), with 36 out of 40 leaders (90%) exhibiting a tendency to employ more messages requesting high, rather than low, performance reports.

3.4. Structural Estimates of Leader Influence

To complement the above reduced-form analysis of leaders' influence on followers, we estimated a structural model in the spirit of equation (1). This also allows us to explore possible heterogeneity in workers' responses to leaders' actions.

Our design has $i = 1, \dots, I$ workers in $f = 1, \dots, F$ firms, each of them providing a performance report in each of $t = 0, \dots, 10$ periods, where $t = 0$ corresponds to Stage 1 and $t \geq 1$ to Stage 2. Once a worker rolls the die, she can choose to report an outcome, r_{ift} . Assume that the worker does so with the objective of maximizing an expected utility function based on that in equation (1), with

$$E [u(r_{ift}, L)] = E[\pi(r_{ift}, L^I)] - c(r_{ift}, L^S). \tag{6}$$

We rewrite this general utility function with one specifically suited to our experimental data:

$$E [u(r_{ift}, L)] = \lambda_i \left[\left(0.25 + \gamma_i \Delta Y_{ift} \right) S_{ift} V_{jt} \right] + \left(\alpha_i + \beta_i M_{ift} \right) C_{ift} + \varepsilon_{ift}, \tag{7}$$

using the subscript, j , to denote the case in which $r_{ift} = j \in \{1, \dots, 6\}$. Hence, V_{jt} is the expected total prize available for all firms in the session in the case where the worker reports a performance of j and group f receives an expected group share of

23. [Online Appendix](#) Figures A.3A and A.3B provide analogous graphs using the actual average Stage 2 reported performance, rather than the change in reported performance. The relationships are similar.

this prize, S_{jft} . The parameter, λ_i , captures how much the worker cares about her own monetary payoff. We assume that expectations, where necessary, are formed by each worker believing that all the other workers will behave as observed in the previous period. Hence, given a performance report r_{ift} by worker, i , then V_{jt} and S_{jft} follow immediately from this assumption.

The leader's use of incentives is captured by ΔY_{ijft} , which represents the expected departure for the worker from a 25% share of the group's earnings from reporting performance, j , based on the worker's expectation of how the leader will distribute this share. As in our reduced-form analysis, we assume that workers look back at the history of the leader's allocation of the bonus in prior periods, and compute an expected share of the total group reward for each possible performance report. Hence, γ_i measures the extent to which the worker's expectations of her own earnings are influenced by the leader's distribution of bonuses in the past. When γ_i equals zero, for example, the worker might not attend to how the leader has distributed bonuses in the past when forming expectations.

The second term models how the leader's statements influence workers' perceptions of the appropriateness of misreporting. Let C_{ijft} measure of the degree to which reporting outcome $r_{ift} = j$ is unethical. Then, $\alpha_i + \beta_i M_{ijft}$ represents the degree to which the worker experiences psychological (dis)utility from taking this act. Our design does not allow us to observe directly the degree to which a worker is lying, which is the difference between r_{ift} and the worker's actual performance in that period. Hence, we proxy "dishonesty" as the deviation of a worker's cumulative reported performance up to period t from 3.5. For each worker/firm/period/reported outcome we compute a measure of "cumulative dishonesty":

$$C_{ijft} = \begin{cases} 0, & \text{if } \sum_{\tau=0}^t \frac{r_{if\tau}}{1+\tau} \leq 3.5, \\ \sum_{\tau=0}^t \frac{r_{if\tau}}{1+\tau} - 3.5, & \text{if } \sum_{\tau=0}^t \frac{r_{if\tau}}{1+\tau} > 3.5. \end{cases} \quad (8)$$

This index will be 0 if the average of the worker's reported rolls up to period t is less than 3.5, assuming that there is no clear psychological impact from reporting outcomes lower than or equal to the expected value of a die roll. This index will increase the higher is the worker's average. Hence, C_{ijft} is a simple measure of how much the worker's reports, r_{if0}, \dots, r_{ift} , reflect dishonestly high reporting.

The parameter, α_i , identifies how much the worker cares, intrinsically, about C_{ijft} . A worker who dislikes appearing dishonest will have $\alpha_i < 0$. This intrinsic weight can be modified by a leader's statements. Specifically, let,

$$M_{ijft} = \begin{cases} 0, & \text{if leader sends no/ambiguous statement,} \\ 1 - 0.4(r_{ifft} - 1), & \text{if leader requests low reports,} \\ -1 + 0.4(r_{ifft} - 1), & \text{if leader requests high reports,} \end{cases} \quad (9)$$

capture the leader's use of statements in a period to influence reports. If the leader sends no message or an ambiguous message, then $M_{ijft} = 0$ and a worker who intrinsically cares about appearing honest ($\alpha_i < 0$) will experience disutility $\alpha_i C_{ijft}$. If the leader makes a request for either high or low numbers,²⁴ then we define $M_{ijft} = 1$ when the worker follows this request (i.e., $r_{ifft} = 1$ in response to a low request, $r_{ifft} = 6$ in response to a high request) and $M_{ijft} = -1$ when the worker makes the most contrary report (i.e., $r_{ifft} = 6$ in response to a low request, $r_{ifft} = 1$ in response to a high request). The term, β_i , then identifies the extent to which a worker's concern for appearing dishonest, when $C_{ijft} > 0$, is influenced by the leader's statements.²⁵

We estimate the model in equation (7) as a two-component finite mixture model using maximum likelihood. We let ε_{ijft} be a random shock distributed as type 1 extreme value. We construct the following log-likelihood function:

$$L(p, \theta) = \sum_i \log \left\{ \sum_k p_k \left[\prod_t \left(\prod_j Pr \left(r_{ikft} = j | \theta_k, X_{ijft} \right)^{d_{ijft}} \right) \right] \right\}, \quad (10)$$

where $Pr(r_{ikft} = j | \theta_k, X_{ijft})$ is the logit probability of reporting a roll j based on the utility function specified above, conditional on the k type-specific parameters $\theta_k = \{\lambda_k, \alpha_k, \gamma_k, \beta_k\}$; p_k are the type probabilities, $X_{ijft} = \{\Delta Y_{ijft}, S_{ijft}, V_{jt}, M_{ijft}, C_{ijft}\}$ are the variables identified from the data as described above, and d_{ijft} are dummies for actual reported performance.

