Journal of Economic Behavior and Organization 188 (2021) 288–306

Contents lists available at ScienceDirect

Journal of Economic Behavior and Organization

journal homepage: www.elsevier.com/locate/jebo

When leading by example leads to less corrupt collaboration

Rainer Michael Rilke[a,1,*], Anastasia Danilov[b,1], Ori Weisel[c], Shaul Shalvi[d], Bernd Irlenbusch[e]

[a] WHU - Otto Beisheim School of Management, Vallendar, Burgplatz 2, 51679 Vallendar, Germany
[b] Humboldt-Universität zu Berlin, School of Business and Economics, Unter den Linden 6, 10099 Berlin, Germany
[c] Coller School of Management, Tel Aviv University, Tel Aviv, 6997801, Israel
[d] Amsterdam School of Economics, University of Amsterdam, Roetersstraat 11, 1018WB Amsterdam, The Netherlands
[e] University of Cologne, Faculty of Management, Economics, and Social Sciences, Department of Corporate Development, Albertus-Magnus-Platz, 50923 Cologne, Germany

ARTICLE INFO

Article history:
Received 3 May 2020
Revised 1 April 2021
Accepted 5 May 2021

JEL classification:
D01
C90
C92

Keywords:
Honesty
Leadership
Image concerns

ABSTRACT

We contribute to the pressing question of how organizational design influences corporate wrongdoing by studying different decision structures — simultaneous vs. sequential — in experimental coordination games. Participants can report private information honestly, or lie to increase their own, as well as the group's, payoff. In simultaneous decision structures, all group members report at the same time, without information about the reports of others, whereas in sequential decision structures there is a first mover who decides first. We find that the presence of a first mover decreases dishonesty levels in repeated interactions (but not in one-shot settings). We argue that this effect is primarily driven by image concerns of decision leaders.

1. Introduction

Observing a large number of recent corporate scandals leads to the impression that employees are particularly prone to unethical behavior when acting together.[2] Consider situations where members of an organization can engage in acts of collaborative manipulations, such as profit overstatement in annual reports as happened at the supermarket chain Tesco in 2014, opening fake cross-selling accounts at the financial service company Wells Fargo between 2011 and 2015, or installing defeat devices by Volkswagen in 2008 and Audi in 2019. Different explanations for such "corrupt collaborations" have been suggested. For example, employees may use the goal of increasing the organizations' benefits as an excuse for wrongdoing or may feel that the involvement of colleagues makes wrongdoing more socially appropriate.[3]

* Corresponding author.
E-mail addresses: rainer.rilke@whu.edu (R.M. Rilke), anastasia.danilov@hu-berlin.de (A. Danilov), oriw@tauex.tau.ac.il (O. Weisel), s.shalvi@uva.nl (S. Shalvi), bernd.irlenbusch@uni-koeln.de (B. Irlenbusch).

[1] Equal contribution

[2] Several experiments documented that group interaction leads to more unethical behavior than decisions taken by individuals (e.g., Kocher et al., 2018; Weisel and Shalvi, 2015; Conrads et al., 2013; Danilov et al., 2013; Wiltermuth, 2011).

[3] According to EY's 14th Global Fraud Survey 2016, one of the most frequently used justifications for immoral behavior is to help an organization, a team, or a group to get the "right" result (EY, 2016). Other studies investigate other motives, e.g., free-riding or pivotality, for unethical conduct (Barr and Michailidou, 2017; Bartling and Fischbacher, 2012; Falk et al., 2020; Gross et al., 2018).

https://doi.org/10.1016/j.jebo.2021.05.007
0167-2681/

We aim at contributing to the pressing research agenda of how organizational design influences corporate wrongdoing by focusing on an essential design element — the time structure of decision processes within the organization — and investigating its role in corrupt collaboration.

The corporate scandals mentioned above suggest that coordination is a highly important element in many cases of corrupt collaboration. When members of the group act in concert and coordinate well, wrongdoing is likely to remain undiscovered and results in higher benefits for the involved parties, but potentially also in more damage to the organization as a whole, or to society. Consider, for example, a situation where a group of employees has an opportunity to engage in accounting manipulations. Such manipulations may involve sales figures; but may also require that employees coordinate on a specific manipulation scheme that will unethically increase profits. Many corporate scandals require mutual agreement and coordination on particular unethical procedures or conduct. For example, when the management of a carmaker decides on a specific defeat device, all involved individuals have to agree and coordinate on its technical implementation. Alternatively, think of situations which require employees to coordinate by remaining silent about unethical practices in the organization, or to coordinate on a particular cover-up story to hide some wrongdoing.

Corrupt collaboration can occur in decision processes with different time structures. There are situations where employees in a group are required to act simultaneously and there are situations in which they act in succession. The first may draft an annual report, and the others respond one after the other by signing off the report in sequence. If the first employee can propose a manipulation scheme that dishonestly inflates the group's accomplishments, and all others agree, everyone stands to profit. Coordination in such a sequential setting might be easier than in a simultaneous setting, since in the sequential setting there is, for example, a 'first mover' who decides first, and whose action is visible to others. On the other hand, coordination might be more difficult in a sequential setting than in a simultaneous one since in the sequential setting, for example, nobody might want to take the lead in wrongdoing.[4]

Another element of organizational design that we consider is whether the group interaction takes place just once or repeatedly. By doing so we capture the difference between decision processes that occur very rarely (in the extreme case, only once), for example by ad-hoc groups that are specifically created for a given task, and decision processes that occur regularly among members of on-going groups.

We approach our research questions by experimentally investigating behavior in coordination games involving three participants. We use a 2 (one-shot vs. repeated interaction) x 3 (decision time structures) experimental design, with the number of interactions manipulated within subjects and the decision time structure between subjects.

We operationalize corrupt collaboration as coordinated dishonesty. In all treatments group members individually report one of three possible outcomes of private die-rolls. A payoff (equal to all group members) is generated if and only if all group members coordinate on reporting the same number, and it increases in the reported number. Subjects thus face a trade-off between being honest, i.e., reporting the true outcomes of the die-rolls they observe, and increasing their payoffs by dishonestly reporting high outcomes and/or matching other's reports.

In different treatments, we vary the time structure in which decisions are taken: In the *Simultaneous* treatment, all members of the group report the outcome of their private die-roll simultaneously, i.e., before they observe their peers' reports. In the *Partially Sequential* and *Fully Sequential* treatments, one member of the group — the "first mover" — reports first. In *Partially Sequential*, the two followers observe the first mover's report and then report their own outcomes simultaneously. In *Fully Sequential*, a second mover observes the first mover's report and then reports his die-outcome to the third group member, who in turn observes both reports before reporting herself. We find substantial levels of corrupt collaboration in the one-shot interactions which do not significantly differ between different decision structures. When the interaction is repeated, subjects lie most when they report simultaneously, as compared to the two sequential treatments (*Partially Sequential* and *Fully Sequential*). In repeated interactions in the sequential settings, first movers set an honest example relatively frequently by not reporting the highest possible number.

Our results are important for two main reasons. First, prior research on dishonesty in groups ignored the role of decision structures. Second, research on coordination typically highlights that sequential decision structures make it easier to coordinate. Our results reveal that sequential decision structures can have another benefit: The reduction of coordinated dishonesty.

Our study directly contributes to two streams of literature. One that focuses on how leading by example influences cooperation (Güth et al., 2007; Gächter et al., 2012), and another that investigates how collaboration can breed dishonest behavior in groups (Weisel and Shalvi, 2015; Wiltermuth, 2011).

We combine major features of this research and aim to contribute to a better understanding of how leading by example can help to reduce corrupt collaboration. We use a framework in which first movers can lead by providing an example, i.e., by taking binding decisions and apply this to the existing corrupt collaboration paradigms, where individuals in groups can collaborate and increase each other's payoffs by behaving dishonestly.

