Social norms on unethical behaviors in the workplace: a lab experiment

Forthcoming in *International Review of Economics*

Social norms on unethical behaviors in the workplace: A lab experiment

Alice Guerra[a,*], Enya Turrini[b]

[a] Department of Economics, University of Bologna, piazza Scaravilli 2, 40126 Bologna, Italy. ORCID iD: 0000-0003-1956-0270.

[b] Department of Economics, Royal Holloway University of London, Horton Building, Egham TW20 0EQ United Kingdom. ORCID iD: 0000-0002-9686-5636.

[*] Corresponding author: University of Bologna, Department of Economics, piazza Scaravilli 2, 40126 Bologna, Italy. Phone: +39 0541 434 266. Email: alice.guerra3@unibo.it

Email addresses: alice.guerra3@unibo.it (A. Guerra); enya.turrini@rhul.ac.uk (E. Turrini)

Funding and/or Conflicts of interests/Competing interests: The authors have no funding and no competing interests to declare.

Data Availability. The data and code used in this research are available on request.

Acknowledgments: We thank Mario Daniele Amore, Giuseppe Attanasi, Joel B. Carnevale, Marco Fabbri, Orsola Garofalo, Natalia Montinari, Lisa Spantig for valuable suggestions, and Tom Lane and Daniele Nosenzo for providing useful material. For insightful comments, we thank the participants of the 2023 CNEE in Copenhagen, 2023 EALE in Berlin, 2023 ESA Europe in Exeter, 2023 ESA World in Lyon, the 2023 IAREP-SABE in Nice, the 2023 IMEBESS in Lisbon, the 2023 SEET conference in Valencia, and the EcHo seminar at the Royal Holloway University of London. We acknowledge proofreading and copyediting of the manuscript by Richard Forsythe (freelance proofreader), and financial support from the Department of Economics at the University of Bologna.

ABSTRACT

We analyze social norms on unethical behaviors in the workplace using a laboratory experiment. We conducted a norm-elicitation experiment in which we considered two unethical actions as observed in an earlier behavior experiment by Amore et al. (2023) [*J Bus Ethics*]: leaders' and workers' untruthful reporting, and workers' misalignment with their leader's truthful reporting. We presented participants with Amore et al.'s background: in experimental firms (1 leader and 3 workers), each member can report their performance via automatic or self-reporting, where the latter allows for profitable and undetectable earnings manipulation. Using the Krupka-Weber procedure, we asked participants to assess the social appropriateness of the reporting decisions that the subjects in Amore et al. (2023) could have taken. We find prevailing norms against self-reporting for artificial profit inflation, and workers' self-reporting when the leader used automatic reporting. Yet, despite these norms, many subjects in the previous experiment engaged in such unethical misreporting for personal gain. These findings reveal a disconnection between the prevailing social norms and the observed unethical behaviors.

Keywords: Ethical norms, Business ethics, Misreporting, Social norms, Norm-behavior consistency, Gender stereotypes, Krupka-Weber method

INTRODUCTION

Do individuals follow social norms? Long-standing empirical and experimental evidence would suggest a qualified yes: people follow norms, even when their violation is not observed, unlikely to be sanctioned, and profitable for the decision-maker (e.g., Bicchieri, 2005; Krupka & Weber, 2009, 2013; Banerjee, 2016; Barr et al., 2018; Lynn, 2018; d'Adda et al., 2020; Govindan, 2022; te Velde & Louis, 2022; Krupka et al., 2022; Bartling & Özdemir, 2023; Eckel et al., 2023).[1]

However, the existence of a social norm in itself does not automatically imply that individuals will abide by it. Recent empirical evidence has revealed that in some settings norms have little or no influence at all on individual behavior, leading to problematic norm-behavior inconsistencies (Morris et al., 2015; Gächter et al., 2017; Grimm, 2019; Traub et al., 2023).[2]

This is especially true when it comes to illegal or morally questionable decisions, e.g., earnings manipulations, tax evasion, and bribery, to name a few (Kocher et al., 2018; Fochmann et al., 2021; Aycinena et al., 2022; Feess et al., 2023). For example, Huber and Huber (2020) reveal that observed variations in dishonest financial reports can only be partly attributed to differences in social norms; and Guerra and Zhuravleva (2021) show that the predominant social norm against bribery does not align with observed bribing behavior.

---

[1] One of the most prominent explanations for norm-abiding behavior lies in the internalization process of the respective norm, which can occur through different channels including socialization, concerns about social and self-image, or observing others' behaviors (for reviews: Legros & Cislaghi, 2020; Gross & Vostroknutov, 2022).

[2] Some other contributions support the evidence that social norms do not always predict average behavior (e.g., Bicchieri & Xiao, 2009; Carpenter & Matthews, 2009; Bicchieri et al., 2011; Kocher et al., 2017; Kassas & Palma, 2019; Kölle & Quercia, 2021).

Norm-behavior inconsistency is problematic because it signals a failed internalization of the relevant social norm (probably weak, not salient, or not efficiently enforced) that likely boosts unethical behaviors, with negative externalities on the whole society. Scholars are still trying to ascertain the contexts and conditions under which norm-behavior discrepancies can be expected (Legros & Cislaghi, 2020).

In this paper, we contribute to this research effort by analyzing social norms on (un)ethical reporting in the workplace using a laboratory experiment, and investigating whether those norms are correlated with (un)ethical reporting decisions available to other subjects in an earlier behavior experiment (Amore et al., 2023). We believe it is crucial to explore this situational context for three main reasons. First, the choice of specific reporting methods is crucial to unravel the occurrence of fraudulent behaviors, because the probability of detecting fraud—or cheating more broadly—is endogenous to the method used to report a given outcome (Kleven et al., 2011; Rosaz & Villeval, 2012; Gill et al., 2013; Behnk et al., 2019; Feltovich, 2019; Santoro, 2021; Vranka et al., 2021; Amore et al., 2023; Lang & Schudy, 2023; Cagala et al., 2024). Second, as highlighted by Huber and Huber (2020), dishonesty in the financial industry often generates greater economic consequences—e.g., corporate fraud, insider trading, etc.—than cheating in more "innocuous" settings. Third, despite fraudulent behaviors being a recurring and prominent issue in numerous business contexts (consider, e.g., the major accounting scandals at Enron, WorldCom, Dieselgate; Burks & Krupka, 2012; Amiram et al., 2018; Dorfleitner et al., 2022), relatively little is known about the mechanisms driving such behaviors. An important question that has remained unanswered thus far is: *Can social norms about what constitutes a (un)ethical behavior in a given business situation, provide predictions as to how people behave in such a scenario?*

To address this question, we conducted a norm-elicitation experiment à la Krupka-Weber (2013) in which we considered two unethical actions available to other subjects in

Amore et al. (2023): (i) leaders' and workers' untruthful reporting, and (ii) workers' misalignment with their leader's truthful reporting. More specifically, we ran a virtual lab experiment meant to simulate a workplace, in which we presented participants with Amore et al.'s (2023) experimental background. In "experimental firms"—i.e., four-member groups with one leader and three workers—the final profit is given by the sum of each member's reported performance and split equally among members. Each member can report their performance via automatic or self-reporting, where the latter allows for profitable and undetectable earnings manipulation.[3] This is similar to Pate (2018) and Feltovich (2019), where subjects can choose between two reporting methods: a computer draw or a self-reported die roll. As in our setting, in Feltovich (2019) and Amore et al. (2023) self-reporting captures unethical behavior (see the concluding section for a discussion).

As in Amore et al. (2023), we experimentally varied the leader's ability to choose the reporting method, namely voluntarily chosen vs exogenously assigned. As an additional analysis, we further test for prejudicial discrimination against female leaders (Koburtay et al., 2019), by exogenously manipulating the leader's gender. Specifically, we implemented a 3 ("Reporting Method") × 2 ("Leader's Gender") design, within subjects for the "Reporting Method" scenarios, and between subjects for the "Leader's Gender" treatments. Regarding the "Reporting Method" scenarios, we varied the leader's ability to choose the reporting method: "Voluntarily" chosen, i.e., s/he can freely choose the reporting method, vs "Mandatorily" assigned, i.e., s/he is exogenously assigned to a specific method. Regarding the "Leader's Gender" treatments, we varied the leader's gender: in the "Gender Neutral" (GN) treatment the

[3] For comprehensive discussion on the relevance of reporting methods for truth-telling, see, e.g., Huber and Huber (2020); Vranka et al. (2021); Amore et al. (2023). More broadly on preferences for truth-telling, see, e.g., Gibson et al. (2013); Abeler et al. (2019).

leader is addressed with words that are valid for both genders (e.g., "he/she," "his/her"), whereas in the "Female Leader" (FL) treatment, the leader is uniquely addressed with feminine words (e.g., "she," "her"). Each subject was randomly assigned to either the GN or FL treatment (between-subject) and went through all of the "Reporting Method" scenarios (within-subject).

We elicited social norms using the Krupka-Weber procedure, asking participants to assess the social appropriateness of each reporting decision available to firm members under different scenarios.[4] To analyze norm-behavior consistency, we compared the results from our norm-elicitation experiment with a set of behavioral patterns observed in Amore et al. (2023). The latter revealed a strong prevalence of untruthful reporting decisions through self-reporting (which was chosen by approximately 80% of subjects, especially among those assigned to the role of leaders). Nonetheless, observing a leader's truthful ("automatic") reporting choice made workers statistically more likely to choose the truthful reporting method as well, albeit only when the leader made a voluntary choice. Their experimental data further indicated that workers' misalignment was only punished by the leader when the latter chose self-reporting and workers chose automatic reporting ("positive" misalignment).

Here, we seek to investigate whether social norms are aligned with those observed behaviors, by answering three main questions: (i) is there a social norm against unethical behaviors, (ii) are social norms affected by reporting choices, and (iii) are social norms affected by the leader's gender. By relying on prior theoretical and empirical evidence on (i) norms

[4] The Krupka-Weber technique has been used to analyze social norms on a wide set of phenomena, including ethical conduct of financial advisers, discrimination, and gendered occupational choices, and it has been proved to be largely robust to the presence of alternative focal points (Nosenzo & Görges, 2020; Fallucchi & Nosenzo, 2022).

against dishonesty (e.g., Xu & Schriesheim, 2018), and (ii) societal prejudices against female leadership (e.g., Eagly & Karau, 2002; Gangadharan et al., 2016, 2019; De Paola et al., 2022), we could ex-ante predict (i) the presence of social norms against dishonesty (here, untruthful reporting via self-reporting); and that (ii) those norms statistically vary with the leader's gender, with unethical actions being rated as more socially inappropriate in the presence of female leaders.

However, we refrain from making precise predictions for the following reasons. Evidence on social norms in the workplace—and truthful reporting in general—is too scant and sparse to make clear predictions (Goldstein & Cialdini, 2011; d'Adda et al., 2017; Rilke et al., 2021; Dorfleitner et al., 2022; Amore et al., 2023). Furthermore, the way in which prior studies have conceptualized and operationalized social norms considerably varies both within and across disciplines, e.g., economics vs social psychology (Cialdini & Trost, 1998; Nosenzo & Görges, 2020). Perhaps more importantly, evidence on norm-abiding behavior is mixed and controversial, especially when related to unethical or illegal decisions (e.g., Huber & Huber, 2020; Guerra & Zhuravleva, 2021; Aycinena et al., 2022).

Regarding gender-specific norms, prior theoretical works have pointed to the presence of societal expectations towards female leaders—who should display stronger ethical concern than their male counterparts—and incongruities between the female gender role in society and the leadership role in businesses (Eagly & Karau, 2002; Kennedy & Kray, 2014). Nonetheless, some empirical evidence has rather revealed that exposure to a female leader does not statistically affect individuals' expectations or behaviors (Ferreira & Gyourko, 2014; Lane, 2016; Zhang et al., 2024).

Taken all together, those discrepancies between theory and evidence favor an inductive over a deductive approach, i.e., directly exploring the patterns in the data without fixing any prior predictions about what they should be.[5]

Our results show a prevailing social norm against self-reporting, albeit only if used to artificially inflate earnings, i.e., deliberate misreporting. A clear norm emerges against workers' "negative" misalignment, i.e., choosing self-reporting when the leader used automatic reporting (voluntarily or mandatorily), even if not intended for earnings inflation. Additionally, social norms do not vary with exposure to female leaders, nor with the leader's exogenous vs endogenous reporting method.

Overall, these findings reveal that social norms (i) do not align with but are rather disconnected from the predominant unethical behaviors of the participants in Amore et al. (2023), and (ii) do not vary with exposure to female leaders, as theories on societal prejudices against female leaders would rather suggest.

Our study contributes to various areas of literature. To our knowledge, this is the first experimental research to both identify social norms on unethical decision-making in the workplace, and analyze their correlation with observed behaviors. This expands our current understanding of the contexts under which norm-behavior inconsistencies should be expected (Batzke & Ernst, 2022; Bicchieri et al., 2022, 2023; Gross & Vostroknutov, 2022), and contributes to the growing experimental literature on leadership (Drouvelis, 2021, Chapter 6).