Table 6 reports the results.²⁶ The model estimates 74% of workers as Type 1, who care about their expected revenue ($\lambda > 0$) and dislike (the appearance of) lying ($\alpha < 0$). This type is responsive both to expectations of how the leader will distribute the bonus ($\gamma > 0$) and to the leader's statements ($\beta > 0$). Hence, consistent with our reduced form analysis, a large number of workers are influenced by their leaders through both

24. We define a leader requesting high (low) reports in a period if the leader uses more of *any* of the message categories that we earlier classified as “high” (“low”) request than of *any* the categories that correspond to “low” (“high”) requests, that is, $\omega_h > 0$ ($\omega_l < 0$). A leader who makes a statement with the same number of both types of categories or with neither type of category ($\omega_h = 0$) is classified as sending no or an ambiguous message.

25. Suppose, for instance, that the leader requests a high number, then a worker reporting a performance of 6 will experience $\beta_i(C_{i6ft} - 0.6C_{i5ft})$ higher utility from reporting a 6 than a 5. If, instead, the leader requests a low number, then the worker experiences a $-\beta_i(C_{i6ft} - 0.6C_{i5ft})$ utility impact from reporting a 6 than a 5.

26. If we estimate a one-component model, the main variables of interest, γ and β , which measure leader influence, are both positive and statistically significant, consistent with the estimates in Table 6. However, α is positive and statistically significant, suggesting that subjects want to be seen as reporting high numbers. Instead, we interpret this as the model's attempt to capture a disproportionately high frequency of six reports (in combination with the presence of lower reports) than can be captured by the utility component alone. The two-component model allows for heterogeneity in propensity to report 6, which better accommodates the data. Moreover, we find a significantly higher value of the likelihood with two types than one type (LR test statistic: $\chi^2(5) = 252.8, p < 0.001$). We also experimented with three components, finding a third type with zero probability, which suggests the bimodal distribution is sufficient to capture workers' preferences in the data.

TABLE 6. Estimates for worker utility model.

	Type 1	Type 2
Type probability	0.740*** (0.012)	0.260*** (0.012)
λ –Own revenue	0.054*** (0.005)	0.542*** (0.015)
α –(Dis)utility from lying	–0.235*** (0.018)	–0.023 (0.024)
γ –Weight on leader incentive use	0.071*** (0.017)	0.000 (0.011)
β –Weight on leader statements	0.515*** (0.014)	0.972*** (0.019)
Number of observations		14,400
Number of workers		240
Log likelihood		–3,653.91

Note: Parameter estimates of model in equation (7) using data from all periods in all conditions with active leaders. An observation is a worker's decision of whether to make a particular performance report in a period. Standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

channels. The second type, consisting of 26%, places relatively greater weight on own revenue and little weight on appearing dishonest. Presumably, these are subjects who exhibit a high propensity to report high scores. They are not influenced by how the leaders distribute incentives, but are sensitive to the leader's statements in the same manner as Type 1.

These estimates confirm our earlier observations that workers respond to leaders. Leaders influence a majority of workers both through how they reward reported performance and through their statements encouraging either high or low performance. The results also indicate that the different channels available to a leader might influence different types of workers. In particular, the incentive channel may be more influential toward workers intrinsically inclined to care about being or appearing dishonest.

3.5. Strategies Employed by Unethical Leaders

We next investigate whether unethical leaders are more likely to employ strategies that encourage dishonest reporting. For this, we return to the earlier classification of leaders as “dishonest” who reported performance of 6 in Stage 1. We then compare dishonest with other leaders' use of statements and incentives to encourage dishonesty. For incentive use, we compare the overall correlation between bonus shares and performance shares exhibited by a leader across all 10 periods, ρ_{f10} . Dishonest leaders exhibit more positive correlation between bonus and performance shares (0.269) than other leaders (0.043). For statement use, we use the average tendency to make statements requesting high, rather than low performance reports, $\bar{\omega}_f = \sum_{t=1}^{10} \omega_{ft}/10$. Again, we find a greater tendency to encourage dishonesty among dishonest leaders

TABLE 7. Leaders' strategies by leader type.

Dependent variable	Incentive use in period (ρ'_{ft})		Statement use in period (ω_{ft})	
	(1)	(2)	(3)	(4)
Dishonest leader	0.152 (0.134)	0.579*** (0.188)	0.150*** (0.055)	0.176** (0.086)
Period		0.032* (0.016)		0.023*** (0.005)
Dishonest leader \times period		-0.077*** (0.025)		-0.005 (0.014)
Constant	0.034 (0.083)	-0.140 (0.123)	0.201*** (0.042)	0.074 (0.050)
Number of observations	400	400	400	400
Number of firms	40	40	40	40
R^2	0.013	0.040	0.044	0.078
Wald	$\chi^2(1) = 1.30$ ($p = 0.254$)	$\chi^2(3) = 11.1$ ($p = 0.011$)	$\chi^2(1) = 7.45$ ($p = 0.006$)	$\chi^2(3) = 30.3$ ($p < 0.001$)

Notes: Models 1 and 2 include data from both conditions in which leaders allocate bonuses (Leader, Leader Incentives Only). Incentive use in period refers to the correlation between performance shares and bonus shares in a period. Models 3 and 4 include data from all periods of both conditions in which leaders send statements (Leader, Leader Statements Only). Statement use in period is the tendency to use message categories encouraging high (1) or low (-1) performance reports. All models report panel linear regressions that use a leader's action in a period as an observation and include firm random effects, with standard errors clustered at the firm level. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

(0.351) than other leaders (0.201). [Online Appendix](#) Figure A.4 provides graphical representations of how the different types of leaders used incentives and statements, both on aggregate and over time. The differences are similarly large in magnitude and they are statistically significant for statements (Mann–Whitney: $z = 2.182$, $p = 0.03$) but not for incentive use (Mann–Whitney: $z = 1.080$, $p = 0.28$). If we look only at the first period, the difference in incentive use is much larger (0.73, $z = 3.380$, $p = 0.007$), but this difference shrinks quickly over time. The difference is fairly constant across periods for statement use.