[4] We focus on the organizational design of the time structure of decision processes. Note that the time structure of decision processes is not necessarily identical to the hierarchical structure of an organization; subordinates may take decisions earlier than their superiors or vice versa.

First, our study builds on experiments by Güth et al. (2007) and Gächter et al. (2012), where leaders influence followers by leading by example when contributing to public goods.[5] The results from these experiments show that leading by example tends to have a positive influence on cooperation behavior.

Prior research established the role of incentives for the positive effect of leading by example to unfold. For example, Güth et al. (2007) finds that contribution levels are higher when leaders have the power to exclude followers. Moreover, when a leader's behavior can be attributed to a different motive than own profit-maximization, followers are more likely to follow (Glöckner et al., 2011). Similarly, Cappelen et al. (2016) show that when leaders earn more than followers, contribution levels are lower. When given the option to choose incentives to cooperate for the group, leaders tend to favor reward incentives rather than punishment incentives (Gürerk et al., 2009). Moreover, when followers are of similar identity (Nosenzo et al., 2015) and when they can endogenously vote to select their leader, they tend to contribute more (Rivas and Sutter, 2011; Dannenberg, 2015).

All these studies are linked by their focus on cooperative behavior in the well-established multi-period public goods setting. None of these studies, however, have systematically investigated the underlying motives of leaders. In public goods experiments on leading by example, first movers can set a good example by doing the "right" thing, i.e., contributing to a team outcome that can increase the followers' payoff. To resemble ethical dilemmas of leaders outside the laboratory in our experiment first movers face a trade-off between being honest and not maximizing their own and the other group members' payoffs or being dishonest and maximizing their own and the other group members' payoffs.

Second, our study contributes to the literature on collaborative corruption. This stream of research shows how group interactions can foster dishonest behavior (Weisel and Shalvi, 2015; Conrads et al., 2013). Experiments on collaborative corruption have used different decision structures to operationalize collaboration. While Weisel and Shalvi (2015) use a sequential decision structure, where one player moves first, and the other follows, Conrads et al. (2013) implement a simultaneous move structure.

Further experiments investigate how communication (Kocher et al., 2018) and leading by word (D'Adda et al., 2017) influence dishonesty in groups. While in the first paper, group members can communicate before making decisions simultaneously, in the latter, a leader can send a message to the followers, who then themselves have to decide simultaneously.[6] In contrast to this literature, we contribute by abstracting away from the influence of communication (Eisenkopf, 2020) and focus on leading by example. We consider both simultaneous and sequential decision structures and systematically contrast one-shot and repeated interactions to illuminate the underlying motives that drive leading by examples in ethical dilemmas.

The rest of the paper is organized as follows: In Section 2 we describe the experimental design, hypotheses, and procedural details. In Section 3 we present the results. In Section 4 we provide an additional analysis of the potential mechanisms underlying our results. Section 5 concludes and discusses implications.

2. Experimental design

2.1. Task

Our experiment employs a novel extension of the die-rolling paradigm (Fischbacher and Föllmi-Heusi, 2013; Weisel and Shalvi, 2015). The experimental task for participant i is to privately roll a fair six-sided die, observe the outcome o_i, and report the outcome r_i on the computer screen. The die sides are labeled with numbers one, two, and three (each number appears on two sides), such that $o_i \in \{1, 2, 3\}$ and with $P(o_i = k) = 1/3$ for $k \in \{1, 2, 3\}$. Importantly, the report r_i does not need to be equal to the actual outcome o_i.

2.2. Stages

The experiment consists of three stages. Subjects learn the details of each stage after the completion of the previous one. One stage is randomly selected at the end of the experiment and paid out in cash. Stage 1 ("Individual - One-shot") includes one round in which subjects roll the die, report the outcome, and receive an individual bonus of $3 \cdot (r_i - 1)$ €. Afterwards, they state their beliefs about the behavior of others by answering the following questions: "How many of the other subjects in this room entered the number 1 (2,3) as their outcome?" Subjects could enter any number between 0 and the number of other people in the session.[7] Behavior in the individual die rolling game and their belief about subjects from Stage 1 will later serve as control variables in the group interaction. As prior studies indicate (e.g., D'Adda et al., 2017) leaders intrinsic level of honesty can influence followers. Thus, we included this stage at the very beginning to keep this measure clean from feedback about other participants behavior.

[5] See Eichenseer (2019) for a review of this literature.

[6] We are also aware of the literature on how different situational and personal characteristics influence leaders' pro-social behavior (e.g., Bendahan et al., 2015). Our aim is, however, to focus on the effect of being in the role of a first mover and the underlying motives for dishonest behavior.

[7] Belief elicitation was incentivized. Subjects received an additional 1 € if their estimation deviated by less than 1 unit from the mean of other subjects actual reports. Feedback regarding the correct estimation of their beliefs was given only at the end of the experiment. Per experimental session we had between 18 and 21 subjects.

In Stage 2 (“Group - One-Shot”) subjects are randomly assigned to groups of three (triads), which remain constant for the remainder of the experiment.[8] Again, every subject has her own die that he has to roll, observe the outcome, and report it on the screen. Payoffs depend on the reports; subjects receive a positive payoff if and only if all group members report the same number, i.e., report a triple.[9] Each group member receives 3 € when triple 1s are reported, 6 € for triple 2s, 9 € for triple 3s, and 0 € otherwise. After all subjects completed Stage 2, they receive feedback about the reported individual numbers and the payoffs in their group. Since the die roll is private, subjects do not receive feedback about the outcome of the others’ die rolls.

In Stage 3 (”Group - Repeated”), subjects interact in the same groups of three (triad), with the same procedures as in Stage 2, for 30 consecutive periods. The payoff in each period is 0.10 € if all group members report 1, 0.20 € if all group members report 2, 0.30 € if all group members report 3, and 0 € otherwise. After each period, subjects see an overview of the reported individual numbers and payments achieved in their group (and not the actual outcome of the others’ die rolls). Since the payoff in each of the 30 periods in Stage 3 is exactly $\frac{1}{30}$ of the payoff in Stage 2, the expected payoff assuming honesty ($\frac{2}{3}$ €), as well the minimum (0 €) and maximum (9 €) payoff in Stage 2 are identical to those in Stage 3.

Designing an experiment to compare one-shot interactions (Stage 2) with repeated interactions (Stage 3) is not trivial, because it is not possible to keep constant both (i) the probability of a given decision to be payoff-relevant, and (ii) the payoff associated with each decision (if it turns out to be relevant), while keeping the expected value of the one-shot and the repeated interactions identical. We chose to keep the probability of each decision being payoff-relevant the same in Stage 2 and Stage 3. Since the decision in the one-shot interaction in Stage 2 is payoff-relevant with certainty (because there is only one decision), each decision in the repeated interaction in Stage 3 is payoff-relevant as well, but the potential payoff from each decision was 1/30 of the payoff in Stage 2 (because there were 30 periods in Stage 3). This feature of the design (i.e., paying for all periods in Stage 3, but reducing the payoff to 1/30 compared to Stage 2) is constant in all of our treatments, and our main comparisons are indeed between treatments.

The three stages allow us to measure several dimensions of lying behavior. Behavior in Stage 1 is a proxy for the individual lying tendency when consequences of lying only affect the decision-maker. Behavior in Stage 2 informs us about the lying tendency in the context of a one-shot group interaction. Stage 3 extends the setting of Stage 2 by repeating the interaction. Controlling for Stage 1 behavior in our regressions allows us to make statements conditional on subjects’ inclination towards individual lying behavior. This is important in our setting, where we have no other indicator of individual subjects’ honesty. Comparing Stage 2 and Stage 3 behavior allows us to make inferences about the importance of repeated interactions.