Furthermore, our research enhances scholarship on gender-based ethical norms. Established theoretical frameworks have underscored societal expectations for female leaders to exhibit heightened ethical concern compared to their male counterparts, alongside the incongruities between the female gender role and leadership positions (Heilman, 2012; Grosser et al., 2014; Pullen & Vachhani, 2021; Stajkovic & Stajkovic, 2024). Our study challenges

[5] This study was not pre-registered, and so it must be considered exploratory.

these notions and stereotypical expectations. Additionally, our research demonstrates that social norms against unethical behaviors are evaluated similarly by both men and women, offering no support for the idea of differential responsiveness of males and females to social norms, specifically in the context of business ethics (Hollway, 2007).

It is noteworthy that also d'Adda et al. (2017) and Amore et al. (2023) analyzed social norms by conducting the Krupka-Weber task immediately after the main experiment, with conflicting results. Amore et al. (2023) investigated whether leaders influence workers' view of how appropriate is to choose a reporting system not aligned with the leader's one, and to inflate the die-rolling outcome. They found no significant differences in workers' mean responses across treatments, suggesting that leaders can influence workers' behavior through voluntary decisions and punishment power, but do not shape their norms. Conversely, d'Adda et al. (2017) showed that leaders influence not only workers' behaviors but also social norms. Indeed, when leaders encouraged their workers to inflate earnings for firm profits, those workers also exhibited a weaker social norm against unethical behaviors.

However, these studies present some limitations. In Amore et al. (2023), the null result from the norm-elicitation task might be due to the limited number of observations, and hence might be misleading, as acknowledged by the authors themselves who call for "more research about the effect of leadership on social norms within organizations. (Amore et al., 2023, p. 507)". Additionally, in both studies, social norms are elicited immediately after the main experiment, by using the same subject pool for both the analyses of actual behavior and perceived social norms. This can be problematic, because participants' assessment of social norms might have been affected by their prior behavior during the main experiment, e.g., by a desire to justify their earlier actions or maintain a consistent self-image. Our research overcomes these limitations by collecting a greater number of observations (1,804 vs 240 in Amore et al., 2023), and by using a distinct subject pool for the specific analysis of social

norms. This ensures that the social norms data we collect are indicative of genuine perceptions rather than post hoc rationalizations.

The paper is structured as follows. The next section outlines the experimental design, procedures, and subject pool. Following that, we analyze the results, which are discussed in the concluding section along with potential avenues for future research. The supplementary material contains additional tables (Appendix A) and the experimental instructions, along with related supporting information and screenshots (Appendix B).

THE EXPERIMENT

In the following, we describe the experimental design, procedures, and subject pool. An English translation of the instructions as provided to the participants is available in Appendix B of the supplementary material.[6]

Design

We designed a norm-elicitation experiment using the Krupka-Weber task, with a 3 ("Reporting Method") × 2 ("Leader's Gender") design. We implemented a within-subject design for the "Reporting Method" scenarios, and a between-subject design for the "Leader's Gender" treatments. Regarding the "Reporting Method" scenarios, we varied the leader's ability to choose the reporting method: "Voluntarily" chosen, i.e., s/he can freely choose the

[6] In the sake of comparability, in our experimental instructions we used the same framing and terms as in Amore et al. (2023). This includes the terms "leader" and "worker," which one might argue to be potentially problematic as may generate an expectation for the workers to follow the leader. Though, this expectation is not supported by empirical evidence, as revealed by Amore et al. (2023).

reporting method, vs "Mandatorily" assigned, i.e., s/he is exogenously assigned to a specific method (more details below).

Regarding the "Leader's Gender" treatments, we varied the leader's gender: in the "Gender Neutral" (GN) treatment the leader is addressed with words—including pronouns, adjectives, verbs, etc.—that are valid for both genders (e.g., "he/she," "his/her"), whereas in the "Female Leader" (FL) treatment, the leader is uniquely addressed with feminine words (e.g., "she," "her").[7]

Each subject was randomly assigned to either the GN or FL treatment (between-subject design) and went through all of the "Reporting Method" scenarios (within-subject design). The within-subject design enables us to control for significant sources of variability and mitigate potential issues associated with demand effects (Zizzo, 2010; Eckel et al., 2023).

[7] We shall recall that the experiment was conducted in Italy with instructions in Italian language. Differently from English, Italian does have grammatical gender—and this allowed us to clearly distinguish gender vs female treatments.

Figure 1

Overview of the experimental design

Part 1: Background description (Amore et al., 2023)

- Four-member experimental firms (one leader, three workers).
- The final profit is given by the sum of each member's reported performance and equally split among members.
- Each member can choose to report their performance via automatic or self-reporting, where the latter allows for profitable and undetectable earnings manipulation.
- The leader is the first mover, while the three workers act simultaneously as second movers after having observed their leader's reporting method.
- Different scenarios based on (i) the leaders' ability to choose the reporting method (exogenously assigned vs voluntarily chosen), and (ii) the leaders' gender (neutral vs female). See Part 2.

Part 2: Social norm-elicitation task

- Subjects are asked to evaluate the social appropriateness of eleven reporting choices (listed in Table 1), which firm members can take under the scenarios described below.
- 3 ("Reporting Method") × 2 ("Leader's Gender") experimental design
  - Within-subject design for the "Reporting Method" scenarios
  - Between-subject design for the "Leader's Gender" treatments

Leader's Gender	Reporting Method		
	Voluntary Reporting	Mandatory self-reporting (mSR)	Mandatory automatic reporting (mAR)
Gender Neutral (GN)	The leader can freely choose between self-reporting and automatic reporting	The leader is exogenously assigned to self-reporting	The leader is exogenously assigned to automatic reporting
Female Leader (FL)	The female leader can freely choose between self-reporting and automatic reporting	The female leader is exogenously assigned to self-reporting	The female leader is exogenously assigned to automatic reporting

More specifically, the experiment comprises two main parts (Figure 1). In the first part, we described Amore et al.'s (2023) experimental background with three scenarios, corresponding to the three "Reporting Method" treatments. More specifically, participants were presented with "experimental firms"—i.e., four-member groups with a leader and three workers—whose final profit is the sum of the performance reported by the group members and is equally split among them. Individual performance is measured as the outcome of the rolling

of a fair six-sided die. Each member can report their performance via two methods: "self-reporting" (SR) or "automatic reporting" (AR).

Self-reporting means that a firm member rolls autonomously and privately a *physical* six-faced die, and inputs the result themselves. In this case, no one—neither the firm members nor third parties—would ever be able to verify the actual die-rolling outcome. Automatic reporting means that a firm member rolls a *virtual* six-faced die (via the computer), and the outcome is automatically reported by the computer, without any possibility to manipulate the result obtained. While self-reporting makes cheating possible, profitable (since firm and individual profits increase with the reported die-rolling outcomes), and undetectable (since the actual die-rolling outcome is unobservable by any third party), automatic reporting represents the ethical choice as it prevents any earnings manipulation, hence signaling truthful reporting.[8]

While workers are always free to choose between the two reporting methods, the leader can also be exogenously assigned (mandatorily) to a specific method. The leader is the first mover, while the workers act simultaneously as second movers after being informed about the reporting method used (exogenously assigned or voluntarily chosen) by their leader. The final profit of the firm is computed as the sum of the die-rolling outcomes reported by firm members, and it is equally split among them.

The second part comprises the social norm-elicitation task. After having described Amore et al.'s (2023) experimental background, we asked our participants to evaluate eleven

[8] As specified in Amore et al. (2023, p. 498), "misreporting increases one's individual or group payoff without generating negative externalities on anyone else aside from the experimenter […]. Hence, each group would be better off if all members fraudulently overstate earnings by self-reporting the highest die-rolling outcome, adhering to a norm of cooperation (group payoff maximization) at the expense of a norm of honesty."

different reporting choices available to the participants in Amore et al.'s (2023) under the three different scenarios, as listed in Table 1. Subjects were instructed to rate each action according to its "social appropriateness" by selecting one option on a four-point scale: 'very socially inappropriate,' 'somewhat socially inappropriate,' 'somewhat socially appropriate' or 'very socially appropriate.' By applying a within-subject design, each participant was presented with all three "Reporting Method" scenarios, each with a set of actions to assess. To mitigate potential ordering effects (d'Adda et al., 2016), we randomly varied the order of the three scenarios.

Table 1
Decisions to be rated according to their social appropriateness

Scenario	Reporting choice	Description
Voluntary	"L SR"	The [*female* (in the FL treatment)] leader chooses self-reporting rather than automatic reporting.
	"L SR>1"	The [*female*] leader chooses self-reporting, obtains 1 from the die-rolling, but reports a value above 1.
	"L SR, W AR"	The worker sees the [*female*] leader choosing self-reporting but chooses automatic reporting.
	"L AR, W SR"	The worker sees the [*female*] leader choosing automatic reporting but chooses self-reporting.
	"L SR, W SR>1"	The worker sees the [*female*] leader choosing self-reporting, chooses self-reporting, obtains 1 from the die-rolling, but reports a value above 1.
	"L AR, W SR>1"	The worker sees the [*female*] leader choosing automatic reporting, chooses self-reporting, obtains 1 from the die-rolling, but reports a value above 1.
Mandatory self-reporting ("mSR")	"L mSR>1"	The [*female*] leader is exogenously assigned to self-reporting, obtains 1 from the die-rolling, but reports a value above 1.
	"L mSR, W AR"	The worker sees that the [*female*] leader is exogenously assigned to self-reporting but chooses automatic reporting.
	"L mSR, W SR>1"	The worker sees that the [*female*] leader is exogenously assigned to self-reporting, chooses self-reporting, obtains 1 from the die-rolling, but reports a value above 1.
Mandatory Automatic reporting ("mAR")	"L mAR, W SR"	The worker sees that the [*female*] leader is exogenously assigned to automatic reporting but chooses self-reporting.
	"L mAR, W SR>1"	The worker sees that the [*female*] leader is exogenously assigned to use automatic reporting, chooses self-reporting, obtains 1 from the die-rolling, but reports a value above 1.

Abbreviations: L stands for Leader, W for Worker, (m)SR for (mandatory) Self-Reporting, (m)AR for (mandatory) Automatic Reporting, GN for Gender Neutral treatment, FL for Female Leader treatment.

Subjects were instructed to indicate not what they *personally* think is the right thing to do, but rather what they expect *the majority of people* would agree on, and repeatedly reminded in the instructions that we were not asking "*about what you personally think should be done, but what you think the majority of people believe to be socially appropriate or inappropriate.*". They were informed that "socially appropriate" means "*the choice that the majority of people agree to be the "correct" or "ethical" thing to do*," and that they would receive a monetary reward if they rated the actions in the same way *as most other participants* in the task.

As in Krupka and Weber (2009), at the end of the experiment, one of the eleven actions was selected at random and each subject's evaluation of that action was compared with how most other participants had evaluated the same action. If his/her evaluation matched that of most other participants—i.e., *the modal response*—s/he was paid an additional 7 euros on top of their 2 euros fixed participation fee; otherwise, s/he was only paid the 2 euros participation fee.[9] This incentive structure—used by Krupka and Weber (2013) and many related contributions (Banerjee, 2016; Huber & Huber, 2020; Schmidt et al., 2020; Guerra & Zhuravleva, 2021; Fallucchi & Nosenzo, 2022; Huber et al., 2023; Rhodes et al., 2024; Charness et al., 2024; Kimbrough et al., 2024)—transforms the task into a pure coordination game where participants have to tacitly coordinate with others in terms of how they rate actions, i.e., to reveal what they perceive to be the socially recognized perceptions of the appropriateness of the described action, rather than their own personal perception of appropriateness. This is important because social norms are collectively recognized rules of

[9] See Table A5 in the supplementary material, where we report the percentage of subjects perfectly matching the modal response for a given action, along with the percentage of those highly mismatching it.

behavior, rather than personal opinions about behaviors (Bicchieri, 2005, 2010; Fallucchi and Nosenzo, 2022; Krupka et al., 2022).

To ensure that all subjects understood the instructions, a computer-based quiz with true/false comprehension questions was conducted before starting the norm-elicitation task, with direct feedback and explanations in case of an incorrect answer. After the norm-elicitation task, subjects were asked to complete a short questionnaire collecting basic demographic information, with no effects on their earnings, and this concluded the experiment.

Procedure

A total of four sessions were conducted at the end of March 2022, on two separate days. To avoid potential correlations between a specific moment of the day—or the day per se—and participants' behavior, we randomized the order of treatments across sessions, i.e., on one day the GN treatment was carried out before the FL treatment, while on the following day we implemented the opposite order. For each session, we recruited a different pool of subjects (between-subject design), while keeping the same experimenters and lab staff to avoid any possible confounding effect (Fréchette, 2012).

The participants were recruited from the subject pool database of the Bologna Laboratory for Experiments in Social Science (BLESS) of the University of Bologna in Italy, via ORSEE (Greiner, 2015). We applied two recruitment filters: (i) to avoid any cultural effect, we only targeted Italian citizens born in the North or Center of Italy;[10] and (ii) to achieve gender

[10] The Italian Regions to be included in such geographical distinction were taken following the definition provided by Istat (Italian National Institute of Statistics). The North includes Piedmont, Aosta Valley, Lombardy, Liguria, Trentino-South Tyrol, Friuli-Venezia Giulia, Emilia-Romagna, and Veneto. The Center includes Tuscany, Lazio, Umbria, and Marche.

balance across treatments, we pursued a comparable share of male and female participants in each session.