These patterns are confirmed in the regression analysis in Table 7, which studies the behavior of a leader in a period. The first two models study incentive use, with the independent variable the correlation between performance and bonus shares in a period—rather than cumulative through a period.²⁷ We observe a greater tendency for dishonest leaders to use incentives to encourage dishonesty, but this is not significant in model 1. In model 2, the difference is initially large and significant, but declines across periods. Models 3 and 4 study statement use, employing our earlier measure of a leader's tendency in a period to use messages encouraging high

27. As discussed in footnote 16, we use a value of zero when either the bonus share or performance share exhibits no variation. This is sensible in cases where performance varies but the bonus share does not (i.e., cases in which the leader could have rewarded high or low performers but chose not to do so). In [Online Appendix](#) Table B.7 we repeat the analysis in Table 7, dropping all periods in which the performance shares of the three workers exhibited no variation. The results are very similar.

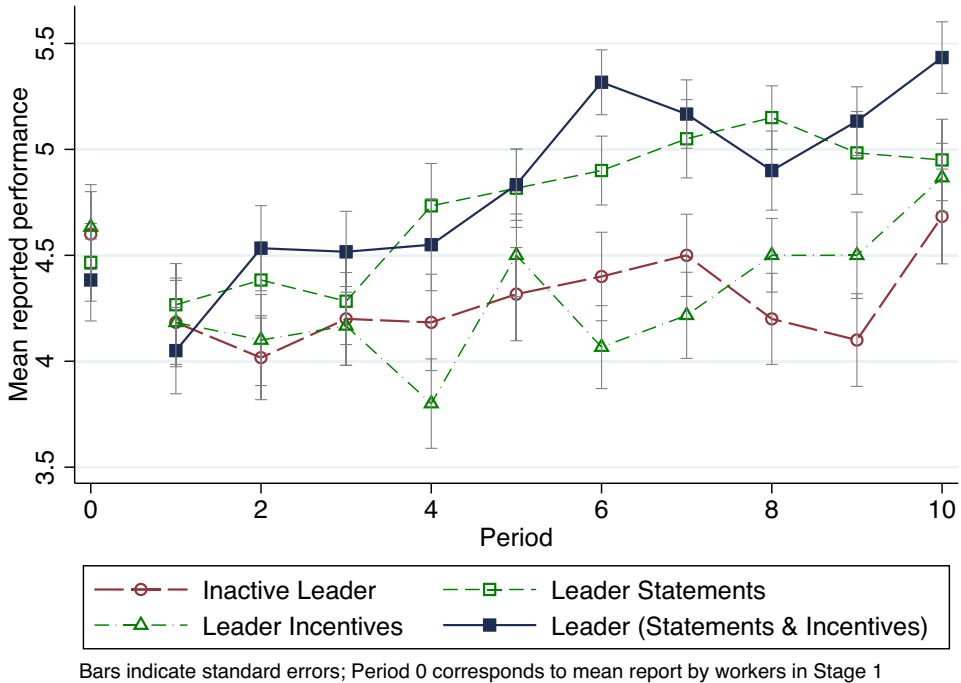


FIGURE 3. Average reported performance over time by leader type.

or low performance reports. Models 3 and 4 confirm that dishonest leaders make more statements encouraging dishonesty. This difference does not decrease over time, though there is a tendency for more statements encouraging dishonesty across periods for both types of leaders.

Hence, we find evidence that leaders who are more dishonest are more likely to use strategies that encourage followers to act dishonestly, particularly through their statements. In combination with the analysis in Section 3.3 and the structural analysis in Section 3.4, which both show that these leader strategies influence performance reports, this provides a basis for the relationship that we identified in Section 3.2 between a leader's type—dishonest or other—and worker's performance reports.

3.6. *The Aggregate Effects of Leadership*

We have seen that leaders influence the conduct of followers. We next focus on the overall effect of leadership, as an institutional feature. Our experimental conditions, which exogenously vary the presence and abilities of leaders, allow us to do so. Figure 3 shows the average reported performance across periods in each of the four conditions. This graph does not draw a distinction between dishonest and other leaders, but instead shows aggregate performance reports across all firms in a condition.

The figure indicates that both conditions in which leaders make public statements tend to yield higher misreporting than when leaders are inactive or can only determine incentives. Average reported performance when leaders can only make statements to workers largely overlaps that of groups in the Leader condition, where they also control incentives. In contrast, both the level and the trend of group performance in the Leader Incentives Only condition follow closely that of groups in the Inactive Leader condition.²⁸ Regression analysis (see [Online Appendix Table B.8](#)) confirms that these aggregate treatment effects arise with the leaders' abilities to make statements to followers.

Hence, although we saw earlier that leaders can influence followers through their ability to make statements as well as determine incentives, the overall effects of leadership seem to affect outcomes primarily through the former. Although surprising, this is consistent with our earlier observation that leaders—even leaders we do not classify as dishonest—are more likely to encourage dishonesty through their statements than through their incentive use, and that such use of statements increases over time (see [Online Appendix Figure A.4](#)). Hence, at least in our experiment, the extent to which leadership encourages dishonesty seems to be more strongly driven by leaders' statements. Of course, as these are post hoc observations and interpretations, we note them cautiously.