In principle one could assume that there is a possibility of learning (e.g., about preferences of others) in Stage 2 which potentially influences behavior (of first movers) in Stage 3. Although we cannot unambiguously reject this possibility, we doubt whether first movers are able to learn about the truth-telling preferences of second and third-movers from their behavior in the single one-shot interaction in Stage 2. Because any combination of reports in Stage 2 is rather likely to occur, it is hard to make strong inferences about truth-telling preferences. In other words, whatever the reports in Stage 2, it is hard to conclude whether they are truthful or not.

2.3. Treatments

To investigate how different decision structures influence dishonest coordination in groups, triads are randomized into one of three treatments: *Simultaneous, Partially Sequential*, and *Fully Sequential*. Fig. 1 provides an overview. Stage 1 is identical in all treatments. Stages 2 and 3 differ between treatments in the following features: In the *Simultaneous* treatment (our baseline) all three group members *simultaneously* roll their dice and report their outcome without knowing the reports of the other triad members.[10] In *Partially Sequential*, one member of each triad is chosen randomly and assigned to the role of the first mover, and has to roll and report the die-roll outcome before the other two group members do so. The other two players — the second movers — are informed of the first-mover’s report before they (simultaneously) roll and report themselves without knowing what the respective other second mover reports. In *Fully Sequential*, the first mover rolls and reports to the second mover, who is also exogenously selected in each group. The second mover rolls and reports to the third mover. Finally, the third mover is informed about the reports of the first and second movers and rolls and reports herself. In Stage 2, the respective setup is played one-shot, while in Stage 3 the respective interaction, with the same group and decision structure, is repeated for 30 periods. Subjects kept their role in the decision structure between Stage 2 and Stage 3.

[8] To avoid effects of framing or context, we keep the role descriptions neutral in the instructions by calling them the “first”, “second”, and “third” group member, respectively. In the paper, we will refer to them as first, second, or third movers, respectively.

[9] We used a die with only three possible outcomes because we wanted to have a reasonably high probability of obtaining a triple. With a conventional six-sided die the probability is $6 \cdot (\frac{1}{6})^3 = 0.0278$, as compared to $3 \cdot (\frac{1}{3})^3 = 0.11$ when there are three possible outcomes.

[10] For ease of exposition we sometimes denote all three subjects in *Simultaneous* as first movers. Note that in *Simultaneous* there are no second and third movers.

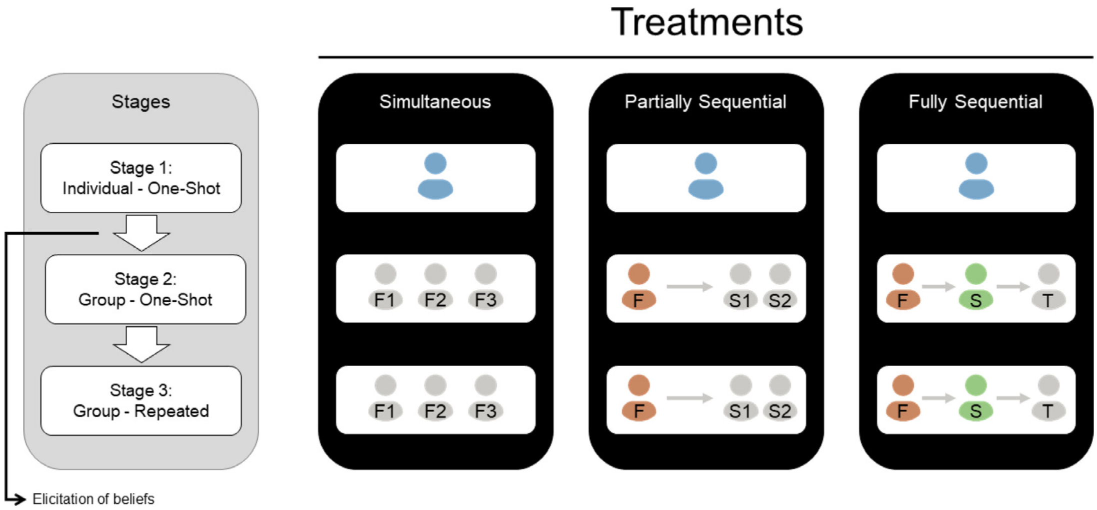


Fig. 1. Stages and treatments. The figure shows the arrangement of stages for each of our three treatments *Simultaneous, Partially Sequential*, and *Fully Sequential*. For illustration purposes, we label first movers with F, second movers with S, and third movers with T.

2.4. Procedures

A total of 369 subjects (48% female, average age: 23.6 years) were recruited via ORSEE (Greiner, 2015) to take part in the experiment. We ran 18 sessions with up to 21 subjects each in the Laboratory of Experimental Economics at a University in Germany. To avoid session effects, we used a within-session and between-subject design, meaning that all three treatments were conducted simultaneously in each session. Subjects sat in private cubicles that were equipped with a computer and a die, respectively. To ensure maximal privacy regarding the outcome of the die roll, dice were placed in lidded cups. The lid had a hole so that only the person sitting in the cubicle could observe the die-roll outcome. All subjects received the same paper instructions that were distributed at a stage-by-stage basis. The treatment specific informations appeared directly on the screen. The translated version of the instructions is provided in the Appendix.[11]

To make sure subjects realized that the dice were fair, they were instructed to roll a few times and observe the outcomes at the very beginning of the experiment.

To ensure that subjects understood the payoff function, an on-screen payoff calculator appeared at the beginning of Stages 2 and 3. Here subjects could simulate a group of three and see the earnings for different combinations. At the end of the experiment (but before earnings were revealed), subjects were asked to complete a short socio-demographic survey. Each subject received a show-up fee of 2.50 € and a payment for one randomly selected stage that consisted of earnings based on her decisions and an endowment of 6 €.[12] On average, each subject earned 16.15 €. A session lasted roughly one hour.

2.5. Hypotheses

As a first benchmark, we consider behavior assuming myopic players that are purely money-maximizing, with zero costs of lying. For such players, the setting in Stage 2, as well as each period of Stage 3, is a simple 3x3 coordination game with three pure strategy Nash equilibria. The equilibrium where every subject reports a 3, however, is payoff-dominant, so in all treatments and both Stages 2 and 3, we expect such myopic subjects to always report 3.[13]

As highlighted by previous research, moral inclinations and the cost of lying play a significant role in shaping behavior, and often prevent people from engaging in money-maximizing over-reporting (Abeler et al., 2019; 2014; Fischbacher and Föllmi-Heusi, 2013; Gneezy et al., 2018; Gibson et al., 2013). If the costs of lying are prohibitively high (i.e., there is a very strong preference for honesty), such that no potential gain can offset them, then players will always report truthfully, regardless of the treatment and whether the interaction takes place just once or repeatedly. Thus, for homogeneous popu-

[11] The instructions made it clear that any money saved in this particular experiment could be used for other research projects by the experimenters. This note hinted at the possible negative external effects of lying. Some participants in our experiment might not perceive this externality as being severe or immoral and do not consider coordination on the highest outcome as unethical. Imposing additional externalities, however, by (not) donating money to a charity does not seem to reduce lying. For example, comparing the lying rates from the aligned outcome treatment in Weisel and Shalvi (2015) to the lying rates from Gross et al. (2018) who use a similar setup with a suffering charity does not yield significant differences in dishonesty.

[12] We add this additional endowment to insure that the average hourly pay in case of honest reporting matches at least 8.5 €. Subjects learned about the endowment at the beginning of each stage.

[13] Similarly, a player in Stage 1 will always report 3.

lations consisting of either payoff-maximizing players or fully honest players, the behavior should be consistent across our treatments.[14]

In addition to the cost of lying, people may have a preference for setting a good example that might influence whether other subjects lie or not. In our setting subjects might have a preference for behaving honestly if there are followers who might be influenced by their action. Such a preference can come in two flavors: it can be intrinsic or can be image-driven. Each of these options leads to a different hypothesis, as outlined in the following paragraphs.