We set up an ad-hoc virtual lab: we programmed the experiment in Qualtrics and invited participants via email to a Zoom meeting where the instructions were read aloud by the experimenter.[11] Participants were recommended in advance to have a stable internet connection and avoid small screens (cell phones, tablets). Throughout the entire duration of the Zoom meetings, the instructor had the webcam turned on, with the surrounding environment kept identical in all sessions, and used the microphone and chat function to communicate with the subjects. In the interest of anonymity, subjects were instructed to always keep their webcams off, from the very beginning to the end of the Zoom meeting. They were allowed to ask questions by typing them in the Zoom chat. Importantly, to further ensure anonymity, participants were asked to replace their names in the Zoom meeting with an ID code, which we randomly assigned them and privately communicated via email through Qualtrics the day before the experimental session. Screenshots of the Zoom meeting and further details of the experimental procedure are provided in the supplementary material.

On average, each Zoom session lasted about 55 minutes (of which it took 25 minutes to complete the norm-elicitation task), including the one-by-one admission of the participants from the waiting room to the main Zoom meeting, reading the instructions, and closing the session with information about the payments via PayPal.

As in Krupka and Weber (2009), once all the participants had completed the tasks, one of the eleven questions was randomly selected and after having compared all the participants'

[11] This procedure has been refereed in the literature as "a lablike experiment" or "Lab on the Web" (Buso et al., 2020, 2021). For a similar procedure, see also Guerra and Zhuravleva (2021). See screen shots of the Zoom meetings in the supplementary material.

answers with the modal response, subjects were privately paid for correct matches (see Table A5 in the supplementary material for the percentage of correct matches across treatments). Here, we took particular care in ensuring transparency. Specifically, the experimenter randomly drew a number, and then asked the participant with the corresponding ID code to randomly choose a number between 1 and 11 and write it in the Zoom chat to all participants. The number written by the participant was used to select the question for the final payments. Participants were paid via PayPal at the end of all experimental sessions. The average payment per subject was 5 euros, with a total amount of 851 euros.

Subject Pool

A total of 164 subjects were recruited, of which 79 in the GN treatment and 85 in the FL treatment: 36 subjects in the first session (GN treatment), 45 and 40 subjects in the second and the third sessions, respectively (both FL treatments), and 43 participants in the fourth session (GN treatment). As is standard in lab economic experiments, the majority of subjects were university students (74.39% at either the undergraduate or graduate level), mostly from Economics. Pooling across sessions, 53.66% of the subjects were female, and the mean age was 25 years old. Table A1 in the supplementary material reports the sample descriptives—both pooled across treatments and for each treatment separately.

A series of balance tests—with p values computed according to Chiapello (2018) and reported in Table A1 in the supplementary material—confirmed that the sample is balanced across treatments in terms of personal characteristics (gender, birthplace, employment status, field of study), except for two factors, namely age and higher education. Specifically, the average age is higher under the "Gender Neutral" vs the "Female Leader" treatment (26 vs 24 years old). The mean value of higher education (i.e., having at least a Bachelor's Degree) is greater under the "Gender Neutral" vs the "Female Leader" treatment (0.759 vs 0.518). These differences are statistically significant at the 5% and 1% significance levels, respectively. In

our analyses, we have accounted for these unbalanced factors by controlling for them in the regression analyses and clustering standard errors at the individual level.

RESULTS

Following the literature—including the original Krupka-Weber approach and related contributions (e.g., Barr et al., 2018; Huber & Huber, 2020; Guerra & Zhuravleva, 2021; Fallucchi & Nosenzo, 2022)—we assign evenly-spaced numeric values to the four appropriateness ratings: −1 to 'very socially inappropriate,' −1/3 to 'somewhat socially inappropriate,' +1/3 to 'somewhat socially appropriate' and +1 to 'very socially appropriate.'

Table 2 displays the distribution of the evaluations for each reporting choice across the three scenarios, for each "Leader's Gender" treatment as well as pooled across treatments. For each scenario and treatment, Table 2 further reports the overall average ratings and their standard deviation. To ease the comparison, the mean ratings are also displayed in Figure 2, which plots the average appropriateness ratings assigned to each reporting choice in the GN treatment (hollowed square marker) and FL treatment (solid circle marker). The corresponding detailed summary statistics, including 95% confidence intervals, are reported in Table A4 in the supplementary material.

Table 2
Distribution of appropriateness ratings

(a) "Voluntary" Scenario: The leader can freely choose between self-reporting (SR) and automatic reporting (AR)

	Leader chooses SR instead of AR "L SR"			Leader chooses SR and misreports "L SR>1"			Leader chooses SR, Worker misaligns (AR) "L SR, W AR"		
	GN	FL	Pooled	GN	FL	Pooled	GN	FL	Pooled
– – (%)	7.59	5.88	6.71	**55.70**	**55.29**	**55.49**	5.06	4.71	4.88
– (%)	**43.04**	**47.06**	**45.12**	25.32	31.76	28.66	18.99	21.18	20.12
+ (%)	30.38	28.24	29.27	11.39	9.41	10.37	**39.24**	**45.88**	**42.68**
+ + (%)	18.99	18.82	18.90	7.59	3.53	5.49	36.71	28.24	32.32
Mean rating	*0.072*	*0.067*	*0.069*	*-0.527*	*-0.592*	*-0.560*	*0.384*	*0.318*	*0.350*
Std. Dev.	*0.589*	*0.575*	*0.580*	*0.633*	*0.535*	*0.584*	*0.582*	*0.554*	*0.570*

	Leader chooses AR, Worker misaligns (SR) "L AR, W SR"			Leader chooses SR, Worker aligns (SR) and misreports "L SR, W SR>1"			Leader chooses AR, Worker misaligns (SR) and misreports "L AR, W SR>1"		
	GN	FL	Pooled	GN	FL	Pooled	GN	FL	Pooled
– – (%)	15.19	12.94	14.02	25.32	23.53	24.39	**49.37**	**52.94**	**51.22**
– (%)	**43.04**	**47.06**	**45.12**	**41.77**	**55.29**	**48.78**	30.38	34.12	32.32
+ (%)	29.11	31.76	30.49	22.78	15.29	18.90	11.39	9.41	10.37
+ + (%)	12.66	8.24	10.37	10.13	5.88	7.93	8.86	3.53	6.10
Mean rating	*-0.072*	*-0.098*	*-0.085*	*-0.215*	*-0.310*	*-0.264*	*-0.468*	*-0.576*	*-0.524*
Std. Dev.	*0.598*	*0.541*	*0.570*	*0.620*	*0.529*	*0.575*	*0.644*	*0.533*	*0.590*

(b) "mSR" Scenario: The leader is exogenously assigned to self-reporting (SR)

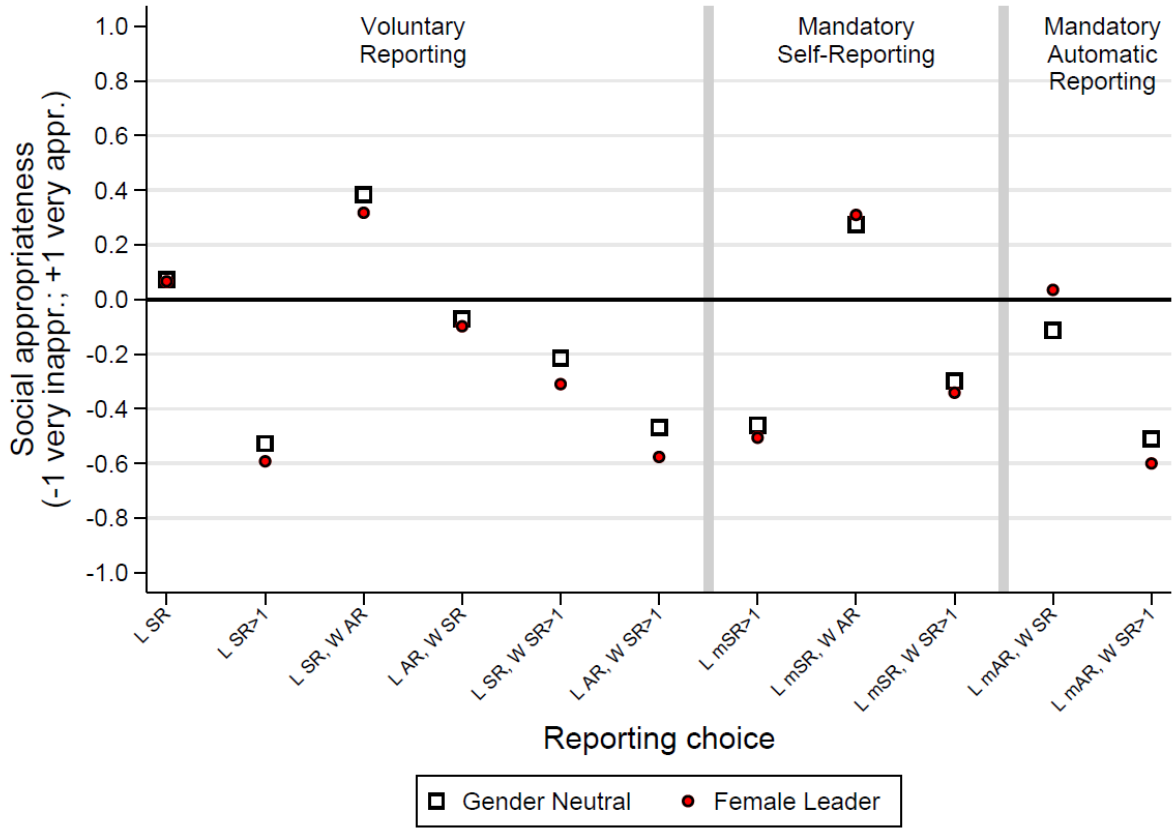
	Leader misreports "L mSR>1"			Worker misaligns (AR) "L mSR, W AR"			Worker aligns (SR) and misreports "L mSR, W SR>1"		
	GN	FL	Pooled	GN	FL	Pooled	GN	FL	Pooled
– – (%)	**45.57**	**44.71**	**45.12**	5.06	10.59	7.93	30.38	27.06	28.66
– (%)	34.18	40.00	37.20	26.58	16.47	21.34	**43.04**	**51.76**	**47.56**
+ (%)	13.92	11.76	12.80	**40.51**	**38.82**	**39.63**	17.72	16.47	17.07
+ + (%)	6.33	3.53	4.88	27.85	34.12	31.10	8.86	4.71	6.71
Mean rating	*-0.410*	*-0.506*	*-0.484*	*0.274*	*0.310*	*0.293*	*-0.300*	*-0.341*	*-0.321*
Std. Dev.	*0.605*	*0.536*	*0.569*	*0.577*	*0.646*	*0.612*	*0.612*	*0.529*	*0.569*

(c) "mAR" Scenario: The leader is exogenously assigned to automatic reporting (AR)

	Worker misaligns (SR) "L mAR, W SR"			Worker misaligns (SR) and misreports "L mAR, W SR>1"		
	GN	FL	Pooled	GN	FL	Pooled
– – (%)	15.19	7.06	10.98	**50.63**	**55.29**	**53.05**
– (%)	**48.10**	38.82	**43.29**	32.91	31.76	32.32
+ (%)	25.32	**45.88**	35.98	8.86	10.59	9.76
+ + (%)	11.39	8.24	9.76	7.59	2.35	4.88
Mean rating	*-0.114*	*0.035*	*-0.037*	*-0.511*	*-0.600*	*-0.557*
Std. Dev.	*0.582*	*0.499*	*0.544*	*0.611*	*0.516*	*0.564*

Notes: Number of participants: 164 (Pooled), of which 79 in the GN treatment and 85 in the FL treatment. Each cell shows the percentage of subjects selecting a specific appropriateness rating for a given action (reporting choice in each scenario). Responses are 'very socially inappropriate' (– –) = −1; 'somewhat socially inappropriate' (–) = −1/3; 'somewhat socially appropriate' (+) =+1/3; 'very socially appropriate' (+ +) = +1. The grey-shaded cells indicate the modal response for each given action. Each sub-table also displays the overall mean rating and standard deviation. The statistics are reported both pooled across the Leader's Gender treatments, and separately by treatment. Abbreviations: L stands for Leader; W for Worker; (m)SR for (Mandatory) Self-Reporting; (m)AR for (Mandatory) Automatic Reporting; SR>1 for 'obtaining 1 from die-rolling but declaring a value above 1'.

Figure 2
Mean appropriateness ratings by Leader's Gender treatment



Notes: This figure displays the mean social appropriateness ratings of each reporting choice in the two Leader's Gender treatments (hollowed square marker for GN; solid circle marker for FL). Mean ratings are taken by assigning values of −1, −1/3, +1/3, and +1 for the ratings 'very socially inappropriate,' 'somewhat socially inappropriate,' 'somewhat socially appropriate,' and 'very socially appropriate,' respectively, and averaging the values for all participants in a given treatment. Abbreviations: L stands for Leader; W for Worker; (m)SR for (Mandatory) Self-Reporting; (m)AR for (Mandatory) Automatic Reporting; SR>1 for 'Self-Reporting, obtaining 1 from the die-rolling, but reporting a value above 1.' For detailed summary statistics, including confidence intervals, see Table A4 in the supplementary material.