Finally, we also explore whether the effects of leadership extend to a broader change in perceptions of appropriate and inappropriate conduct. At the end of the experiment, we elicited workers' social norms regarding the appropriateness of misreporting the outcome of the die roll in Stage 1. We used the incentivized norm elicitation method introduced by Krupka and Weber (2013). At the end of each session, subjects were asked to rate the appropriateness of overstating one's performance in Stage 1, on a scale from "very inappropriate" (1) to "very appropriate" (4). Subjects were rewarded if their answer matched that of a randomly drawn other participant in the session. This procedure captures *social* perceptions of appropriateness, a central component of injunctive social norms. Consistent with leaders having some form of persistent influence, being in any condition with active leaders yielded perceived norms of conduct that are more lax with respect to misreporting than being in the Inactive Leader condition (see [Online Appendix Figure A.5](#)).²⁹ Thus, exposure to a leader who has some channel through which to exert influence on the group changes workers' perceptions regarding the appropriateness of acting unethically in other environments,

28. Two-sided Mann–Whitney tests show that reported performance is significantly different between the Inactive Leader and the Leader Statements Only and Leader conditions ($p < 0.001$ and $p < 0.001$), as well as between the Leader Incentives Only and the Leader Statements Only and Leader conditions ($p < 0.001$ and $p < 0.001$). But the difference between the Inactive Leader and Leader Incentives Only condition is insignificant ($p = 0.734$) as well as the difference between the Leader Statements Only and the Leader condition ($p = 0.144$).

29. We do not find a significant difference between the appropriateness ratings given by workers who were led by a dishonest leader compared to those led by an honest leader. This might be because the incentives were to match the choice provided by another participant in the session, rather than in one's firm.

where the leader is not active. This suggests that leaders do more than change behavior by exerting direct influence; they may also change norms and values.

4. Experiment 2: Absence of Intergroup Competition

An important feature of our main experiment is that widespread inflation of performance reports creates a negative externality for all firms and workers, and decreases the total surplus. As we note earlier, this corresponds to many settings—such as several of our motivating examples—involving competition, where high levels of unethical conduct destroy aggregate value. However, it is also worthwhile to investigate to what extent leaders' effects on followers' ethical conduct extend to other kinds of environments. For this purpose, we conducted a second experiment.

4.1. Experimental Design

The design closely follows that of the first experiment, apart from the fact that the negative externality from cheating is not borne by the other groups, but by a passive third party. Each session now features 20 active participants and 5 inactive participants, with roles fixed throughout the experiment. Inactive participants do not make any decisions, but their payoffs are determined by active participants' reported die roll outcomes.

In Stage 1, each inactive participant is randomly matched with four active ones. Each participant begins with 50 ECU. As before, active participants each privately roll a die and report the outcome. They each receive a payoff equal to $\pi_i^{active} = 50 + 5(r_i - 3.5)$. The latter term corresponds to the amount that each active participant takes from her matched inactive participant. Hence, each inactive participant, who did not roll a die, receives a payoff determined by the reported scores of the four matched active participants: $\pi^{inactive} = 50 - \sum_{i=1}^4 5(r_i - 3.5)$. Thus, for example, reporting $r_i = 6$ means that participant i receives a payoff of 62.5, but if all four active participants report this outcome the inactive participant receives a payoff of zero. At the end of Stage 1, each subject observes the average reported performance by all subjects, her own payoff and, if active, her own reported performance.

In Stage 2, firms are again comprised of one supervisor and three workers, randomly drawn and assigned to firm roles from the set of active participants. Workers roll dice to determine performance and are free to enter any performance report. In each period, the firm is also randomly matched with an inactive participant. As in our main experiment, the firm's total payoff is determined by the sum of the three workers' performance reports. Now, however, the firm benefits only by taking from the matched passive participant. Specifically, a firm's payoff in a period is, $\pi_f = 200 + 20 ((\sum_{j=1}^3 r_j / 3) - 3.5)$, meaning that the firm benefits by 20 ECU for each unit increase in the performance report. Conversely, the inactive participant's

payoff is $\pi^{inactive} = 50 - 20 ((\sum_{i=1}^3 r_j/3) - 3.5)$, so that the firms' benefits from inflated performance reports come directly from the passive third party.³⁰

We included only two leadership conditions from the first experiment: Inactive Leader and Leader. In the former, leaders are passive observers and all firm members receive equal shares of the firm's payoff. In the latter, as before, leaders make statements to workers at the beginning of the period and distribute 45% of the total payoff as a bonus between the three workers.

Information at the end of a period is similar to that in our first experiment. After each period in Stage 2, each active participant observes the reported individual performance levels and payoff of each worker in her firm; the average performance reported by all firms in that period and in all previous periods; her firm's payoff; and the payoff of the inactive participant matched with her firm for that period. Inactive participants only observe their own payoff and the average reported performance level for each firm in the current and previous periods.

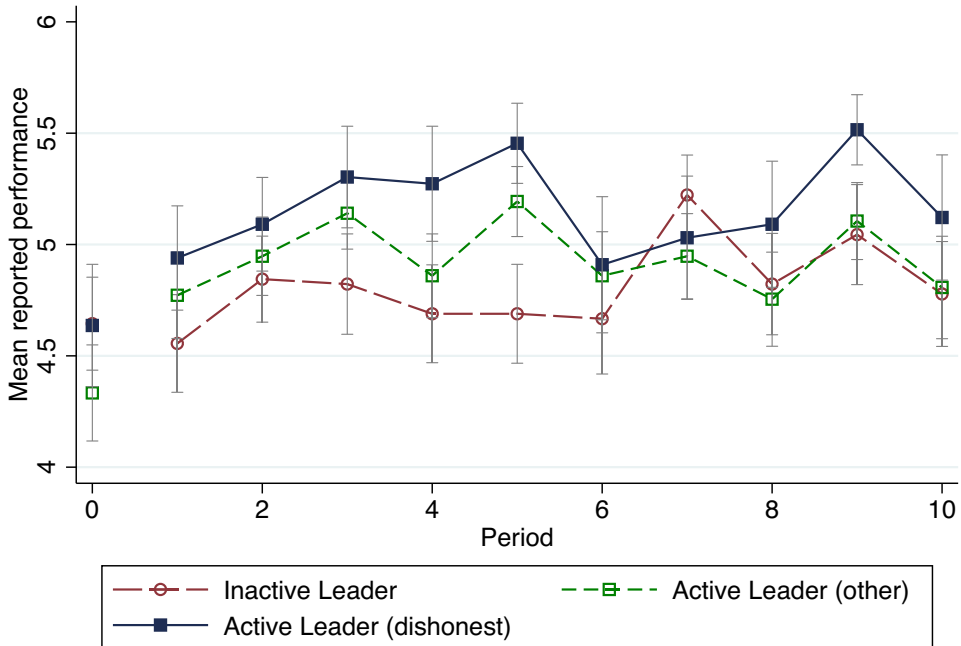
We conducted the experiment at the Laboratory for Experimental and Behavioral Economics at the University of Zurich. In total, we conducted three sessions (15 firms, 60 active participants, and 15 inactive participants) of the No Leader condition and six sessions (30 firms, 120 active participants, and 30 inactive participants) of the Leader condition. Sessions were conducted in English, using only subjects proficient in English. At the end of the session, subjects' payoffs were converted into cash at a rate of 15 ECU = 1 Swiss Fr.