A preference for leading by a good example is *intrinsic* if a person experiences a utility gain when setting a good example by being honest. Consider, for example, a subject who wants to earn money and is additionally motivated to set a good example if she has the opportunity to do so. Such a subject is more likely to behave honestly as a first mover in the sequential treatments (*Fully Sequential* and *Partially Sequential*), where there is an opportunity to set a good example, than in *Simultaneous*, where no such opportunity exists. A strong intrinsic preference for setting a good example would lead a subject with followers to behave honestly regardless of whether they are in a one-shot or a repeated setting.

Hypothesis 1 (Intrinsic motivation to lead by setting a good example)**.** First movers report more honestly in *Fully Sequential* and *Partially Sequential* compared to subjects in *Simultaneous* (in one-shot and repeated settings).

An alternative rationale also suggests that first movers lie less in the sequential settings compared to *Simultaneous*: Honesty - in terms of forgone payoffs - is less costly for first movers in *Partially Sequential* and *Fully Sequential* than in *Simultaneous*. The rationale behind this explanation is the following. In *Simultaneous* there is one prominent focal point on which group members can coordinate, i.e., the payoff-dominant equilibrium when all players choose 3. A player that reports a number other than 3 effectively ruins the group's chance to coordinate, leaving the group a fairly small chance of reporting a triple, even if the other two members of the group are willing to lie to increase profits. In contrast, in the treatments with sequential decision structures, the report of the first mover becomes a visible focal point on which her followers can subsequently coordinate. If the other group members are willing to lie, the group will still report a profitable triple (even if not the most profitable triple 3's). Thus, assuming that other players are willing to dishonestly increase payoffs, foregone payoffs associated with being honest are smaller for first movers in *Partially Sequential* and *Fully Sequential* compared to players in *Simultaneous*. To the degree that this reasoning holds, first movers are honest in *Partially Sequential* and *Fully Sequential* merely because it is less costly than in *Simultaneous*, and not because they care about their good image as an honest decision leader.

Fischbacher and Föllmi-Heusi (2013) include a treatment where subjects report a number (which determined their payoff) without rolling a die. In contrast to what one would expect from a solely self-interested decision-maker, not all subjects report the highest possible payoff. The authors conclude that a desire not to appear greedy may explain this finding. Barr and Michailidou (2017), however, who study a coordination game with simultaneous reporting, find that 97% of all players report the highest possible outcome in a control condition with a payoff-dominant equilibrium (Harsanyi and Selten, 1988) and zero off-diagonal payoffs. Taken together, these results suggest that if a preference for not appearing greedy plays a role in our group setting, it is probably minor.

First movers' ability to influence the decisions of others may create a feeling of responsibility which leads to an additional rationale, see, for example, the literature on responsibility and guilt sharing in group decisions as discussed by Feess et al. (2020). It can operate through at least two distinct channels: On the one hand, players in *Simultaneous* feel less responsible for the groups' outcome because their report cannot influence any other player. On the other hand first movers in the sequential treatments can "nudge" the second and third movers to report a higher number and, thus, feel more responsible for the occurred corruption (compared to the subjects in *Simultaneous*). In this sense, a feeling of responsibility could lead to more honesty among the first-movers' in the sequential treatments.

Evidence from prior experiments on dishonesty highlights that in addition to an intrinsic preference for behaving honestly, image concerns influence people's decision to lie. We argue that individuals experience image concerns as an intrinsic cost when they could be perceived as liars (Abeler et al., 2019; Gneezy et al., 2018; Khalmetski and Sliwka, 2019).

Although in all our settings the three players are equally pivotal[15], the decision structures that we use - sequential or simultaneous - may affect players with image concerns in different ways. In the sequential treatments the first mover's decision becomes ex-ante visible to the second and third movers. In contrast, in *Simultaneous* all three players can share the 'social image' burden by reporting high numbers simultaneously. Therefore, they might suffer less from harming their social image when reporting high numbers than first movers with social-image in the sequential treatments. This aspect might

[14] One could also assume different compositions of honest and dishonest types. It can be shown that coordination in *Partially* and *Fully Sequential* is expected to be higher as compared to *Simultaneous*. This is most apparent when considering a group that is composed of a single honest player and two dishonest players. In *Simultaneous* the probability of coordination in such a group is 1/3 (the dishonest players report 3, the honest one does so with probability 1/3). In *Partially* and *Fully Sequential*, if the honest player is the first mover, the probability of coordination is 1; the dishonest players will match whatever the honest one reports in the first move. If the honest player is not the first mover, the probability is 1/3. Similar reasoning shows that for any composition of honest and dishonest players, the likelihood of coordination in *Partially* and *Fully Sequential* is equal to or higher than the likelihood of coordination in *Simultaneous*.

[15] See Bartling et al. (2015) for an experiment investigating responsibility attribution where only one player is pivotal. Interestingly, Feess et al. (2020) find that unethical behavior increases with the number of pivotal players.

dominate behavior also in the repeated setting where players who move second and/or third could also have reputation concerns that develop over time.

Arguably, in our one-shot interaction in Stage 2 a first-mover barely harms his/her own social image when dishonestly reporting a high number. The reason being that from the perspective of the followers it might well be true that this high number was indeed the outcome of the die role. However, reporting high numbers repeatedly as a first mover in Stage 3 is likely to harm one's own social image, because obtaining high numbers in multiple rounds is much less likely than in a single roll.

Thus, one can expect that (i) reported outcomes of first movers in Stage 2 are higher than in Stage 3; and that (ii) in Stage 3 reported outcomes of first movers in the sequential treatments are lower than reported outcomes of players in *Simultaneous*.

Hypothesis 2 (Image driven motivation to lead by setting a good example)**.** First movers report more honestly in *Fully Sequential* and *Partially Sequential* compared to subjects in *Simultaneous* only when the interaction is repeated. In the one-shot setting honesty does not differ between the decision structures.

3. Results

The presentation of the results is organized around our main research question: How do different decision structures affect (dis)honesty? We start by briefly describing behavior in Stage 1, where participants report individually, and then proceed to the analysis of the group setting in Stage 2 (one-shot interaction) and Stage 3 (repeated interactions). In our analysis of behavior in groups, we first analyze behavior at the group level and then turn our attention to first movers, i.e., decision leaders. Our chapter concludes with an observation of the followers behavior.[16]

3.1. Individual behavior in Stage 1

Analysis of behavior in Stage 1 reveals two main findings. First, in all three treatments reports significantly exceed the expected value of honest reporting 2 (the average reports are 2.77, 2.71, and 2.71 in *Simultaneous, Partially Sequential*, and *Fully Sequential*, respectively; $p < 0.0001$ for all three; Wilcoxon Signed Rank tests, henceforth WSR).[17] Second, reported numbers do not differ between the three treatments ($p = 0.541$, Kruskal Wallis test, henceforth KW), nor do payoffs (5.32 €, 5.12 €, 5.12 €; $p = 0.541$, KW). The latter finding confirms that treatment randomization was successful.[18]

3.2. Group behavior

In this section we analyze group behavior. The main dependent variables are groups' average reports, and the degree to which groups coordinate on triples.

3.2.1. One-shot interaction (Stage 2)

Average reports in Stage 2 significantly exceed 2 (expected value of honest reporting) in all treatments (2.86 in *Simultaneous*, 2.81 in *Partially Sequential*, 2.89 in *Fully Sequential*, $p < 0.001$, WSR for each treatment separately) and significantly exceed reports in Stage 1 (2.77 in *Simultaneous*, 2.71 in *Partially Sequential*, 2.71 in *Fully Sequential*, $p \leq 0.0517$, WSR for each treatment separately).[19] Comparing reports in Stage 2 between all treatments does not reject the null hypothesis that they are identical ($p = 0.349$, KW). In total, 71% of groups report a triple in *Simultaneous*, compared to 80% in *Partially Sequential*, and 93% in *Fully Sequential* ($p = 0.038$, χ^2 test). Pairwise comparisons yield that coordination success is significantly higher in *Fully Sequential* compared to *Simultaneous* ($p = 0.01$).[20]

Result 1. (Group behavior - one-shot)**.** In Stage 2 there is no difference in dishonesty in terms of average reports between the treatments; but coordination on triples is higher in *Fully Sequential* than in *Simultaneous*.