In the following, we first look at the social appropriateness ratings of reporting choices under each "Reporting Method" scenario and "Leader's Gender" treatment. Next, we statistically analyze treatment effects. Specifically, we compare social norms between the "Reporting Method" scenarios (Voluntary vs Mandatory Reporting) under each "Leader's Gender" treatment. Subsequently, we conduct an additional analysis of gender differences by comparing social norms between "Leader's Gender" treatments under each "Reporting Method" scenario.

Is There a Social Norm Against Unethical Behaviors?

We begin with a descriptive overview of the social appropriateness ratings of reporting choices. In each treatment and scenario, the mean and modal evaluations follow the same general pattern. As shown in Table 2, a prevailing social norm exists against self-reporting *and* misreporting (i.e., inflating earnings; see all "SR>1" choices). Looking at the leader's action, this norm becomes stronger when the leader *voluntarily* chooses—vis-à-vis is exogenously assigned to—self-reporting (and inflate earnings; see "L SR>1" vs "L mSR>1"). Along the same line, when looking at workers' actions, their decision of self-reporting *and* inflating earnings is perceived as less inappropriate when their leader also *voluntarily* chose self-reporting compared to the scenario in which the leader is exogenously assigned to self-reporting ("L SR, W SR>1" vs "L mSR, W SR>1").

Instead, choosing self-reporting *per se* (i.e., without earnings manipulation) rather than automatic reporting is not perceived as less inappropriate compared to choosing self-reporting *to manipulate earnings*. This holds from both the leader's ("L SR" vs "L SR>1") and worker's perspectives ("L AR, W SR" vs "L AR, W SR>1;" "L mAR, W SR" vs "L mAR, W SR>1").

Clear social norms also emerge in favor of "positive" misalignment, i.e., workers choosing automatic reporting when the leader used self-reporting ("L SR, W AR" and "L mSR, W AR"), and against "negative" misalignment, i.e., workers choosing self-reporting when the leader used automatic reporting ("L AR, W SR" and "L mAR, W SR"), especially when this comes jointly with earnings manipulation ("L AR, W SR>1" and "L mAR, W SR>1"), coherently with previous results.

Figure 2 also shows that, generally, reporting choices are rated similarly across the two "Leader's Gender" treatments, except for negative misalignment under the "Mandatory" scenario (choosing self-reporting when the leader was exogenously assigned to automatic

reporting; see "L mAR, W SR"), which is perceived as less inappropriate in the presence of a female leader.

Overall, these results can be summarized as follows:

Result 1: Prevailing social norms emerge against untruthful reporting (i.e., self-reporting, if used to inflate earnings), against workers' "negative" misalignment (i.e., choosing untruthful reporting when their leader used truthful reporting), but in favor of "positive" misalignment (i.e., workers choosing truthful reporting when the leader used untruthful reporting).

Are Social Norms Affected by Reporting Choices?

We statistically analyze whether social norms change with reporting choices by contrasting the leader's *voluntary* decisions against the *mandatory* assignments (under each "Leader's Gender" treatment). Specifically, we test whether the average and distribution of social appropriateness ratings are statistically different between the "Voluntary" scenario vis-à-vis the "Mandatory Self" or "Mandatory Automatic" scenarios, while keeping the "Leader's Gender" treatment constant.

For this purpose, following Barr et al. (2009) and Guerra and Zhuravleva (2021), we conducted a set of transformed Somers' D-tests (Newson, 2002, 2006), in which we accounted for the non-independence of observations within each session by clustering at the subject level. The estimates are assumed to have a *t* distribution and the confidence intervals are based on Fisher's Z transformation (Upton, 1992).

In the following, we list the tests and the corresponding results.

- *The leader misreports under "Voluntary" vs "Mandatory" self-reporting ("L SR>1" vs "L mSR>1")*. Under each "Leader's Gender" treatment, for both "L SR>1" and "L mSR>1" the modal response for a leader's misreporting is 'very socially inappropriate,' with the mean rating being statistically lower when the leader *voluntarily* chooses to misreport. The

Somers' D-test shows that the difference in the average social appropriateness ratings between the two reporting choices is statistically significant (p=0.009 in the pooled sample; p=0.075 in GN; p=0.059 in FL).

- *The worker misaligns choosing automatic reporting when the leader voluntarily chooses vis-à-vis being exogenously assigned to self-reporting ("L SR, W AR" vs "L mSR, W AR").* The decision of the worker to misalign with the leader's self-reporting by choosing automatic reporting instead is rated as 'somewhat socially appropriate,' regardless of whether the leader voluntarily chose or was exogenously assigned to the self-reporting system (Somers' D-test: p=0.236 in the pooled sample; p=0.673 in FL). The difference is statistically significant in the GN treatment (p=0.048), in which workers' "positive" misalignment is evaluated as more socially appropriate when the leader voluntarily chose self-reporting (vis-à-vis when the leader was exogenously assigned to self-reporting).

- *The worker misaligns by choosing self-reporting when the leader voluntarily chose automatic reporting vs when the leader was exogenously assigned to automatic reporting ("L AR, W SR" vs "L mAR, W SR").* Only in the FL treatment is the average rating in "L mAR, W SR" statistically greater than in "L AR, W SR" (p=0.020 in the FL; but p=0.497 in GN and p=0.236 in the pooled sample). The modal response is 'somewhat socially appropriate' in the "L mAR, W SR," whereas it turns into 'somewhat socially inappropriate' in the "L AR, W SR." Instead, the modal response remains 'somewhat socially inappropriate' in GN and the pooled sample.

- *The worker misreports when the leader voluntarily chose self-reporting vs when the leader was exogenously assigned to self-reporting ("L SR, W SR>1" vs "L mSR, W SR>1");* and *The worker misreports when the leader voluntarily chose automatic reporting vs when the leader was exogenously assigned to automatic reporting ("L AR, W SR>1" vs "L mAR, W SR>1").* Subjects consider workers' misreporting to always be 'somewhat socially

inappropriate,' with no statistically significant difference between the voluntary and mandatory scenarios (at the 5% significance level). This holds if the leader chose/was assigned to either self-reporting (Somers' D-test: p=0.130 in the pooled sample; p=0.078 in GN; p=0.599 in FL) or automatic reporting (Somers' D-test: p= 0.450 in the pooled sample; p=0.597 in GN; p=0.588 in FL).

These results are confirmed by OLS regression analyses (Table 3), in which we controlled for a set of individual characteristics and clustered standard errors at the subject level to adjust for possible heteroscedasticity and correlation of individual-level responses, as we had multiple observations for each participant. Given the nature of the dependent variable (ordinal, with four levels), we also checked that the OLS results are robust to using an ordered logit model. This is indeed the case, as shown by the estimates reported in Table A2 in the supplementary material.

Table 3
OLS regression of appropriateness ratings

DV: Social appropriateness rating	Leader's action	Worker's action				
	(1)	(2)	(3)	(4)	(5)	(6)
Misreport (Misrep)	-0.591***	-0.567***	-0.567***	-0.567***	-0.376***	-0.402***
	(0.047)	(0.050)	(0.050)	(0.050)	(0.044)	(0.047)
Voluntary (V)		0.004	0.004	0.004	0.004	-0.049
		(0.031)	(0.031)	(0.031)	(0.031)	(0.045)
V × Misrep		0.041	0.041	0.004	0.041	0.093*
		(0.036)	(0.036)	(0.044)	(0.036)	(0.045)
Negative Misalignment (NMisal)			-0.315***	-0.333***		
			(0.036)	(0.044)		
V × Misrep × NMisal				0.073		
				(0.052)		
Positive Misalignment (PMisal)					0.382***	0.329***
					(0.061)	(0.067)
V × PMisal						0.106+
						(0.058)
Constant	-0.032	0.210+	0.367**	0.376**	0.018	0.045
	(0.245)	(0.115)	(0.117)	(0.118)	(0.119)	(0.121)
Age	0.010	-0.003	-0.003	-0.003	-0.003	-0.003
	(0.011)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Gender	-0.044	0.018	0.018	0.018	0.018	0.018
	(0.078)	(0.048)	(0.048)	(0.048)	(0.048)	(0.048)
Higher Education	-0.175*	-0.028	-0.028	-0.028	-0.028	-0.028
	(0.085)	(0.049)	(0.049)	(0.049)	(0.049)	(0.049)
Employed	-0.045	0.017	0.017	0.017	0.017	0.017
	(0.090)	(0.052)	(0.052)	(0.052)	(0.052)	(0.052)
Observations	492	1,312	1,312	1,312	1,312	1,312

Notes: This table shows the estimated coefficients from linear regressions on social appropriateness ratings of leader's (col. 1) and workers' (cols. 2-6) actions, on binary variables indicating the particular scenario and reporting choice. The dependent variable (DV) is the social appropriateness rating for a given action and can take four possible values: 'very socially inappropriate'= −1; 'somewhat socially inappropriate'= −1/3; 'somewhat socially appropriate'= +1/3; 'very socially appropriate'= +1. The independent variables are the following: "Misreport (Misrep)" takes value 1 if the firm member misreports the outcome, 0 otherwise; "Voluntary (V)" takes value 1 for the Voluntary scenario and 0 for the Mandatory scenarios; "Negative Misalignment (NMisal)" takes value 1 if the worker chooses self-reporting but the leader used automatic reporting; "Positive Misalignment (PMisal)" takes value 1 if the worker chooses automatic reporting but the leader used self-reporting. All regressions include control variables on individual characteristics. Standard errors clustered at the subject level are in parentheses. The number of observations in col. 1 is given by the total sample (N=164) times the three questions relative to the leader's action, hence N=492; in cols. 2 to 6, it is given by the total sample (N=164) times the remaining eight questions relative to the worker's action, hence N=1,312.
+ p < 0.10 * p < 0.05 ** p < 0.01 *** p < 0.001.

Overall, these results can be summarized as follows:

Result 2: Social norms as identified in Result 1 are not affected by whether the leader's reporting method was voluntarily chosen or exogenously assigned.

Are Social Norms Affected by the Leader's Gender?

As an additional analysis, we investigate whether social norms vary with exposure to female leaders. This gender analysis is not present in Amore et al. (2023); it is not meant to be compared with any prior evidence but to stand "alone" as a test for societal prejudices against female leadership.

Following Barr et al. (2018) and Fallucchi and Nosenzo (2022), we test for treatment differences in two ways. First, we used Mann–Whitney ("MW") rank-sum tests to compare—for each of the eleven actions—the distributions of ratings between the GN baseline condition and the FL treatment. Second, we test the more specific hypothesis that the FL treatment generated a larger share of socially inappropriate or very inappropriate ratings compared to the GN baseline, since these were the ratings mostly associated with unethical actions. For this second test, we used Fisher's exact tests. Table 4 reports the p values of the MW and Fisher tests, alongside the corresponding effect sizes (Cohen's d) with bootstrapped standard errors and normal-based 95% confidence intervals.

Table 4
"Gender Neutral" vs "Female Leader" treatments: Effect sizes and statistical tests

Reporting Choice	Effect size	Bootstrap SE	Normal-based 95% CI		MW	Fisher
"L SR"	0.009	0.161	-0.306	0.323	0.915	0.876
"L SR>1"	0.111	0.146	-0.175	0.397	0.743	0.393
"L SR, W AR"	0.117	0.163	-0.203	0.437	0.382	0.858
"L AR, W SR"	0.046	0.146	-0.239	0.331	0.830	0.874
"L SR, W SR>1"	0.165	0.156	-0.141	0.470	0.355	0.113
"L AR, W SR>1"	0.184	0.165	-0.140	0.507	0.397	0.292
"L mSR>1"	0.081	0.159	-0.231	0.393	0.787	0.421
"L mSR, W AR"	-0.058	0.161	-0.374	0.258	0.513	0.607
"L mSR, W SR>1"	0.073	0.148	-0.218	0.363	0.817	0.465
"L mAR, W SR"	-0.276	0.158	-0.587	0.034	**0.047**	**0.029**
"L mAR, W SR>1"	0.159	0.148	-0.131	0.448	0.446	0.659

Notes For each action (reporting choice), the table reports the standardized mean differences between the Gender Neutral baseline condition and the Female Leader, measured as Cohen's effect size, the bootstrapped standard errors, and normal-based 95% confidence intervals ("CI"). For each action, the table also reports the p values from Mann–Whitney (MW) and Fisher's exact tests. p values in bold in the grey shaded cells are below the 5% threshold for type I error. Abbreviations: L stands for Leader; W for Worker; (m)SR for (Mandatory) Self-Reporting; (m)AR for (Mandatory) Automatic Reporting; SR>1 for 'Self-Reporting, obtaining 1 from the die-rolling, but reporting a value above 1'.