4.2. Results

We again classify active leaders as "dishonest" or "other" based on whether their reported Stage 1 score is 6. This yields 11 dishonest leaders and 19 not classified as such.

Figure 4 shows the effects of these different kinds of leaders on reported performance, providing similar data to Figure 1. The ordering of the three different kinds of firms is similar to that in Experiment 1: workers in groups with dishonest leaders again exhibit the highest performance reports, on average, across all 10 periods (5.17), followed by groups with other leaders (4.94). The lowest performance reports tend to be for firms with No Leader (4.81). The distributions also show the same patterns as in Experiment 1, with a shift toward higher reports for firms with dishonest leaders (see [Online Appendix Figure A.6](#)). For example, the frequency of performance

30. Although we attempted to vary as little as possible relative to our main experiment, some changes were hard to avoid. For example, this payoff system no longer makes misreporting inefficient. Attempting to build this inefficiency into the design would require the loss for the inactive participant exceeding the gain for the firm, and either leaving the inactive participant with negative payoffs or having this participant start off with more wealth (per capita) than the firm members. Our design thus potentially makes misreporting more attractive than in our first experiment, which is consistent with our data.



Bars indicate standard errors; Period 0 corresponds to mean report by workers in Stage 1

FIGURE 4. Average reported performance over time by leader type (Experiment 2).

reports of 6 across all 10 periods is 66% for firms with leaders classified as dishonest, 54% for firms with other leaders, and 48% under No Leader.³¹

Table 8 provides tobit regressions analyzing the effects of different leader types, analogous to Table 3 for Experiment 1. The magnitude of the coefficients for dishonest leaders is always positive and is similar in size to those in Table 3; but they are less precisely estimated and, therefore, not statistically significant.³² Nevertheless, relative to the No Leader control, we observe statistically significantly higher performance reports in the firms led by dishonest leaders. For example, the sum of the first two coefficients in model 1—which, in combination, capture the effect of active and dishonest leaders relative to No Leader—is significantly different from zero

31. Interestingly, the overall levels of dishonesty tend to start off higher in Experiment 2 than in Experiment 1, but increase less over time. Although we are cautious to make too much of comparisons between the experiments, a natural interpretation is that efficiency concerns lead workers to misreport less initially in Experiment 1, but that competition in Experiment 1 creates an incentive for firms to “outperform” each other over time.

32. We again employ a tobit specification given the rate of censoring at 6. We provide analogous linear and ordered probit results in Appendix Tables C.6a and b. Results are again substantively similar.

TABLE 8. Leader honesty and reported performance (Experiment 2).

Dependent variable Unit of analysis	Reported Stage 2 performance			
	(1) Worker in a period	(2)	(3) Worker mean across periods	(4) Group mean across periods
Active leader	0.533 (0.429)	0.785* (0.470)	0.242 (0.238)	0.218 (0.192)
Dishonest (active) leader	0.844 (0.621)	0.595 (0.611)	0.315 (0.320)	0.173 (0.263)
Period	0.060*** (0.022)	0.080*** (0.029)		
Active leader × period		-0.047 (0.047)		
Dishonest leader × period		0.047 (0.076)		
(Mean) Stage 1 report by worker(s)	0.602*** (0.115)	0.602*** (0.115)	0.263*** (0.062)	0.247** (0.111)
Constant	2.871*** (0.477)	2.760*** (0.488)	3.676*** (0.271)	3.666*** (0.481)
Number of observations	1350	1350	135	45
Obs. left censored (≤ 1)	56	56	0	0
Obs. right censored (≥ 6)	737	737	28	2
Number of workers, firms	135, 45	135, 45	135, 45	-, 45
Log likelihood	-1722.8	-1722.3	-184.8	-42.8
Wald/ F	$\chi^2(4) = 39.1$ ($p < 0.001$)	$\chi^2(6) = 41.8$ ($p < 0.001$)	$F_{3,132} = 8.04$ ($p < 0.001$)	$F_{3,42} = 3.73$ ($p = 0.018$)

Notes: Active leader equals 1 in Leader condition; active (dishonest) leader equals 1 when the group has an active leader and this leader reported a 6 in Stage 1. All models report tobit regressions. Models 1 and 2 are panel specifications that use a worker's performance report in a period as an observation and include worker random effects, with bootstrapped standard errors clustered at the firm level. Model 3 uses the average report by a worker across all periods, with standard errors clustered at the firm level. Model 4 uses the average report within a firm across all periods, with robust standard errors. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

($\chi^2(1) = 5.26, p = 0.02$).³³ Thus, as with Experiment 1, we find evidence, though weaker, that unethical leaders yield groups that tend to act unethically.

We also again find evidence that leaders' actions influence followers. Tables B.9, B.10, and B.11, in the [Online Appendix](#), provide analogous results to the reduced-form analysis reported in Tables 4 and 5 and the structural estimates in Table 6. For brevity, Table 9 summarizes the principal coefficients measuring the effects of leaders' actions on followers' reported performance. For comparison, the table also presents

33. This comparison, between groups with no leaders and groups with dishonest leaders, is also statistically significant in model 2 ($\chi^2(1) = 5.04, p = 0.02$) and is marginally statistically significant in model 3 ($F_{(1,132)} = 3.20, p = 0.08$) and model 4 ($F_{(1,42)} = 2.36, p = 0.13$).