3.2.2. Repeated interactions (Stage 3)

[16] Our main experiment comprises three treatments with three pairwise comparisons between treatments. To be conservative with hypotheses testing, we take into account multiple hypotheses testing by correcting the conventional p-value threshold of 0.05 with the Bonferroni correction to 0.05/3 = 0.016. Thus, in the following we refer to a difference between treatments as significant if the p-value is below the corrected threshold (Bonferroni, 1935).

[17] In this paper all reported p-values are two-sided. If no test is mentioned the same test is applied as the test before.

[18] After their decision in Stage 1 subjects entered their beliefs about the other participants' reports in Stage 1 of the experiment. A KW test comparing the distribution of these beliefs between the treatments does not reject the null hypothesis that they are equally distributed ($p = 0.4404$). Furthermore, there is a strong correlation between beliefs and behavior on Stage 1 (Spearman's $\rho = 0.4989$, $p < 0.0001$).

[19] This observation is in line with over-reporting tendencies reported in studies with similar group situations (e.g., Weisel and Shalvi, 2015; Soraperra et al., 2017) and in line with studies reporting higher levels of dishonesty in groups compared to individual settings (e.g., Conrads et al., 2013; Kocher et al., 2018).

[20] A comparison of *Partially Sequential* and *Simultaneous* does not reject the null hypothesis that they are equal ($p = 0.304$). Coordination success are also reflected in payoffs which averaged 6.37 €, 7.10 €, and 8.12 € in *Simultaneous, Partially Sequential*, and *Fully Sequential*, respectively. Pairwise comparisons reveal that payoffs in *Fully Sequential* are not significantly higher compared to *Simultaneous* ($p = 0.069$, Mann Whitney U Test, henceforth MWU), and that earnings in *Partially Sequential* are not significantly different than in *Fully Sequential* or *Simultaneous* ($p = 0.221$, $p = 0.524$, respectively).

Table 1
Group behavior between stages.

Independent variables	Dependent variable: Reported number $r_{i,t}$ (1)	(2)
Partially Sequential	−0.19***	−0.05
	(0.06)	(0.07)
Fully Sequential	−0.24***	0.02
	(0.06)	(0.06)
Partially Sequential x Stage 3		−0.14**
		(0.06)
Fully Sequential x Stage 3		−0.28***
		(0.06)
Stage 3	−0.08***	0.06
	(0.03)	(0.04)
Constant	3.00***	2.86***
	(0.03)	(0.04)
Observations	11,439	11,439
R^2	0.038	0.039

Notes: GLS regression (random effects). Robust standard errors clustered on the group level are in parentheses. The reference group is treatment *Simultaneous*. Significance indicators: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Now we focus on group behavior in the 30 repeated interactions of Stage 3. As in the prior stages, the average reports in all treatments exceed the prediction assuming honest reports (2.92 in *Simultaneous*, 2.73 in *Partially Sequential*, 2.67 in *Fully Sequential*, $p < 0.001$, WSR test with group averages as independent observations). Pairwise comparisons reveal that reports in *Simultaneous* are significantly higher than in *Fully Sequential* and *Partially Sequential* ($p = 0.008$; Cohen's $d = 0.76$ and $p = 0.006$; $d = 0.95$, Mann Whitney U Test, henceforth MWU), and that the difference between reports in *Fully Sequential* and *Partially Sequential* is not significant ($p = 0.396$; $d = 0.17$).[21]

Triples are reported in 82% of all periods (87% in *Simultaneous*, 79% in *Partially Sequential*, 80% in *Fully Sequential*), more than seven times than the expected 11.11% assuming honesty. In most cases, groups report triple 3's. The frequency of triple 3's is significantly higher in *Simultaneous* (86%) compared to *Partially Sequential* (70%, $p = 0.012$; $d = 0.57$) and *Fully Sequential* (67%, $p = 0.011$; $d = 0.66$).[22]

Result 2. (Group behavior - repeated interactions)**.** In Stage 3 there is less dishonesty, in terms of both average reports and coordination on triples, in *Partially Sequential* and *Fully Sequential* than in *Simultaneous*.

3.2.3.One-shot (Stage 2) vs. repeated interaction (Stage 3)

A joint analysis of group behavior in Stages 2 and 3 confirms the pattern emerging from Results 1 and 2. Model 1 in Table 1 shows that, overall, reports in *Partially Sequential* and *Fully Sequential* tend to be lower than in *Simultaneous*. The small and non-significant coefficients of the treatment dummies in Model 2, together with the negative and significant interaction terms indicate that the overall effect is driven by Stage 3.

The stage–treatment interaction is also apparent when directly comparing both stages. In *Simultaneous*, Stage 2 and Stage 3 do not significantly differ in average reported numbers ($p = 0.805$, WSR, $d = -0.25$), frequencies of triples ($p = 0.419$, $d = -0.30$), frequencies of triple 3's ($p = 0.419$, $d = -0.28$), or payoffs ($p = 0.419$, $d = -0.35$). In the sequential treatments, however, reported numbers, the frequency of triple 3's, and payoffs are higher in Stage 2 than in Stage 3 (*Partially Sequential*: $p = 0.011$, $d = 0.31$; $p = 0.07$, $d = 0.26$; $p = 0.061$, $d = 0.12$; *Fully Sequential*: $p < 0.001$, $d = 0.67$; $p < 0.001$, $d = 0.57$; $p = 0.001$, $d = 0.52$). In *Fully Sequential* there are also more triples in Stage 2 ($p = 0.001$, $d = 0.54$; *Partially Sequential*, $p = 0.164$, $d = 0.13$).

Result 3. (Group behavior - one-shot vs. repeated interactions)**.** Dishonesty levels in *Partially Sequential* and *Fully Sequential* are lower in Stage 3 than in Stage 2. In *Simultaneous*, dishonesty is similar in both stages.

3.3. First mover behavior

In this section we analyze first mover behavior. The main dependent variable is first movers' average reports. Fig. 2 provides an overview across treatments and stages for all player roles.

3.3.1. One-shot interaction (Stage 2)

[21] These treatment effects are supported by a regression analysis where we included lagged variables, a linear time-trend and behavior from Stage 1 and 2 as predictors (see Table A.3).

[22] In line with the frequency of triples, the average payoff in Stage 3 is 7.77 € in *Simultaneous*, marginally significantly more than in *Partially Sequential* (6.78 €, $p = 0.029$, $d = 0.42$; MWU) and *Fully Sequential* (6.72 €, $p = 0.043$, $d = 0.45$). A possible concern is that behavior in Stage 2 influences behavior in Stage 3, such that coordination success in Stage 2 makes coordination more likely in Stage 3. Note that such an effect, however, should work against our findings, since coordination in Stage 2 is somewhat higher in *Fully Sequential* while in Stage 3 it is higher in *Simultaneous*.