Both the MW rank-sum tests and Fisher's exact tests only show significant differences between GN and FL in one case, when the worker misaligns by choosing self-reporting but the leader was instead exogenously assigned to automatic reporting ("L mAR, W SR;" MW p=0.047; Fisher p=0.029). This means that if the leader is female and was exogenously assigned to automatic reporting, participants rated a worker's negative misalignment as more socially acceptable. As Fisher's exact test reveals, the share of subjects who selected the rating 'very socially inappropriate' or 'somewhat socially inappropriate' for the "L mAR, W SR" action statistically decreases in the FL treatment compared to GN. Interestingly, this difference is uniquely driven by male participants (MW: p=0.020 for males; p=0.510 for females; Fisher: p=0.034 for males; p= 0.394 for females; see Table A3 in the supplementary material).

Regarding the effect size, this difference is relatively strong in magnitude compared to the others. However, this unique significant result is likely to be a false positive among 11 tests. Indeed, overall the effect size remains quite small, and in all other cases, the differences in the distribution of ratings—and the share of the 'very socially inappropriate' or 'somewhat socially inappropriate' ratings—between "Leader's Gender" treatments are much smaller and statistically not significant (Table 4). Additionally, there are also no significant differences between participants' genders: i.e., males and females do not have differential social norms when exposed to female leaders (Table A3).

Overall, these results can be summarized as follows:

Result 3: Social norms as identified in Result 1 do not change with exposure to female leaders.

DISCUSSION

In this paper, we have analyzed social norms on unethical behaviors in experimental firms (groups with 1 leader and 3 workers), and investigated whether those norms are aligned with individuals' actual unethical behaviors. For this purpose, by using the Krupka-Weber

(2013) norm elicitation method we asked participants to assess the social appropriateness of the reporting decisions available to other participants in an earlier behavior experiment by Amore et al. (2023): (i) leader's and workers' untruthful reporting; and (ii) workers' misalignment with their leader's truthful reporting. As an additional analysis, we further tested for societal prejudices against female leadership by investigating whether social norms vary with exposure to female leaders.

Our research reveals three main results. First, a prevailing norm exists against untruthful reporting (see Result 1). Clearly, this norm does not align but is rather disconnected with the behavior observed in Amore et al. (2023), in which most subjects (approx. 80%)—being assigned to the role of either leaders or workers—chose self-reporting and used it to manipulate earnings for personal profits.

Moreover, a clear norm emerges against workers' "negative" misalignment, i.e., choosing untruthful reporting when their leader rather used truthful reporting, and in favor of workers' "positive" misalignment (see Result 1). Also this norm does not align with the behavior observed in Amore et al. (2023), in which only workers' "positive" misalignment was punished by the leader whereas their "negative" misalignment was not.

Second, social norms against untruthful reporting are not affected by the leader's voluntary truthful reporting, i.e., they do not statistically vary with the "Reporting Method" scenarios (see Result 2). This is in contrast with the behavior observed in Amore et al. (2023), in which a leader's truthful reporting choice also made workers statistically more likely to choose truthful reporting, *albeit only* when the leader made a *voluntary* choice.

Third, social norms do not statistically vary with the leader's gender (see Result 3). While being in contrast with long-standing theories on societal prejudices against female leaders (Eagly & Carli, 2003; Kennedy & Kray, 2014; Koburtay et al., 2019), our "null" result is in line with the growing evidence showing that being exposed to female leaders does not

statistically affect behaviors, attitudes, or outcomes (e.g., Lane, 2016; Li et al., 2021). Notably, our finding is also in line with Robin and Babin (1997), showing that males and females have similar ethical judgments in business situations.

To conclude, let us return to our main research question: Are social norms aligned with unethical decision-making? Our results suggest a qualified no: while there exist strong social norms against unethical actions, individuals *do* behave unethically for personal and group gains. This norm-behavior inconsistency shall not appear surprising if jointly read with recent experimental studies revealing that norms do not always perfectly translate into behavior, especially when it comes to illegal or unethical actions (e.g., Huber & Huber, 2020; Guerra & Zhuravleva, 2021; Fallucchi & Nosenzo, 2022). This disconnect may also explain the reason why, under certain circumstances, nudging or addressing social norms has small, unstable, or no effect at all on behavior (e.g., Krupka & Croson, 2016; Sinning & Zhang, 2023).

What are the possible causes for the emergence of this kind of inconsistency? We do not yet have sufficient evidence nor a widely accepted theory that can explain the reasons why and the conditions under which social norms and behaviors diverge. Possible answers—which here are purely speculative, and hence to be tested in future research—could be sought in weak saliency degrees of social norms against illegal behaviors, slower norm-internalization processes, weak internal motivational factors, as well as the specific task or subject pool.[12]

Our study highlights the need for future research in different directions. First of all, further evidence is needed to relate norms and behaviors in statistical analysis, i.e., actually testing for norm-behavior consistency, and to identify the reasons why we observe discrepancies between norms and behavior.

[12] For related discussions, see Batzke and Ernst (2022) and Gross and Vostroknutov (2022).

Another research direction is related to one of our results, that choosing to self-report *and misreport* for one's gain is seen as socially inappropriate while choosing to self-report *per se* is not (Result 1). Further work is warranted to explore whether the different situations as presented to our subjects might have affected that result in particular. While the order of questions was randomized between scenarios, this was not possible within scenarios. Therefore, the perceived social inappropriateness in the unethical situation may arise from the direct comparison with the more ethical situation (Fox & Tversky, 1995).[13]

Last, we shall stress that evidence on societal prejudices, negative stereotypes, discrimination, and a glass ceiling against female leadership is still scant (Chhaochharia et al., 2022; De Paola et al., 2022; Gangadharan et al., 2016, 2019). Our study contributes to it; however, as we attempted many comparisons that might suffer from multiple hypothesis testing without corrections, it calls for more research effort. Relatedly, to enhance the effectiveness of gender manipulations in experimental settings, future research could use avatars instead of, or in addition to, wording differences between "he/she" vs singular pronouns like "she."

Resolving norm-behavior inconsistencies represents a joint, challenging task for researchers and policy-makers alike, especially when it comes to illegal behaviors that generate strong negative externalities in our societies. This paper stresses the need to further investigate norm-behavior discrepancies under different, *specific* illegal contexts. In this sense, we reiterate the argument recently put forward by Bicchieri et al. (2022), that it is important to integrate "the broader social context in the study of norm compliance. Most existing experimental research studies norms in abstract, anonymous, and context-neutral decision settings. While using contextually neutral decision environments is one of the hallmarks of

[13] Furthermore, future research could account for the circular reasoning of the Krupka Weber paradigm for norm elicitation by eliciting personal norms first (unincentivized), and then having subjects guess (incentivized) the modal response.

experimental control, […] this comes at the cost of missing important insights about the drivers of norm compliance" (Bicchieri et al., 2022, p. 70).

Policy-makers could, in turn, rely upon research findings to efficiently address social norms in behavior change programs, hence practically implementing effective norm-based interventions in the field (e.g., nudging norms), with the ultimate goal of curtailing unethical behaviors within and beyond business contexts.

REFERENCES

Abeler, J., Nosenzo, D., & Raymond, C. (2019). Preferences for truth-telling. *Econometrica*, 87(4), 1115-1153.

Amiram, D., Bozanic, Z., Cox, J. D., Dupont, Q., Karpoff, J. M., & Sloan, R. (2018). Financial reporting fraud and other forms of misconduct: a multidisciplinary review of the literature. *Review of Accounting Studies*, 23(2), 732-783.

Amore, M. D., Garofalo, O., & Guerra, A. (2023). How leaders influence (un) ethical behaviors within organizations: A laboratory experiment on reporting choices. *Journal of Business Ethics*, 183(2), 495-510.

Aycinena, D., Rentschler, L., Beranek, B., & Schulz, J. F. (2022). Social norms and dishonesty across societies. *Proceedings of the National Academy of Sciences*, 119(31), e2120138119.

Banerjee, R. (2016). On the interpretation of bribery in a laboratory corruption game: moral frames and social norms. *Experimental Economics*, 19(1), 240-267.

Barr, A., Lane, T., & Nosenzo, D. (2018). On the social inappropriateness of discrimination. *Journal of Public Economics*, 164, 153-164.

Barr, A., Lindelow, M., & Serneels, P. (2009). Corruption in public service delivery: An experimental analysis. *Journal of Economic Behavior & Organization*, 72(1), 225-239.

Bartling, B., & Özdemir, Y. (2023). The limits to moral erosion in markets: Social norms and the replacement excuse. *Games and Economic Behavior*, 138, 143-160.

Batzke, M.C.L., Ernst, A. (2022). Explaining and Resolving Norm-Behavior Inconsistencies—A Theoretical Agent-Based Model. In: Czupryna, M., Kamiński, B. (eds) *Advances in Social Simulation*. Springer Proceedings in Complexity. Springer, Cham.

Behnk, S., Barreda-Tarrazona, I., & García-Gallego, A. (2019). Deception and reputation–An experimental test of reporting systems. *Journal of Economic Psychology*, 71, 37-58.

Bicchieri, C. (2005). The grammar of society: The nature and dynamics of social norms. Cambridge, UK: *Cambridge University Press*.

Bicchieri, C. (2010). Norms, preferences, and conditional behavior. *Politics, philosophy & economics*, 9(3), 297-313.

Bicchieri, C., & Xiao, E. (2009). Do the right thing: but only if others do so. *Journal of Behavioral Decision Making*, 22(2), 191-208.

Bicchieri, C., Dimant, E., Gächter, S., & Nosenzo, D. (2022). Social proximity and the erosion of norm compliance. *Games and Economic Behavior*, 132, 59-72.

Bicchieri, C., Dimant, E., Gelfand, M., & Sonderegger, S. (2023). Social norms and behavior change: The interdisciplinary research frontier. *Journal of Economic Behavior & Organization*, 205, A4-A7.

Bicchieri, C., Xiao, E., & Muldoon, R. (2011). Trustworthiness is a social norm, but trusting is not. *Politics, Philosophy & Economics,* 10(2), 170-187.

Brown, M. E., & Mitchell, M. S. (2010). Ethical and unethical leadership: Exploring new avenues for future research. *Business Ethics Quarterly*, 20(4), 583-616.

Burks, S. V., & Krupka, E. L. (2012). A multimethod approach to identifying norms and normative expectations within a corporate hierarchy: Evidence from the financial services industry. *Management Science*, 58(1), 203-217.

Buso, I. M., De Caprariis, S., Di Cagno, D., Ferrari, L., Larocca, V., Marazzi, F., ... & Spadoni, L. (2020). The effects of COVID-19 lockdown on fairness and cooperation: Evidence from a lablike experiment. *Economics Letters*, 196, 109577.

Buso, I. M., Di Cagno, D., Ferrari, L., Larocca, V., Lorè, L., Marazzi, F., ... & Spadoni, L. (2021). Lab-like findings from online experiments. *Journal of the Economic Science Association*, 7(2), 184-193.

Cagala, T., Glogowsky, U., Rincke, J., & Schudy, S. (2024). Commitment requests do not affect truth-telling in laboratory and online experiments. *Games and Economic Behavior*, 143, 179-190.

Carpenter, J., & Matthews, P. H. (2009). What norms trigger punishment?. *Experimental Economics*, 12(3), 272-288.

Charness, G., Di Bartolomeo, G., & Papa, S. (2024). A stranger in a strange land: Promises and identity. *Games and Economic Behavior*, 144, 13-28.

Chhaochharia, V., Du, M., & Niessen-Ruenzi, A. (2022). Counter-stereotypical female role models and women's occupational choices. *Journal of Economic Behavior & Organization*, 196, 501-523.

Chiapello, M. (2018). Balancetable: Stata module to build a balance table. Technical report, Statistical Software Components S458424 - Boston College Department of Economics.

Cialdini, R. B., & Trost, M. R. (1998). Social influence: Social norms, conformity and compliance. In *The Handbook of Social Psychology*, ed. DT Gilbert, ST Fiske, G Lindzey, 2:151–92. Boston: McGraw-Hill. 4th ed.

d'Adda, G., Darai, D., Pavanini, N., & Weber, R. A. (2017). Do leaders affect ethical conduct?. *Journal of the European Economic Association*, 15(6), 1177-1213.

d'Adda, G., Drouvelis, M., & Nosenzo, D. (2016). Norm elicitation in within-subject designs: Testing for order effects. *Journal of Behavioral and Experimental Economics*, 62, 1-7.

d'Adda, G., Dufwenberg, M., Passarelli, F., & Tabellini, G. (2020). Social norms with private values: Theory and experiments. *Games and Economic Behavior*, 124, 288-304.

De Paola, M., Gioia, F., & Scoppa, V. (2022). Female leadership: Effectiveness and perception. *Journal of Economic Behavior & Organization*, 201, 134-162.

Dorfleitner, G., Kreuzer, C., & Sparrer, C. (2022). To sin in secret is no sin at all: On the linkage of policy, society, culture, and firm characteristics with corporate scandals. *Journal of Economic Behavior & Organization*, 202, 762-784.

Drouvelis, M. (2021). *Social Preferences: An Introduction to Behavioural Economics and Experimental Research*. Newcastle upon Tyne, UK: Agenda Publishing.