TABLE 9. Summary of how leaders' actions affect followers' reports (Experiment 2).

	Coefficient (Std. error)	Size relative to Exp. 1 coefficient (%)
<i>Incentive use</i>		
Bonus/perf. shares corr. up to previous period (ρ_{ft-1}) compare to model 1 in Table 4	0.920*** (0.292)	66.0
γ -Weight on leader incentive use for Type 1 (63.6%) compare to Type 1 (74.0%) in Table 6	0.012 (0.015)	16.9
γ -Weight on leader incentive use for Type 2 (36.4%) compare to Type 2 (26.0%) in Table 6	0.007 (0.011)	–
<i>Statement use</i>		
Leader's encouragement of high performance reports (ω_{ft}) compare to model 1 in Table 5	1.270 (0.976)	60.0
β -Weight on leader statements for Type 1 (63.6%) compare to Type 1 (74.0%) in Table 6	0.429*** (0.027)	83.3
β -Weight on leader statements for Type 2 (36.4%) compare to Type 2 (26.0%) in Table 6	0.791*** (0.055)	81.4

Notes: Primary coefficients of interest from Tables B.9, B.10, B.11 are presented; these tables present analogous estimates for Experiment 2 to those in Tables 4–6 for Experiment 1. Standard errors in parentheses. The second column presents the coefficient from Experiment 2 as a percentage of the coefficient from Experiment 1; “–” is reported for the case in which the Experiment 1 variable was estimated as 0.000 (both coefficients are statistically indistinguishable from zero). * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

these coefficients as a percentage of the corresponding coefficients for Experiment 1. Although the statistical significance of the relationships is weaker for Experiment 2, the coefficients generally have the same sign and are at least 60% of the size of the Experiment 1 coefficients. The only major difference appears to be in the structural estimates for γ , which measures workers' sensitivity to how the leader distributes incentives. Both Types 1 and 2 demonstrate insensitivity to how the leader distributes incentives in Experiment 2, whereas this was only the case for Type 2 in Experiment 1.³⁴

Finally, we study whether leaders we classify as dishonest are more likely to employ incentives and statements in a manner that encourages high performance reports. The results are qualitatively similar to those in Experiment 1: leaders we classify as dishonest tend to use incentives and statements in a manner that encourages more misreporting.³⁵ For example, when we re-estimate models 1 and 3 of Table 7,

34. [Online Appendix](#) Figures A.7A and B show scatterplots of the relationships between a leader's overall uses of incentives and statements to encourage dishonesty and the average change in workers' reports between Stage 1 and Stage 2 (analogous to [Figures 2\(a\)](#) and [\(b\)](#)). Both graphs reveal positive relationships; however, consistent with Table 9, they are weak for incentive use.

35. [Online Appendix](#) Figures A.8A and B show the average tendency to use incentives and statements in ways that encourage misreporting, separately for leaders we classify as dishonest and for those we do not. Table B.12 in the [Online Appendix](#) reports regression analysis similar to Table 7.

using Experiment 2 data, the analogous coefficients for “dishonest leader” are, respectively, 0.201 (132% of the analogous Experiment 1 coefficient) and 0.094 (63%) and both are at least marginally statistically significant ($p < 0.11$).

To summarize, our second experiment studies the effect of leadership on followers' ethical conduct in a different environment confronting firms than in our main experiment. In this study, there is no competition between firms and misreporting does not harm other firms. Nevertheless, we find similar effects of leadership, though smaller and statistically weaker, to those in our first experiment.

5. Conclusion

We use controlled laboratory experiments to study how leaders influence the unethical conduct of followers. As we note in the Introduction, the ability of researchers to address such questions using natural data is complicated by the hidden nature of unethical conduct, as well as by endogeneity issues that make causal inferences regarding the effects of leadership difficult.³⁶ The laboratory environment allows us to address many of these concerns, while employing a task that incorporates features of real-world unethical conduct.

Our main finding is that unethical leaders produce unethical behavior on the part of followers. Even though we only observe Stage 1 unethical conduct imperfectly, a noisy classification of leader “types” based on this behavior has explanatory power for how much workers misreport their performance when in groups with active leaders. Two features of our experiments make this finding especially important. First, leaders are appointed at random, which means that the effect of unethical leaders on workers is causal—the selection and endogeneity issues present in the field are eliminated in our laboratory setting. Second, followers are never informed of the leader's Stage 1 behavior. Thus, the effect of leaders must be through what actions they take in their functions as leaders—making statements to workers and distributing incentives. Indeed, our analysis reveals direct connections between these actions by leaders and followers' unethical conduct. We also show that leaders who are more likely to have misreported high performance in the first stage are more likely to employ strategies that encourage misreporting, particularly statements.

36. In fact, three of the authors of this paper attempted unsuccessfully to study a causal relationship between real leaders' ethical characteristics and followers' conduct. Specifically, we developed a simple task (juggling a ball with alternating feet) and gave amateur football players incentives similar to those in our first experiment: they could inflate their own score by misreporting it, thereby obtaining higher monetary earnings for their own team at the expense of other teams. We attempted to study the causal influence of leaders through priming—half the players were read task instructions by their coach and the other half by an experimenter—and tested whether priming in this task allowed us to detect an effect of leaders rated as less ethical (using the Ethical Leadership scale). In a pilot with 14 teams/leaders, we found no effect, while also discovering very little evidence of score misreporting. There are many reasons for a null effect—from a weak priming manipulation, to an unreliable measure of leader ethical traits, to a task in which intrinsic motivations limit dishonest reporting. These highlight the challenges of attempting to identify the relationships we study here in the field.