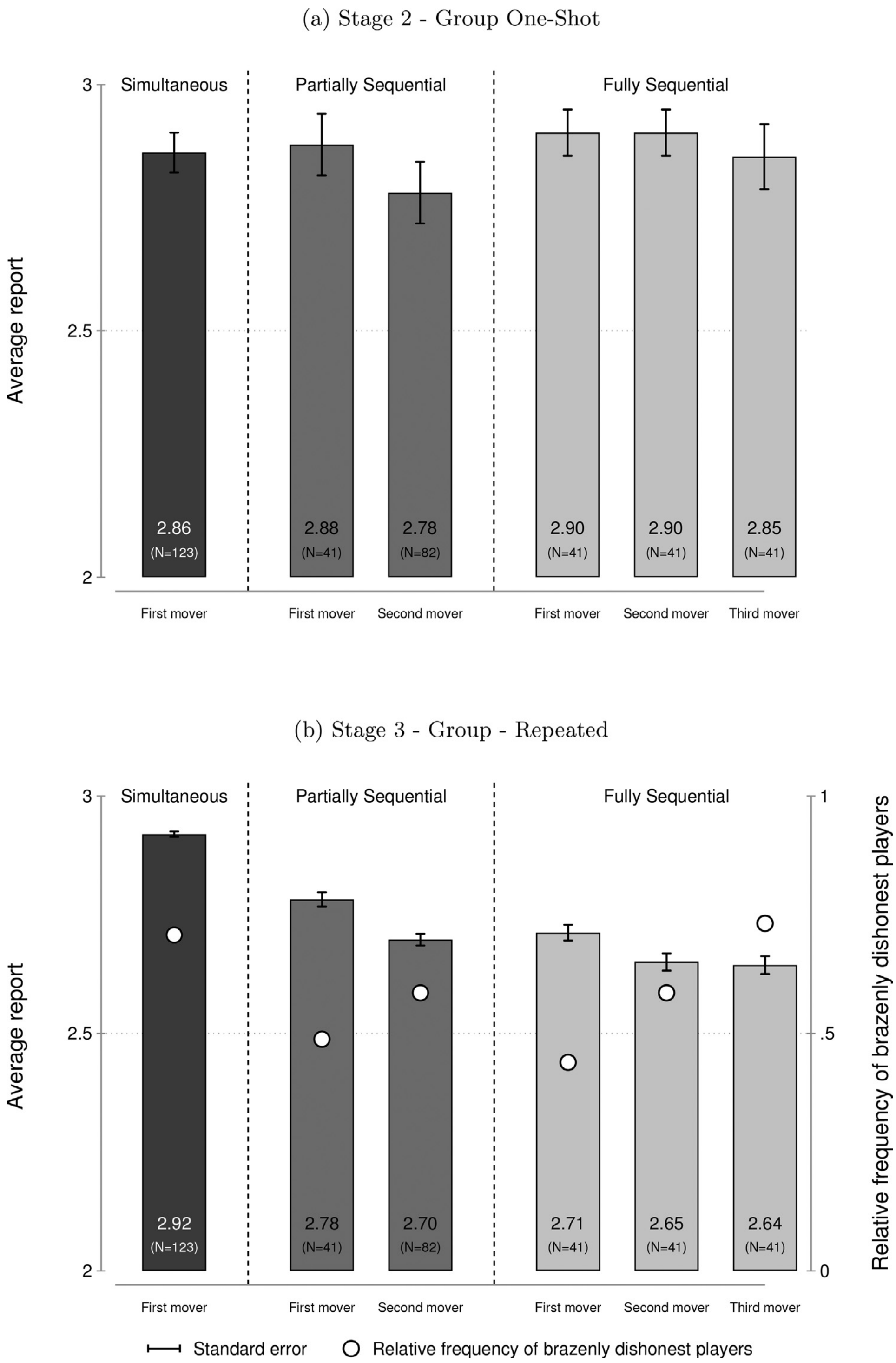


Fig. 2. Average reported numbers by first, second, and third movers. Bars show average reports by first, second, and third movers. Whiskers indicate standard errors. Dots show the relative frequencies of "brazenly dishonest" players. First movers are defined as brazen if they report a 3 in all thirty periods. Second movers are defined as brazen if their report is equal to the first mover's report in all thirty periods. Third movers are defined as brazen if their report is equal to the reports of both the first and the second movers' in all thirty periods. A supplementary analysis of brazen typs can be found in the Appendix. Histograms and scatterplots of behavior can be found in Fig. A.1 and A.2 in the Appendix.

Table 2
First mover behavior in Stage 3.

	Dependent variable: Reported number $r_{i,t}$			
Independent variables	(1)	(2)	(3)	(4)
Partially Sequential	−0.14**	−0.14***	−0.10**	−0.09*
	(0.06)	(0.05)	(0.05)	(0.05)
Fully Sequential	−0.21***	−0.21***	−0.12**	−0.13***
	(0.06)	(0.05)	(0.05)	(0.04)
Period		0.00***	0.00***	0.00***
		(0.00)	(0.00)	(0.00)
Triple 3 (Stage 2)		0.15***	0.14***	0.16***
		(0.05)	(0.05)	(0.05)
Triple 3_{t-1} (Stage 3)		0.09**	0.09**	0.07*
		(0.04)	(0.04)	(0.04)
Ind. report < 3 (Stage 1)		−0.20***	−0.08	−0.04
		(0.06)	(0.07)	(0.06)
Partially Sequential x Ind. report < 3			−0.19	−0.22
			(0.17)	(0.15)
Fully Sequential x Ind. report < 3			−0.42***	−0.37***
			(0.09)	(0.09)
Constant	2.92***	2.74***	2.73***	2.78***
	(0.02)	(0.06)	(0.06)	(0.16)
Observations	6150	5945	5945	5945
R^2	0.038	0.152	0.172	0.199
Additional controls	No	No	No	Yes

Notes: GLS regression (random effects). Robust standard errors (group level) are in parentheses. The reference group are all subjects from the *Simultaneous* treatment. Additional controls not shown in the table are session dummies, age, and female. Model 1 contains data from periods 1 to 30, while Models 2–4 contain data from periods 2 to 30. Significance indicators: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

As Fig. 2a suggests, first mover's average reports in Stage 2 are not significantly different between the treatments (2.86, 2.88, 2.90 in *Simultaneous, Partially Sequential*, and *Fully Sequential*, respectively; $p = 0.997$, KW).

Result 4. (First movers - one-shot)**.** In Stage 2 there is no difference between treatments in average first mover reports.

3.3.2. Repeated interactions (Stage 3)

Fig. 2b displays the average reported numbers for each role (first, second and third movers) and the fraction of brazenly dishonest players. The average report in *Simultaneous* (2.92) is significantly higher than first movers' average report in *Partially Sequential* (2.78; $p = 0.005$, $d = 0.51$; MWU) and *Fully Sequential* (2.71; $p = 0.0002$, $d = 0.77$). The difference between *Partially Sequential* and *Fully Sequential* is not significant ($p = 0.437$, $d = 0.20$).

Result 5. (First movers - repeated interactions)**.** In Stage 3 there is less dishonesty in terms of average first mover reports.

We test the robustness of the pattern reported Result 5 and further investigate the dynamics of first mover behavior in Table 2, which include a number of control variables ("Period", "Ind. report < 3 (Stage 1)", "Triple 3 (Stage 2)" and "Triple $3_{(t-1)}$ (Stage 3)"). The treatment dummies in Model 1 remain significant and almost unchanged with the addition of these variables.

"Period" captures the possible linear time trend over the 30 periods. It is significant and positive, indicating that first movers tend to report higher numbers with time. "Triple 3 (Stage 2)" and "Triple $3_{(t-1)}$ (Stage 3)" are dummy variables that capture the success of the first mover's group in previous stages/periods. The former equals 1 if the group coordinated on triple 3's in Stage 2 (the one-shot group interaction); the latter equals 1 if the group coordinated on triple 3's in the previous period of the repeated interaction in Stage 3. The positive and significant coefficients of both variables imply that a group's previous success in coordinating on triple 3's is positively correlated with the first mover reporting higher numbers (i.e., being less honest) in subsequent periods of Stage 3. The "Ind. report < 3 (Stage 1)" dummy equals 1 when the respective first mover has not reported a 3 in Stage 1 of the experiment, i.e., the player is likely to have reported honestly when making her individual decision. The negative and significant coefficient in Model 2 highlights that individuals who are likely to be honest in Stage 1 tend to report lower numbers (i.e., to be more honest) as first movers in Stage 3.