Eagly, A. H., & Carli, L. L. (2003). The female leadership advantage: An evaluation of the evidence. *The Leadership Quarterly*, 14(6), 807-834.

Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological Review*, 109(3), 573.

Eckel, C. C., Hoover, H. G., Krupka, E. L., Sinha, N., & Wilson, R. K. (2023). Using social norms to explain giving behavior. *Experimental Economics*, 26(5), 1115-1141.

Fallucchi, F., & Nosenzo, D. (2022). The coordinating power of social norms. *Experimental Economics*, 25(1), 1-25.

Feess, E., Kerzenmacher, F., & Muehlheusser, G. (2023). Morally questionable decisions by groups: Guilt sharing and its underlying motives. *Games and Economic Behavior*, 140, 380-400.

Feltovich, N. (2019). The interaction between competition and unethical behaviour. *Experimental Economics*, 22(1), 101-130.

Ferreira, F., & Gyourko, J. (2014). Does gender matter for political leadership? The case of US mayors. *Journal of Public Economics*, 112, 24-39.

Fochmann, M., Fochmann, N., Kocher, M. G., & Müller, N. (2021). Dishonesty and risk-taking: Compliance decisions of individuals and groups. *Journal of Economic Behavior & Organization*, 185, 250-286.

Fox, C. R., & Tversky, A. (1995). Ambiguity aversion and comparative ignorance. *The Quarterly Journal of Economics*, 110(3), 585-603.

Fréchette, G. R. (2012). Session-effects in the laboratory. *Experimental Economics*, 15(3), 485-498.

Gangadharan, L., Jain, T., Maitra, P., & Vecci, J. (2016). Social identity and governance: The behavioral response to female leaders. *European Economic Review*, 90, 302-325.

Gangadharan, L., Jain, T., Maitra, P., & Vecci, J. (2019). Female leaders and their response to the social environment. *Journal of Economic Behavior & Organization*, 164, 256-272.

Gächter, S., Gerhards, L., & Nosenzo, D. (2017). The importance of peers for compliance with norms of fair sharing. *European Economic Review*, 97, 72-86.

Gibson, R., Tanner, C., & Wagner, A. F. (2013). Preferences for truthfulness: Heterogeneity among and within individuals. *American Economic Review*, 103(1), 532-48.

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

Goldstein, N. J., & Cialdini, R. B. (2011). Managing normative influences in organizations. In *Social Psychology and Organizations* (pp. 99-118). Routledge.

Govindan, P. (2022). How do informal norms affect rule compliance: Experimental evidence. *Journal of Behavioral and Experimental Economics*, 96, 101795.

Greiner, B. (2015). Subject Pool Recruitment Procedures: Organizing Experiments with ORSEE. *Journal of the Economic Science Association*, 1, 1–12.

Grimm, S. (2019). Effects of choice observability on risk taking: The role of norms. *Journal of Behavioral and Experimental Economics*, 80, 34-46.

Gross, J., & Vostroknutov, A. (2022). Why do people follow social norms?. *Current Opinion in Psychology*, 44, 1-6.

Grosser, K., Moon, J., Freeman, R. E., & Nelson, J. (2014). Special issue on: Gender, business ethics, and corporate social responsibility. *Business Ethics Quarterly*, 24(3), 497-500.

Guerra, A., & Zhuravleva, T. (2021). Do bystanders react to bribery?. *Journal of Economic Behavior & Organization*, 185, 442-462.

Heilman, M. E. (2012). Gender stereotypes and workplace bias. *Research in Organizational Behavior*, 32, 113–135.

Hollway, W. (2007). *The capacity to care: Gender and ethical subjectivity*. London, New York, NY: Routledge

Huber, C., & Huber, J. (2020). Bad bankers no more? Truth-telling and (dis) honesty in the finance industry. *Journal of Economic Behavior & Organization*, 180, 472-493.

Huber, C., Litsios, C., Nieper, A., & Promann, T. (2023). On social norms and observability in (dis) honest behavior. *Journal of Economic Behavior & Organization*, 212, 1086-1099.

Kassas, B., & Palma, M. A. (2019). Self-serving biases in social norm compliance. *Journal of Economic Behavior & Organization,* 159, 388-408.

Kennedy, J. A., & Kray, L. J. (2014). Who is willing to sacrifice ethical values for money and social status? Gender differences in reactions to ethical compromises. *Social Psychological and Personality Science*, 5(1), 52-59.

Kimbrough, E. O., Krupka, E. L., Kumar, R., Murray, J. M., Ramalingam, A., Sánchez-Franco, S., ... & Hunter, R. F. (2024). On the stability of norms and norm-following propensity: A cross-cultural panel study with adolescents. *Experimental Economics*, 27, 351-378.

Kleven, H. J., Knudsen, M. B., Kreiner, C. T., Pedersen, S., & Saez, E. (2011). Unwilling or unable to cheat? Evidence from a tax audit experiment in Denmark. *Econometrica*, 79(3), 651-692.

Koburtay, T., Syed, J., & Haloub, R. (2019). Congruity between the female gender role and the leader role: A literature review. *European Business Review,* 31(6), 831-848.

Kocher, M. G., Schudy, S., & Spantig, L. (2018). I lie? We lie! Why? Experimental evidence on a dishonesty shift in groups. *Management Science*, 64(9), 3995-4008.

Kocher, M. G., Martinsson, P., Myrseth, K. O. R., & Wollbrant, C. E. (2017). Strong, bold, and kind: Self-control and cooperation in social dilemmas. *Experimental Economics*, 20(1), 44-69.

Kölle, F., & Quercia, S. (2021). The influence of empirical and normative expectations on cooperation. *Journal of Economic Behavior & Organization,* 190, 691-703.

Krupka, E. L., & Croson, R. T. (2016). The differential impact of social norms cues on charitable contributions. *Journal of Economic Behavior & Organization*, 128, 149-158.

Krupka, E., & Weber, R. A. (2009). The focusing and informational effects of norms on pro-social behavior. *Journal of Economic Psychology*, 30(3), 307-320.

Krupka, E. L., & Weber, R. A. (2013). Identifying social norms using coordination games: Why does dictator game sharing vary?. *Journal of the European Economic Association*, 11(3), 495-524.

Krupka, E. L., Weber, R., Croson, R. T., & Hoover, H. (2022). "When in Rome": Identifying social norms using coordination games. *Judgment & Decision Making*, 17(2).

Lane, T. (2016). Discrimination in the laboratory: A meta-analysis of economics experiments. *European Economic Review*, 90, 375-402.

Lang, M., & Schudy, S. (2023). (Dis) honesty and the value of transparency for campaign promises. *European Economic Review*, 159, 104560.

Legros, S., & Cislaghi, B. (2020). Mapping the social-norms literature: An overview of reviews. *Perspectives on Psychological Science*, 15(1), 62-80.

Li, Y., Sbai, E., & Chaudhuri, A. (2021). An experimental study of gender differences in agency relationships. *Journal of Behavioral and Experimental Economics*, 90, 101650.

Lynn, M. (2018). How motivations for tipping vary with occupational differences in descriptive tipping norms. *Journal of Behavioral and Experimental Economics*, 77, 1-10.

Morris, M. W., Hong, Y. Y., Chiu, C. Y., & Liu, Z. (2015). Normology: Integrating insights about social norms to understand cultural dynamics. *Organizational Behavior and Human Decision Processes*, 129, 1-13.

Newson, R. (2002). Parameters behind “nonparametric” statistics: Kendall's tau, Somers’ D and median differences. *The Stata Journal*, 2(1), 45-64.

Newson, R. (2006). Confidence intervals for rank statistics: Somers’ D and extensions. *The Stata Journal*, 6(3), 309-334.

Nosenzo, D., & Görges, L. (2020). Measuring social norms in economics: why it is important and how it is done. *Analyse & Kritik*, 42(2), 285-312.

Pullen, A., & Vachhani, S. J. (2021). Feminist ethics and women leaders: From difference to intercorporeality. *Journal of Business Ethics*, 173(2), 233-243.

Rhodes, L. A., Sánchez, G. E., Espinoza, N. E., & Borja, V. (2024). The role of social norms on the willingness to act and donate against sexual harassment. *International Review of Economics*, 71(3), 257-271.

Rilke, R. M., Danilov, A., Weisel, O., Shalvi, S., & Irlenbusch, B. (2021). When leading by example leads to less corrupt collaboration. *Journal of Economic Behavior & Organization*, 188, 288-306.

Robin, D., & Babin, L. (1997). Making sense of the research on gender and ethics in business: A critical analysis and extension. *Business Ethics Quarterly*, 7(4), 61-90.

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

Santoro, F. (2021). To file or not to file? Another dimension of tax compliance-the Eswatini Taxpayers’ survey. *Journal of Behavioral and Experimental Economics*, 95, 101760.

Schmidt, R., Schwieren, C., & Sproten, A. N. (2020). Norms in the lab: Inexperienced versus experienced participants. *Journal of Economic Behavior & Organization*, 173, 239-255.

Schram, A., Di Zheng, J., & Zhuravleva, T. (2022). Corruption: A cross-country comparison of contagion and conformism. *Journal of Economic Behavior & Organization*, 193, 497-518.

Sinning, M., & Zhang, Y. (2023). Social norms or enforcement? A natural field experiment to improve traffic and parking fine compliance. *Journal of Economic Behavior & Organization*, 210, 43-60.

Stajkovic, K., & Stajkovic, A. D. (2024). Ethics of Care Leadership, Racial Inclusion, and Economic Health in the Cities: Is There a Female Leadership Advantage?. *Journal of Business Ethics*, 189(3), 699-721.

Traub, S., Schwaninger, M., Paetzel, F., & Neuhofer, S. (2023). Evidence on need-sensitive giving behavior: An experimental approach to the acknowledgment of needs. *Journal of Behavioral and Experimental Economics*, 105, 102028.

Trevino, L. K. (1992). Experimental approaches to studying ethical-unethical behavior in organizations. *Business Ethics Quarterly*, 2(2), 121-136.

te Velde, V. L., & Louis, W. (2022). Conformity to descriptive norms. *Journal of Economic Behavior & Organization*, 200, 204-222.

Upton, G. J. (1992). Fisher's exact test. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 155(3), 395-402.

Vranka, M., Hudík, M., Frollová, N., Bahník, Š., Sýkorová, M., & Houdek, P. (2021). Honesty of online workers: A field experiment shows no evidence of self-selection of cheaters to a cheating-enabling work environment. *Journal of Behavioral and Experimental Economics*, 94, 101730.

Xu, C., & Schriesheim, C. A. (2018). Deceptive influence in the workplace. *Academy of Management Proceedings*, 1, 14976.

Zhang, P., Zhang, Y., & Palma, M. A. (2024). Social roles and competitiveness: My willingness to compete depends on who I am (supposed to be). *Games and Economic Behavior*, 143, 125-151.

Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. *Experimental Economics*, 13(1), 75–98.

Supplementary material

Social norms on unethical behaviors in the workplace: A lab experiment

This supplementary material contains additional tables (Appendix A), experimental instructions and supporting information (Appendix B).

Appendix A: Additional tables

Table A1: Sample descriptives and balance tests

Variable	Pooled Sample	Gender Neutral (GN)	Female Leader (FL)	Difference (FL – GN)
Age	25.238	26.127	24.412	-1.715**
	(4.965)	(5.419)	(4.376)	(0.773)
Male	0.463	0.443	0.482	0.039
	(0.500)	(0.500)	(0.503)	(0.078)
North of Italy	0.963	0.962	0.965	0.003
	(0.188)	(0.192)	(0.186)	(0.030)
Higher Education	0.634	0.759	0.518	-0.242***
	(0.483)	(0.430)	(0.503)	(0.073)
Employed	0.372	0.430	0.318	-0.113
	(0.484)	(0.498)	(0.468)	(0.076)
Economics student	0.213	0.241	0.188	-0.052
	(0.411)	(0.430)	(0.393)	(0.065)
Observations	164	79	85	164

Notes: The balance table reports the means and difference in means for each variable between treatments, with standard deviations (of means) and p values (of the difference in means) in parentheses. Standard errors are clustered at the individual level, and p values are computed according to Chiapello (2018).
***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

Table A2: Ordered logit regression of appropriateness ratings.