Our experimental design also allows us to compare the relative importance of how leaders use statements and incentives to influence followers. The effect of leaders through statements appears at least as strong as the influence through incentives, confirming other findings that the effect of leaders' communication can be as strong in shaping worker behavior as reliance on incentives (Brandts and Cooper 2007; Brandts, Cooper, and Weber 2015; Antonakis et al. 2016). Importantly, in our main experiment leaders tend to employ more statements requesting higher "performance" and this produces greater dishonesty among followers. Although the ability to distribute incentives also has an effect on unethical conduct, leaders rely less on this mechanism, and its aggregate effect is therefore weaker. Of course, our study includes only one incentive mechanism, and it is possible that alternative, more powerful incentives might produce stronger overall effects on dishonesty.³⁷ It is also the case that our study does not identify precisely why statements affect follower conduct. Moreover, the effectiveness of these two mechanisms may also depend on the environment confronting a group—the effects of leaders on ethical conduct appear stronger when firms are engaged in competition than when they are not. Hence, although our experiments produce evidence that leaders, the channels of influence at their disposal, and the environment confronting firms can jointly influence follower conduct, future work is necessary to fully understand all of these relationships. Put differently, although this paper shows that leaders *can* affect ethical conduct, more work is necessary to understand precisely *when* and *why* they have such effects.³⁸

Our results have potentially important policy implications for the design of organizations and for the selection of leaders and the definition of their responsibilities. Most importantly, we document a causal relationship between unethical leaders and the unethical behavior of those they lead. This suggests that the moral qualifications of a corporate leader can be important, beyond his or her managerial or technical abilities. Moreover, leaders' statements are a central channel through which such influence occurs and can be as strong an influence as financial rewards. Thus, rather than focusing solely on incentives and monitoring as a way to foment more ethical conduct, the "tone at the top" may play an important role. Importantly, our structural analysis suggests that different workers may respond differently to different channels of leadership influence, indicating that workers' characteristics can be an important element for whether they can be led to act unethically. Our findings also underscore the importance of governance and institutional checks on leaders' authority and their internal communication, since leaders can spread unethical behavior within an organization at very little cost, if they wish.

Our study also indicates that the selection of "good" leaders and their use of strategies, including public statements, to encourage ethical conduct may positively effect followers' ethical behavior. Thus, both mechanisms present potential

37. We thank two anonymous reviewers for raising this point.

38. For example, Schweitzer et al. (2004), demonstrate that goal setting can influence unethical conduct. A natural channel for leadership influence may be through the setting of performance targets. Alternatively, a leader's statements may make such targets more salient, by focusing workers' attention on them.

opportunities for leaders and organizations interested in curbing unethical conduct. The trick appears, at least partly, to get leaders to use these techniques at their disposal—particularly, to employ public statements that discourage unethical conduct. Our results suggest that a key approach may be emphasizing the search for leaders who act ethically themselves and are likely to employ these instruments to encourage ethical employee conduct.

References

- Antonakis, John, Giovanna d'Adda, Roberto A. Weber, and Christian Zehnder (2016). "On the Economic Value of Charismatic Leadership." Working Paper, HEC Lausanne, University of Lausanne.
- Antonakis, John, Marika Fenley, and Sue Liechti (2012). "Learning Charisma: Transform Yourself into the Person Others Want to Follow." *Harvard Business Review*, 90, 127–147.
- Baker, George, Robert Gibbons, and Kevin J. Murphy (1994). "Subjective Performance Measures in Optimal Incentive Contracts." *Quarterly Journal of Economics*, 109, 1125–1156.
- Barnard, Chester I. (1938). *The Functions of the Executive*. Harvard University Press, Cambridge, MA.
- Bass, Bernard M. (1990). *Bass and Stogdill's Handbook of Leadership: Theory, Research, and Managerial Applications*, 3rd ed. Free Press, New York.
- Beekman, Gonne, Erwin Bulte, and Eleonora Nillesen (2013). "Corruption and Economic Activity: Micro Level Evidence from Rural Liberia." *European Journal of Political Economy*, 30, 70–79.
- Beresford, Dennis R., Nicholas B. Katzenbach, and C.B.J. Rogers (2003). *Report of Investigation by the Special Investigative Committee of the Board of Directors of WorldCom, Inc.* Worldcom Incorporated, Clinton, MS.
- Bialik, Carl (2015). "How FIFA's Structure Lends Itself to Corruption." *Fivethirtyeight.com*, May 27. Available at: <http://fivethirtyeight.com/features/how-fifas-structure-lends-itself-to-corruption/> (accessed 10/06/2016).
- Brandts, Jordi and David Cooper (2007). "It's What You Say Not What You Pay: An Experimental Study of Manager-Employee Relationships in Overcoming Coordination Failure." *Journal of the European Economic Association*, 5, 1223–1268.
- Brandts, Jordi, David Cooper, and Roberto A. Weber (2015). "Legitimacy, Communication, and Leadership in the Turnaround Game." *Management Science*, 61, 2627–2645.
- Brown, Ken and Ianthe J. Dugan (2002). "Arthur Andersen's Fall From Grace Is a Sad Tale of Greed and Miscues." *Wall Street Journal*. Available at: <http://www.wsj.com/articles/SB1023409436545200> (accessed 31/01/2017).
- Brown, Michael E. and Linda K. Treviño (2006). "Ethical Leadership: A Review and Future directions." *The Leadership Quarterly* 17, 595–616.
- Brown, Michael E., Linda K. Treviño, and David A. Harrison (2005). "Ethical Leadership: A Social Learning Perspective for Construct Development and Testing." *Organizational Behavior and Human Decision Processes*, 97, 117–134.
- Burks, Stephen V. and Erin Krupka (2012). "A Multimethod Approach to Identifying Norms and Normative Expectations Within a Corporate Hierarchy: Evidence from the Financial Services Industry." *Management Science*, 58, 203–217.
- Burns, James M. (1978). *Leadership*. Harper & Row, New York.
- Chattopadhyay, Raghavendra and Esther Duflo (2004). "Women as Policy Makers: Evidence from a Randomized Policy Experiment in India." *Econometrica*, 72, 1409–1443.
- Conrads, Julian, Bernd Irlenbusch, Rainer M. Rilke, Anne Schielke, and Gari Walkowitz (2014). "Honesty in Tournaments." *Economic Letters*, 123, 90–93.
- Conrads, Julian, Bernd Irlenbusch, Rainer M. Rilke, and Gari Walkowitz (2013). "Lying and Team Incentives." *Journal of Economic Psychology*, 34, 1–7.