In Models 3 and 4 we check whether the latter results—the tendency of honest individuals in Stage 1 to be relatively honest in Stage 3—is constant across treatments, by adding two interaction terms between the treatment dummies and the "Stage 1: Ind. report < 3" variable. "Stage 1: Ind. report < 3", now capturing the effect of not reporting 3 in *Simultaneous* is smaller and no longer significant, indicating that the effect found in Model 2 is mainly due to the sequential treatments. Indeed, both interaction variables are negative. Only the interaction with *Fully Sequential* is significant, however, showing that the link between behavior in Stage 1 and Stage 3 is strongest in *Fully Sequential*. The treatment dummies, now capturing the treatment effects on first movers who reported 3 in Stage 1, are smaller than in model 2, but remain negative and significant, indicating that first movers who reported 3 in Stage 1 also report significantly lower numbers in the two sequential treatments compared to the *Simultaneous* treatment.

Table 3
First mover behavior between stages.

Independent variables	Dependent variable: Reported number $r_{i,t}$ (1)	(2)
Partially Sequential	−0.13**	0.02
	(0.06)	(0.07)
Fully Sequential	−0.20***	0.04
	(0.06)	(0.06)
Partially Sequential x Stage 3		−0.15***
		(0.06)
Fully Sequential x Stage 3		−0.25***
		(0.07)
Stage 3	−0.02	0.06
	(0.03)	(0.04)
Constant	2.94***	2.86***
	(0.03)	(0.04)
Observations	6355	6355
R^2	0.036	0.037

Notes: GLS regression (random effects). Robust standard errors (group level) are in parentheses. The reference group are all subjects from the *Simultaneous* treatment. Significance indicators: $^{***}p < 0.01$,** $p < 0.05$,* $p < 0.1$.

3.3.3. One-shot (Stage 2) vs. repeated interaction (Stage 3)

Similar to the analysis of group behavior above (Results 1, 2, and 3), a joint analysis of first mover behavior in Stages 2 and 3 confirms the pattern emerging from Results 4 and 5 (see also Fig. 2). Model 1 in Table 3 shows that, overall, reports in *Partially Sequential* and *Fully Sequential* tend to be lower than in *Simultaneous*. The small and non-significant coefficients of the treatment dummies in Model 2, together with the negative and significant interaction terms indicate that the overall effect is driven by Stage 3. The average reports of first movers displayed in Fig. 2 also support the interaction between treatments and stages. While in *Simultaneous* first movers' average reports are marginally significantly higher in Stage 3 compared to Stage 2 ($p = 0.035$, $d = 0.24$, WSR), in the two sequential treatments first movers' reports are significantly lower in Stage 3 (*Partially Sequential*: $p = 0.003$, $d = 0.33$; *Fully Sequential*: $p = 0.001$, $d = 0.53$).[23]

Result 6. (First movers - one-shot vs. repeated interactions)**.** First mover reports in *Partially Sequential* and *Fully Sequential* are lower in Stage 3 than in Stage 2; in *Simultaneous* first mover reports are higher in Stage 3.

3.4. Follower behavior

Now we analyze second and third mover's reported numbers.

3.4.1. One-shot interaction (Stage 2)

Compared to players in *Simultaneous* second (and third) mover reports in *Partially Sequential* and *Fully Sequential* are not significantly different (*Partially Sequential*: $p = 0.2777$, $d = 0.16$; *Fully Sequential*: second mover $p = 0.8784$, $d = 0.097$; third mover: $p = 0.8219$, $d = 0.01$).

Second mover reports in *Partially Sequential* and *Fully Sequential* are not significantly different from first mover reports, respectively ($p > 0.4352$; $d \leq 0.19$). Third mover in *Fully Sequential* do not report significantly lower numbers compared to first and second mover ($p = 0.7698$, $d = 0.13$).[24]

Result 7. (Follower behavior - one-shot)**.** In Stage 2 there is no difference between treatments in average follower reports.

3.4.2. Repeated interaction (Stage 3)

The average reports in *Simultaneous* (2.92) are significantly higher than second movers' average report in *Partially Sequential* (2.70; $p = 0.0011$, MWU, $d = 0.85$) and *Fully Sequential* (2.65; $p = 0.0026$, $d = 0.92$). Similarly, the average reports in *Simultaneous* are significantly higher than third movers' average report (2.64; $p = 0.0017$, $d = 0.98$).

Follower reports in *Partially Sequential* and *Fully Sequential* tend to be lower compared to first mover reports of the respective treatment, but fail to reach statistical significance. The difference between first and second mover reports in *Partially Sequential* is not significantly different ($p = 0.1016$, $d = 0.25$). Likewise, average reports of second (third) movers in

[23] Note, that our data suggest that learning in Stage 2 does not drive behavior in Stage 3. First, dishonesty is increasing with the degree of sequentiality in Stage 2 (for example, the number of triples is higher in Fully/Partially Sequential compared to Simultaneous), but decreasing with the degree of sequentiality in Stage 3 (for example, the number of triples is higher in Simultaneous compared to Fully/Partially Sequential). This goes against the idea that learning about other group members' behavior in Stage 2 encourages similar behavior in Stage 3. Second, the estimated regression coefficients of the treatment dummies remain essentially the same when we control for coordination success in Stage 2 (see, for example, Table 2). Third, average reported numbers in Stage 2 in the sequential treatments are not systematically lower than those in the simultaneous treatment.

[24] A regression analysis additionally controlling for first (second) mover behavior does not change the results.

Table 4
Follower behavior between stages.

	Dependent variable: Reported number $r_{i,t}$			
	Second mover		Third mover	
Independent variables	(1)	(2)	(3)	(4)
Partially Sequential	0.22***	0.08		
	(0.06)	(0.07)		
Fully Sequential	0.26***	0.04	0.27***	0.01
	(0.06)	(0.06)	(0.06)	(0.08)
Partially Sequential x Stage 3		0.14**		
		(0.06)		
Fully Sequential x Stage 3		0.31***		0.27***
		(0.07)		(0.08)
Stage 3	0.04	0.06	0.01	0.06
	(0.03)	(0.04)	(0.03)	(0.04)
Constant	2.96***	2.86***	2.93***	2.86***
	(0.03)	(0.04)	(0.04)	(0.04)
Observations	7626	7626	5084	5084
Number of sid	246	246	164	164
R^2	0.052	0.053	0.065	0.067

Notes: GLS regression (random effects). Robust standard errors (group level) are in parentheses. The reference group are all subjects from the *Simultaneous* treatment. Significance indicators: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Fully Sequential are not significantly different compared to the first movers of the respective treatment ($p = 0.3513$, $d = 0.17$; $p = 0.2883$, $d = 0.19$).

Result 8. (Follower behavior - repeated)**.** In Stage 3 average follower reports in *Partially Sequential* and *Fully Sequential* are lower than in *Simultaneous*.

3.4.3. One-shot (Stage 2) vs. repeated interaction (Stage 3)

Regression models in Table 4 show that, overall follower reports in *Partially Sequential* and *Fully Sequential* tend to be lower than in *Simultaneous*. The small and non-significant coefficients of the treatment dummies in Model 2 (second movers) and Model 4 (third movers), together with the negative and significant interaction terms indicate that the overall effect is driven by Stage 3.

Result 9. (Follower - one-shot vs. repeated interactions)**.** Follower reports in *Partially Sequential* and *Fully Sequential* are lower in Stage 3 than in Stage 2.