DV: Social appropriateness rating	Leader's action	Worker' action				
	(1)	(2)	(3)	(4)	(5)	(6)
Misreport (Misrep)	-1.883***	-1.763***	-1.896***	-1.898***	-1.256***	-1.326***
	(0.178)	(0.182)	(0.207)	(0.208)	(0.156)	(0.164)
Voluntary (V)		0.009	-0.009	-0.009	-0.010	-0.151
		(0.088)	(0.093)	(0.094)	(0.093)	(0.132)
V × Misrep		0.133	0.154	0.108	0.155	0.296*
		(0.106)	(0.112)	(0.135)	(0.111)	(0.135)
Negative Misalignment (NMisal)			-1.074***	-1.098***		
			(0.138)	(0.159)		
V × Misrep × NMisal				0.100		
				(0.170)		
Positive Misalignment (PMisal)					1.172***	1.027***
					(0.204)	(0.214)
V × PMisal						0.291+
						(0.175)
Age	0.027	-0.013	-0.014	-0.014	-0.015	-0.015
	(0.039)	(0.017)	(0.018)	(0.018)	(0.018)	(0.018)
Gender	-0.094	0.051	0.062	0.062	0.065	0.066
	(0.253)	(0.152)	(0.158)	(0.158)	(0.156)	(0.156)
Higher Education	-0.491+	-0.083	-0.101	-0.101	-0.102	-0.100
	(0.275)	(0.154)	(0.160)	(0.160)	(0.158)	(0.158)
Employed	-0.150	0.091	0.085	0.085	0.087	0.086
	(0.297)	(0.161)	(0.167)	(0.167)	(0.163)	(0.164)
Cut-off 1	-1.685+	-2.488***	-3.188***	-3.201***	-2.036***	-2.103***
	(0.866)	(0.404)	(0.443)	(0.449)	(0.410)	(0.417)
Cut-off 2	0.216	-0.638	-1.234**	-1.248**	-0.145	-0.211
	(0.868)	(0.392)	(0.417)	(0.422)	(0.406)	(0.413)
Cut-off 3	1.608+	0.934*	0.434	0.423	1.532***	1.468***
	(0.899)	(0.388)	(0.405)	(0.408)	(0.414)	(0.420)
Observations	492	1,312	1,312	1,312	1,312	1,312

Notes: This table presents a robustness check for the OLS regression results reported in Table 3, by showing the estimated coefficients from ordered logit regressions on social appropriateness ratings of leader's (col. 1) and workers' (cols. 2-6) actions. The dependent variable (DV) and covariates are the same as in Table 3. Specifically, the DV is the social appropriateness rating for a given action, and can take four possible values: 'very socially inappropriate'= −1; 'somewhat socially inappropriate'= −1/3 ; 'somewhat socially appropriate'= +1/3 ; 'very socially appropriate'= +1. The independent variables are the following: "Misreport (Misrep)" takes value 1 if the firm member misreports the outcome, 0 otherwise; "Voluntary (V)" takes value 1 for the Voluntary scenario and 0 for the Mandatory scenarios; "Negative Misalignment (NMisal)" takes value 1 if the worker chooses self-reporting but the leader used automatic reporting; "Positive Misalignment (PMisal)" takes value 1 if the worker chooses automatic reporting but the leader used self-reporting. All regressions include control variables on individual characteristics. Standard errors clustered at the subject level are in parentheses. The number of observations in col. 1 is given by the total sample (N=164) times the three questions relative to the leader's action, hence N=492; in cols. 2 to 6, it is given by the total sample (N=164) times the remaining eight questions relative to the worker's action, hence N=1,312.
+ p < 0.10 * p < 0.05 ** p < 0.01 *** p < 0.001.

Table A3: "Gender Neutral" vs "Female Leader" treatments by participants' gender: Effect sizes and statistical tests

Reporting Choice	MW		Fisher	
	Male	Female	Male	Female
"L SR"	0.960	0.986	1	0.831
"L SR>1"	0.409	0.779	0.206	1
"L SR, W AR"	0.716	0.125	1	0.792
"L AR, W SR"	0.143	0.140	0.241	0.195
"L SR, W SR>1"	0.442	0.593	0.609	0.145
"L AR, W SR>1"	0.295	0.823	0.259	0.757
"L mSR>1"	0.280	0.556	0.327	1
"L mSR, W AR"	0.544	0.761	1	0.644
"L mSR, W SR>1"	0.874	0.869	1	0.317
"L mAR, W SR"	**0.020**	0.510	**0.034**	0.394
"L mAR, W SR>1"	0.746	0.441	1	0.352

Notes: For each action (reporting choice), the table reports the p values from Mann–Whitney (MW) and Fisher's exact tests. p values in bold in the grey shaded cells are below the 5% threshold for type I error. Abbreviations: L stands for Leader; W for Worker; (m)SR for (Mandatory) Self-Reporting; (m)AR for (Mandatory) Automatic Reporting; SR>1 for 'Self-Reporting, obtaining 1 from the die-rolling, but reporting a value above 1'.

Table A4: Summary statistics

Reporting Choice		Treatment		Total
		Gender Neutral	Female Leader	
"L SR"	Obs.	79	85	164
	Mean	0.07	0.07	0.07
	SE	0.07	0.06	0.05
	95% CI LL	-0.06	-0.06	-0.02
	95% CI UL	0.20	0.19	0.16
"L SR>1"	Obs.	79	85	164
	Mean	-0.53	-0.59	-0.56
	SE	0.07	0.06	0.05
	95% CI LL	-0.67	-0.71	-0.65
	95% CI UL	-0.39	-0.48	-0.47
"L SR, W AR"	Obs.	79	85	164
	Mean	0.38	0.32	0.35
	SE	0.07	0.06	0.04
	95% CI LL	0.25	0.20	0.26
	95% CI UL	0.51	0.44	0.44
"L AR, W SR"	Obs.	79	85	164
	Mean	-0.07	-0.10	-0.09
	SE	0.07	0.06	0.04
	95% CI LL	-0.21	-0.21	-0.17
	95% CI UL	0.06	0.02	0.00
"L SR, W SR>1"	Obs.	79	85	164
	Mean	-0.22	-0.31	-0.26
	SE	0.07	0.06	0.04
	95% CI LL	-0.35	-0.42	-0.35
	95% CI UL	-0.08	-0.20	-0.18
"L AR, W SR>1"	Obs.	79	85	164
	Mean	-0.47	-0.58	-0.52
	SE	0.07	0.06	0.05
	95% CI LL	-0.61	-0.69	-0.62
	95% CI UL	-0.32	-0.46	-0.43

Table A4 (Cont'd.)

Reporting Choice		Treatment		Total
		Gender Neutral	Female Leader	
"L mSR>1"	Obs.	79	85	164
	Mean	-0.46	-0.51	-0.48
	SE	0.07	0.06	0.04
	95% CI LL	-0.60	-0.62	-0.57
	95% CI UL	-0.32	-0.39	-0.40
"L mSR, W AR"	Obs.	79	85	164
	Mean	0.27	0.31	0.29
	SE	0.06	0.07	0.05
	95% CI LL	0.15	0.17	0.20
	95% CI UL	0.40	0.45	0.39
"L mSR, W SR>1"	Obs.	79	85	164
	Mean	-0.30	-0.34	-0.32
	SE	0.07	0.06	0.04
	95% CI LL	-0.44	-0.46	-0.41
	95% CI UL	-0.26	-0.23	-0.23
"L mAR, W SR"	Obs.	79	85	164
	Mean	-0.11	0.04	-0.04
	SE	0.07	0.05	0.04
	95% CI LL	-0.24	-0.07	-0.12
	95% CI UL	0.02	0.14	0.05
"L mAR, W SR>1"	Obs.	79	85	164
	Mean	-0.51	-0.60	-0.56
	SE	0.07	0.06	0.04
	95% CI LL	-0.65	-0.71	-0.64
	95% CI UL	-0.37	-0.49	-0.47

Notes: This table presents the summary statistics of the variables used in Figure 2, and in the analyses presented in Section 3– by treatment and pooled across treatments. The summary statistics include the number of observations (Obs.), means, standard errors (SE), and the 95% confidence intervals (CI), where LL and UL are the lower and upper limits, respectively. Abbreviations: L stands for Leader; W for Worker; (m)SR for (Mandatory) Self-Reporting; (m)AR for (Mandatory) Automatic Reporting; SR>1 for 'Self-Reporting, obtaining 1 from the die-rolling, but reporting a value above 1'.

Table A5
Percentage of subjects matching the modal response

(c) "Voluntary" Scenario: The leader can freely choose between self-reporting (SR) and automatic reporting (AR)

	Leader chooses SR instead of AR "L SR"			Leader chooses SR and misreports "L SR>1"			Leader chooses SR, Worker misaligns (AR) "L SR, W AR"		
	GN	FL	Pooled	GN	FL	Pooled	GN	FL	Pooled
Match (%)	43.04	47.06	45.12	55.70	55.29	55.49	39.24	45.88	42.68
High Mismatch (%)	18.99	18.82	18.90	18.99	12.94	15.85	5.06	4.71	4.88

	Leader chooses AR, Worker misaligns (SR) "L AR, W SR"			Leader chooses SR, Worker aligns (SR) and misreports "L SR, W SR>1"			Leader chooses AR, Worker misaligns (SR) and misreports "L AR, W SR>1"		
	GN	FL	Pooled	GN	FL	Pooled	GN	FL	Pooled
Match (%)	43.04	47.06	45.12	41.77	55.29	48.78	49.37	52.94	51.22
High Mismatch (%)	12.66	8.24	10.37	10.13	5.88	7.93	20.25	12.94	16.46

(d) "mSR" Scenario: The leader is mandatorily assigned to self-reporting (SR)

	Leader misreports "L mSR>1"			Worker misaligns (AR) "L mSR, W AR"			Worker aligns (SR) and misreports "L mSR, W SR>1"		
	GN	FL	Pooled	GN	FL	Pooled	GN	FL	Pooled
Match (%)	45.57	44.71	45.12	40.51	38.82	39.63	43.04	51.76	47.56
High Mismatch (%)	20.25	15.29	17.68	5.06	10.59	7.93	8.86	4.71	6.71

(c) "mAR" Scenario: The leader is mandatorily assigned to automatic reporting (AR)

	Worker misaligns (SR) "L mAR, W SR"			Worker misaligns (SR) and misreports "L mAR, W SR>1"		
	GN	FL	Pooled	GN	FL	Pooled
Match (%)	48.10	45.88	46.95	50.63	55.29	53.05
High Mismatch (%)	11.39	7.06	9.15	16.46	12.94	14.63

Notes: Each cell shows the percentage of subjects perfectly matching ("Match") the modal response for a given action. "High Mismatch" indicates a high mismatch between the individual's answer and the modal response; specifically, it reports the percentage of subjects who ranked the action as somewhat/very socially appropriate while the modal response was somewhat/very socially inappropriate, or viceversa. The statistics are reported both pooled across the Leader's Gender treatments, and separately by treatment. Abbreviations: L stands for Leader; W for Worker; (m)SR for (Mandatory) Self-Reporting; (m)AR for (Mandatory) Automatic Reporting; SR>1 for 'obtaining 1 from die-rolling but declaring a value above 1'. Number of participants: 164 (Pooled), of which 79 in the GN treatment and 85 in the FL treatment.

Appendix B: Experiment instructions and supporting information

This Appendix contains an English translation of the instructions to participants for both the "Gender Neutral" and "Female" leader treatments (Section B1), and the post-experiment questionnaire, which is the same for both treatments (Section B2). The original instructions in Italian are available upon request. A selective set of Screenshots taken during the experimental sessions is provided in Section B3.

B1: Experiment instructions to participants

This section contains an English translation of the instructions to participants for the "Gender Neutral" leader treatment, with the differences for the "Female" leader treatment in bordered boxes.

Screen 1

General Information

Welcome! Thank you for participating in this study.

Please note that it is preferable not to use a mobile device, as the study may not be correctly visualized. Remember that you will not be allowed to go back to the previous page (you cannot click on "Back") during the study.

Please carefully read the instructions. You will receive 2 euros as a show-up fee. In addition, you will receive a payment that will depend on your decisions during the experiment. You will receive the payment via PayPal to the email address you previously provided, within 14 days from the completion of the study. The participants who will abandon the study before its completion, will not receive any payment.

Your identity will remain anonymous, and your choices will remain confidential, i.e., they will not be communicated to the other participants and none can link them to your identity. The anonymized data will be only used for scientific purposes.

We kindly ask you to keep your microphone muted and the webcam switched off during the whole duration of the session.

Feel free to ask any question, at any point in time. In case you have a question, write a private Zoom chat message to one of the collaborators, or click on the "Raise your hand" button (close to the "Reactions of Zoom" button).

Screen 2

In this study, your task is to evaluate the possible choices of various individuals, in different situations. We will ask you to indicate, for each choice, what you expect the majority of people would agree to be inappropriate or appropriate.

The scenario is the following. A firm is composed of four members: one **female leader** and three **workers** (females or males). The final firm profit is the sum of the reported performance of each firm

member. Each member can report their own performance via two methods: **self-reporting** or **"automatic reporting."**
Self-reporting means that a firm member rolls autonomously and privately a physical six-faced die, and inputs the result of the roll. Note that no one (neither the firm members, nor third parties) will be able to verify the actual result of the die roll.

Automatic reporting means that a firm member rolls a virtual six-faced die (via the computer), and the outcome is automatically reported by the computer without any possibility to modify the obtained result.

In some cases, the female leader is free to choose the reporting method, while in other cases she s/he is mandatorily assigned to self-reporting or automatic reporting. The workers, instead, are always free to choose between the two methods.

Specifically:

1. The female leader is the first mover, meaning that she s/he is the first to choose the reporting method, or she s/he has been assigned the method to be used.
2. The workers are the second mover, meaning that they choose the reporting method after being informed about the reporting method used by their female leader.
3. The female leader and each worker roll their own die based on the method chosen/assigned.
4. The sum of the results obtained by the rolls constitutes the firm profit, which will be equally split among firm members.

Screen 3

What should you do?