- Cremer, Andreas and Tom Bergin (2015). "Fear and Respect: VW's Culture under Winterkorn." Reuters.com, October 10. Available at: <http://www.reuters.com/article/us-volkswagen-emissions-culture-idUSKCN0S40MT20151010> (accessed 10/06/2016).
- Detert, James R., Linda K. Treviño, Ethan R. Burris, and Meena Andiappan (2007). "Managerial Modes of Influence and Counterproductivity in Organizations: A Longitudinal Business-Unit-Level Investigation." *Journal of Applied Psychology*, 92, 993–1005.
- Fischbacher, Urs (2007). "z-Tree: Zurich Toolbox for Ready-Made Economic Experiments." *Experimental Economics*, 10, 171–178.
- Fischbacher, Urs and Franziska Föllmi-Heusi (2013). "Lies in Disguise—An Experimental Study on Cheating." *Journal of the European Economic Association*, 11, 525–547.
- Gibson, Ranja, Carmen Tanner, and Alexander F. Wagner (2013). "Preferences for Truthfulness: Heterogeneity Among and Within Individuals." *American Economic Review*, 103(1), 532–548.
- Gino, Francesca, Shahar Ayal, and Dan Ariely (2009). "Contagion and Differentiation in Unethical Behavior: The Effect of One Bad Apple on the Barrel." *Psychological Science*, 20, 393–398.
- Gino, Francesca, Erin Krupka, and Roberto A. Weber (2013). "License to Cheat: Voluntary Regulation and Ethical Behavior." *Management Science*, 59, 2187–2203.
- Gneezy, Uri (2005). "Deception: The Role of Consequences." *American Economic Review*, 95(1), 384–394.
- Greiner, Ben (2015). "Subject Pool Recruitment Procedures: Organizing Experiments with ORSEE." *Journal of the Economic Science Association*, 1, 1–12.
- Güth, Werner, M. Vittoria Levati, Matthias Sutter, and Eline Van der Heijden (2007). "Leading by Example with and Without Exclusion Power in Voluntary Contribution Experiments." *Journal of Public Economics*, 91, 1023–1042.
- Hamman, John R., Roberto A. Weber, and Jonathan Woon (2011). "An Experimental Investigation of Electoral Delegation and the Provision of Public Goods." *American Journal of Political Science*, 55, 738–752.
- Hermalin, Benjamin E. (1998). "Toward an Economic Theory of Leadership: Leading by Example." *American Economic Review*, 88(5), 1188–1206.
- Houser, Daniel, Stefan Vetter, and Joachim Winter (2012). "Fairness and Cheating." *European Economic Review*, 56, 1645–1655.
- Jack, B. Kesley and Maria P. Recalde (2015). "Leadership and the Voluntary Provision of Public Goods: Field evidence from Bolivia." *Journal of Public Economics*, 122, 80–93.
- Jones, Benjamin F. and Benjamin A. Olken (2005). "Do Leaders Matter? National Leadership and Growth Since World War II." *The Quarterly Journal of Economics*, 120, 835–864.
- Kant, Immanuel (1785/1964). *Groundwork of the Metaphysics of Morals* (trans. by H. J. Paton). Harper & Row, New York (Original work published 1785).
- Kosfeld, Michael and Devesh Rustagi (2015). "Leader Punishment and Cooperation in Groups: Experimental Field Evidence from Commons Management in Ethiopia." *American Economic Review*, 105(2), 747–783.
- Kotter, John P. (1990). *Force for Change: How Leadership Differs from Management*. The Free Press, New York.
- Krupka, Erin L. and Roberto A. Weber (2013). "Identifying Social Norms Using Coordination Games: Why Does Dictator Game Sharing Vary?" *Journal of the European Economic Association*, 11, 495–524.
- Lazear, Edward P. (2012). "Leadership: A Personnel Economics Approach." *Labour Economics*, 19, 92–101.
- Mayer, David M., Karl Aquino, Rebecca L. Greenbaum, and Maribeth Kuenzi (2012). "Who Displays Ethical Leadership, and Why Does It Matter? An Examination of Antecedents and Consequences of Ethical Leadership." *Academy of Management Journal*, 55, 151–171.
- Mayer, David M., Samir Nurmohamed, Linda K. Treviño, Debra L. Shapiro, and Marshall Schminke (2013). "Encouraging Employees to Report Unethical Conduct Internally: It Takes a Village." *Organizational Behaviour and Human Decision Processes*, 121, 89–103.
- Mazar, Nina, On Amir, and Dan Ariely (2008). "The Dishonesty of Honest People: A Theory of Self-Concept Maintenance." *Journal of Marketing Research*, 45, 633–644.

- Milgram, Stanley (1963). "Behavioral Study of Obedience." *Journal of Abnormal and Social Psychology*, 67, 371–378.
- Mill, John S. (1863). *Utilitarianism*. Parker, Son and Bourn, London.
- Nagin, Daniel, James Rebitzer, Seth Sanders, and Lowell Taylor (2002). "Monitoring, Motivation and Management: The Determinants of Opportunistic Behavior in a Field Experiment." *American Economic Review*, 92(4), 850–873.
- Potters, Jan, Martin Sefton, and Lise Vesterlund (2007). "Leading-by-Example and Signaling in Voluntary Contribution Games: An Experimental Study." *Economic Theory*, 33, 169–182.
- Prendergast, Canice (1999). "The Provision of Incentives in Firms." *Journal of Economic Literature*, 37, 7–63.
- Schweitzer, Maurice E., Lisa Ordóñez, and Bambi Douma (2004). "Goal Setting as a Motivator of Unethical Behavior." *Academy of Management Journal* 47, 422–432.
- Toffler, Barbara L. and Jennifer Reingold (2004). *Final Accounting: Ambition, Greed and the Fall of Arthur Anderson*. Currency/Doubleday.

Supplementary Data

Supplementary data are available at [JEEA](#) online.