In the following pages, you will read situations similar to this one, and you will be asked to evaluate each of the possible choices of the different firm members (e.g., choosing self-reporting rather than automatic reporting). For each choice, you will have to evaluate if it is socially appropriate, selecting one of the following alternatives: "very socially inappropriate," "somewhat socially inappropriate," "somewhat socially appropriate" and "very socially appropriate."

By **socially appropriate**, we mean the choice that **the majority of people** agree to be the "correct" or "ethical" thing to do.

Remember: we are not asking you what you personally think should be done, but what you think the majority of people believe to be socially appropriate or inappropriate.

How much will you earn?

In total, you will be presented with three situations. For each one of them, there will be a set of choices to be evaluated. In total, you will have to evaluate **eleven different choices**.

At the end of the study, we will randomly select **one of these eleven choices** for your payment. **For this selected choice, we will check the answer that has been more frequently given by all the participants in this study (i.e., the "modal" answer).**

If you guess the modal answer, you will receive 7 euros, in addition to the show-up fee of 2 euros; **otherwise, you will only receive the show-up fee of 2 euros**. In other words, if you will give the answer most frequently given by the other participants in this study, you will receive an additional 7 euros.

Screen 4

Before starting the study, we ask you the following comprehension questions. **To continue, you will have to answer all questions correctly**.

1. The female leader and the three workers can always choose the reporting method.
True False

If answering True, the Feedback window reports the following text:

The correct answer is "False": In some cases, the female leader is free to choose the reporting method, while in other cases she s/he is mandatorily assigned to self-reporting or automatic reporting. The workers, instead, are always free to choose between the two methods.

2. Workers decide the reporting method after having observed the reporting method used by their female leader.
True False

If answering False, the Feedback window reports the following text:

The correct answer is "True": the female leader is the first mover, meaning that she s/he is the first to choose the reporting method, or to be informed about the reporting method she s/he has been assigned to. The workers are the second movers, meaning that they choose the reporting method after having observed the one used by their female leader.

3. Self-reporting allows one to report a different result from the one truly obtained from the die-rolling, whereas automatic reporting does not allow for any change.
True False

If answering False, the Feedback window reports the following text:

The correct answer is "True": self-reporting means that a subject rolls a six-face die autonomously and privately, and reports the result of the roll. Note that nobody (neither the firm members, nor third parties) will be able to verify the actual result of the die-rolling. automatic reporting means that a subject rolls a virtual six-face die (via computer), and the result of the roll is automatically reported by the computer without any possibility for the subject to change the obtained result.

4. Only if you guess the modal answer, you will receive 9 euros (7 euros in addition to the 2 euros as show-up fee); otherwise, you will only get the 2 euros as show-up fee.
True False

If answering False, the Feedback window reports the following text:

The correct answer is "True": if you guess the modal answer, you will receive 7 euros more in addition to the show-up fee of 2 euros; otherwise, you will only receive the show-up fee of 2 euros. In other words, if you will provide the answer most frequently given by the other participants in this study, you will receive an additional 7 euros.

Screen 5

If everything is clear and there are no questions, click on the arrow to **start the study**.

Screen 6

Suppose that the female leader can choose between self-reporting and automatic reporting.

Remember: each worker can always choose between self-reporting and automatic reporting.

For each of the following choices, indicate if it is "very socially inappropriate," "somewhat socially inappropriate," "somewhat socially appropriate," or "very socially appropriate." Remember: we are asking you **what you think the majority of people believe to be socially appropriate or inappropriate**.

	very socially inappropriate	somewhat socially inappropriate	somewhat socially appropriate	very socially appropriate
The female leader chooses self-reporting rather than automatic reporting.	○	○	○	○
The female leader chooses self-reporting, obtains 1 from the die-rolling, but reports a value above 1.	○	○	○	○
The worker sees the female leader choosing self-reporting but chooses automatic reporting.	○	○	○	○
The worker sees the female leader choosing automatic reporting but chooses self-reporting.	○	○	○	○
The worker sees the female leader choosing self-reporting, chooses self-reporting, obtains 1 from the die-rolling, but reports a value above 1.	○	○	○	○
The worker sees the female leader choosing automatic reporting, chooses self-reporting, obtains 1 from the die-rolling, but reports a value above 1.	○	○	○	○

Suppose that the female leader CANNOT choose between self-reporting and automatic reporting, and she is mandatorily assigned to self-reporting.

Remember: each worker can always choose between self-reporting and automatic reporting.

For each of the following choices, indicate if it is "very socially inappropriate" "somewhat socially inappropriate," "somewhat socially appropriate" or "very socially appropriate." Remember: we are

asking you **what you think the majority of people believe to be socially appropriate or inappropriate**.

	very socially inappropriate	somewhat socially inappropriate	somewhat socially appropriate	very socially appropriate
The female leader is mandatorily assigned to self-reporting, obtains 1 from the die-rolling, but reports a value above 1.	○	○	○	○
The worker sees that the female leader is mandatorily assigned to self-reporting but chooses automatic reporting.	○	○	○	○
The worker sees that the female leader is mandatorily assigned to self-reporting, chooses "self-reporting, obtains 1 from the die-rolling, but reports a value above 1.	○	○	○	○

Suppose that the female leader CANNOT choose between self-reporting and automatic reporting, and she is mandatorily assigned to automatic reporting.

Remember: each worker can always choose between self-reporting and automatic reporting.

For each of the choices, indicate if it is "very socially inappropriate," "somewhat socially inappropriate," "somewhat socially appropriate" or "very socially appropriate." Remember: we are asking you **what you think the majority of people believe to be socially appropriate or inappropriate**.

	very socially inappropriate	somewhat socially inappropriate	somewhat socially appropriate	very socially appropriate
The worker sees that the female leader is mandatorily assigned to automatic reporting but chooses self-reporting.	○	○	○	○
The worker sees that the female leader is mandatorily assigned to automatic reporting, chooses self-reporting, obtains 1 from the die-rolling, but reports a value above 1.	○	○	○	○

Screen 7

Thank you for having participated in the study! Now, we ask you to fill out a brief questionnaire.

Once everyone will have completed the questionnaire, we will give you instructions about the payment. So, do not close either the page or Zoom.

B2. Post-experiment questionnaire

Screen 8

Sex

- Female
- Male

Age

manual entry

In which Italian region were you born?

If you were not born in Italy, select "Other"

A drop-down menu with the following options: Aosta Valley, Liguria, Lombardy, Piedmont, Trentino-South Tyrol, Veneto, Friuli Venezia Giulia, Emilia-Romagna, Tuscany, Umbria, Marche, Lazio, Abruzzo, Molise, Campania, Apulia, Basilicata, Calabria, Sicily, Sardinia, Other

What is the highest education level you have currently completed?

- Junior High School
- High School
- Bachelor's Degree
- Master's Degree
- Master
- PhD
- Other

What is your current occupation?

- Student
- Self-employed
- Employed and/or project collaborator
- Student and Employed/Self-Employed
- Unemployed
- Other

What is your major field of study?

- Economics
- Natural, physical, and mathematical science
- Engineering and architecture
- Medicine
- Letters and Philosophy
- Sociology
- Political Sciences
- Other Social Sciences

Was your father born in Italy?

- ○ Yes
- ○ No

Was your mother born in Italy?

- ○ Yes
- ○ No

Have you ever participated in experiments similar to this one?

- ○ Yes
- ○ No

Have you already participated in other academic research? (select one or more answers)

- ☐ YES, economic experiments
- ☐ YES, psychological experiments
- ☐ YES, medical and biological experiments
- ☐ NONE of the previous answers

Screen 9

Thank you for having completed the study!

You will receive your payment via PayPal within 14 days, to the email address you have previously provided.

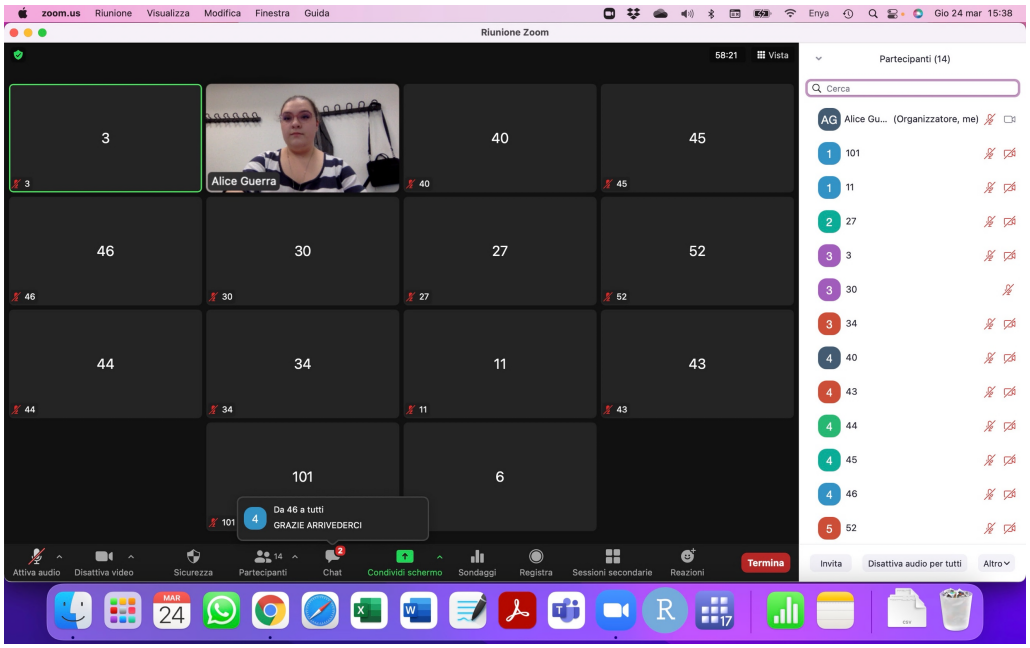
Now you can close this page.

DO NOT leave the Zoom virtual meeting. Please wait for our indications.

Thank you.

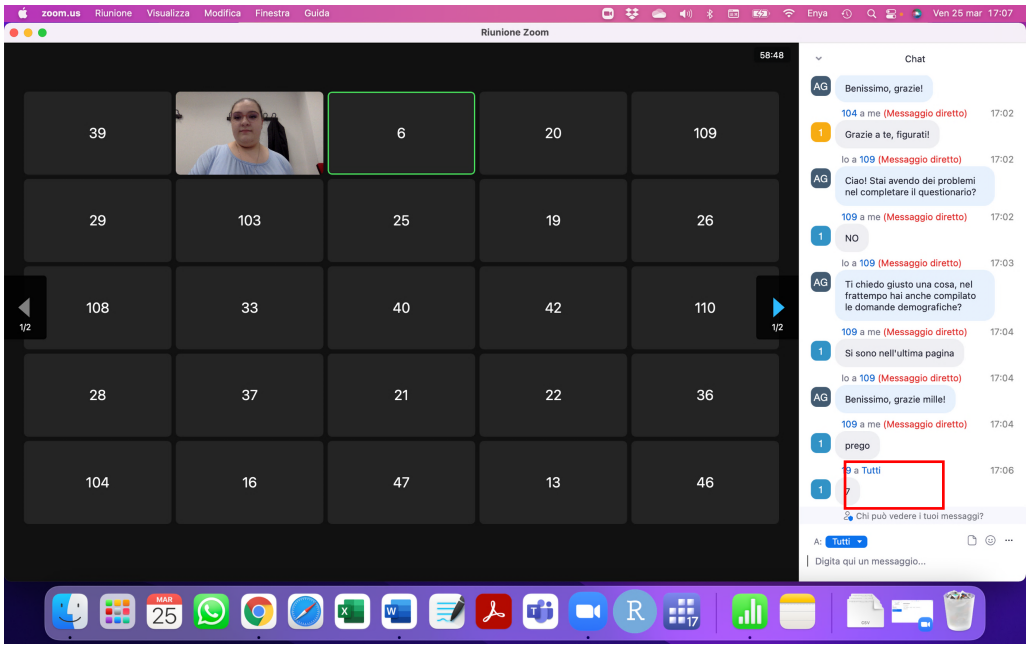
B3: Screenshots taken during the experimental sessions

Screenshot 1: An example screen of the virtual lab on Zoom



Notes: This Screenshot shows our ad-hoc virtual lab on Zoom, where only the instructor ("Organizzatore, me" in the participants' list on the right) had the webcam and microphone turned on. To ensure anonymity, subjects were instructed to always keep their webcams and microphones switched off, and were asked to replace their names with an ID code that we randomly assigned them and privately communicated via email through Qualtrics the day before the experimental session.

Screenshot 2: An example screen of the random selection of the question for participants' payment



Notes: This Screenshot shows how we randomly selected the question to compute participants' payment. Specifically, at the end of each session, with all participants still present, we randomly selected one among the eleven questions to be used for computing participants' payment. To ensure transparency, the experimenter (the only one with the webcam switched on) randomly drew a number, then asked the participant with the corresponding ID code (here, ID #19) to randomly choose a number between 1 and eleven and write it in the Zoom chat to all participants. The number written by participant #19 (here, the number 7, as highlighted with the red box), was used to select the question for the final payments.