4. Excluding an alternative explanation for why decision leadership leads to reduced dishonesty

As mentioned in the hypotheses section there is an alternative explanation why decision leadership leads to reduced dishonesty in the sequential treatments compared to *Simultaneous*: Honesty - in terms of forgone payoffs - is less costly for first movers in *Partially Sequential* and *Fully Sequential* than in *Simultaneous*. Note that this line of reasoning would suggest an analogous difference in reports not only in Stage 3 between the *Simultaneous* and the sequential treatments but also in Stage 2. The latter, however, we do not observe.

To investigate this rationale, we designed an additional treatment: *Fully Sequential 3*. In this treatment, Stage 1, as well as the decision structures in Stages 2 and 3, are identical to *Fully Sequential*. The only difference is the payoff structure: a payoff is generated only by triple 3's (9 € in Stage 2 and 0.30 € in each period of Stage 3), whereas the payoff from triple1's or triple2's is zero. This treatment eliminates the first mover's possibility to set a focal point other than 3, such that it is no longer the case that it is relatively cheap to be honest in the sequential treatments, and in *Fully Sequential* in particular (which is a clean comparison to *Fully Sequential 3*). It follows that if first movers' relative honesty in *Fully Sequential* is driven by lower costs of truth-telling, they should be less honest in *Fully Sequential 3* than in *Fully Sequential*. However, if their honesty is driven by a desire to lead by a good example, (dis)honesty levels in *Fully Sequential* and *Fully Sequential 3* should be similar, as this motivation is relevant in both settings.

The *Fully Sequential 3* treatment was conducted with 102 subjects in 34 groups (42% female, average age: 23.17). In Stage 2 (one-shot), first movers' average reports in *Fully Sequential 3* (2.91) are very similar to those in *Simultaneous* (2.86, $p = 0.490$, $d = 0.16$, MWU) and *Fully Sequential* (2.90, $p = 0.572$, $d = 0.03$; see left panel of Fig. 3). As is apparent from Fig. 3, in Stage 3 first movers' average reports in *Fully Sequential 3* (2.70) are very similar to *Fully Sequential* (2.71, $p = 0.661$, $d = 0.03$), and significantly lower than reports in *Simultaneous* (2.92, $p = 0.0001$, $d = 0.84$, MWU).

The average Stage 3 payoff of first movers in *Fully Sequential* (6.72 EUR) is not significantly different from first movers in *Fully Sequential 3* (6.15 EUR; $p = 0.2282$, MWU test). The proportion of brazen first movers in *Fully Sequential 3* (32%)

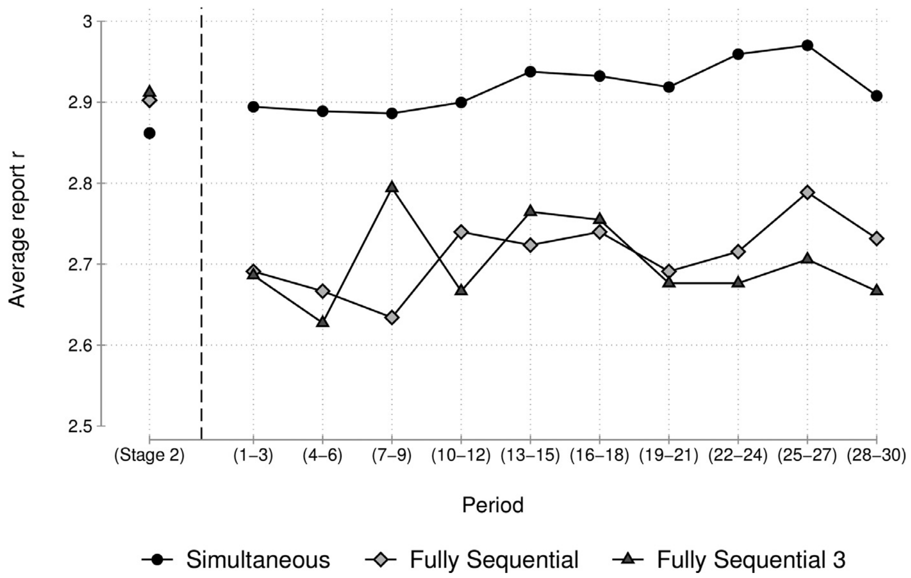


Fig. 3. Average reports of first movers in Stage 2 (one-shot) and Stage 3 (repeated).

is similar to *Fully Sequential* (44%, $p = 0.307$, $d = 0.24$, χ^2-test), and significantly lower than *Simultaneous*(70%, $p = 0.0001$, $d = 0.69$).[25] These results suggest that the reduced dishonesty in the sequential treatments is not a result of lower costs of honesty relative to *Simultaneous*.

5. Conclusion

We investigate the influence of decision leadership — an important design issue in organizations — on collaborative corruption. We examine behavior in experimental coordination games where subjects make choices either simultaneously or sequentially, i.e., one group member moves first, and others follow after observing the leaders' action. Along a second dimension, we vary whether the interaction occurs once or repeatedly.

In the one-shot interaction dishonest coordination on profitable triples seems to be easier in the sequential settings, in particular *Fully Sequential*, compared to the simultaneous decision structure. Differences in lying behavior are more pronounced when the games are played repeatedly, mainly due to first movers who lie significantly less in the sequential decision structures compared to the simultaneous ones. Our results suggest that first movers' relative reluctance to lie in repeated settings is driven by a desire to maintain an image of being honest; repeatedly reporting high numbers to second and third movers is a clear signal of dishonest behavior by the first mover.

Although our findings suggest that in repeated settings first movers (in sequential structures) tend to be more honest compared to subjects in the simultaneous structure, second and third movers are rather dishonest. In particular, when both the first and second movers reported the same number, third movers (in *Fully Sequential*) are reluctant to report a different number. This reluctance might be interpreted as a form of peer-pressure (Ahmad and Loch, 2020).

Our study contributes to the understanding of the role of decision structures in coordinated unethical conduct. In particular, we demonstrate that sequential decision structures can help reduce collaborative corruption in repeated interactions when its success relies on perfect coordination. We argue that sequential decision-making may result in less over-reporting due to first movers' social-image concerns. Our findings suggest that in situations that are susceptible to collaborative corruption, organizations may benefit from limiting simultaneous (leaderless) decision structures. Presumably, such decision structures are more common in flatter organizations (Rajan and Wulf, 2006), which tend to be more and more common in firms.

Further, organizations should aim at making individual decisions more visible to other employees and followers. Since first movers tend to avoid initiating corrupt collaboration repeatedly, making their choices more visible to followers can trigger image concerns that may promote honesty. Besides reviewing decision-making structures in organizations, future research should identify ways to leverage this effect for followers as well. We are not aware of studies that investigate this possibility.

[25] Average reports of second (2.70) and third movers (2.73) in *Fully Sequential 3* are 2.70 and 2.73, respectively. They tend to be somewhat higher compared to followers reports in *Fully Sequential* but do not differ significantly.

Obviously, our experiment is not without limitations.[26] Our stylized die rolling game is only an abstraction of dishonest behavior and group interactions outside the lab. Although the stylised abstraction needs to be kept in mind when interpreting the results for the field, there is already some evidence for the external validity of lab studies on lying (e.g., Dai et al., 2018; Kröll and Rustagi, 2016; Potters and Stoop, 2016). While identifying the sources of dishonesty, fraud and corruption in groups with field data is nearly impossible, experimental lab studies allow at least some inferences about deceptive behavior. Thus, to understand underlying behavioral mechanisms laboratory experiments are a valid choice.

Another potential criticism is that individuals in groups usually can communicate, while in our experiment communication can only occur via individuals' actions. Sometimes, however, communication might be infeasible or not credible and therefore of limited influence for the behavior of individuals. This might be especially true for unethical behavior.

Further, we focus on cases where perfect coordination among individuals is required. There exist, however, situations where individual employees inflate reports that benefit the group without the need for coordination, such that individual contributions can additively lead to higher group outcomes. In additive settings it might well be that other effects occur, for example, observing low reports by rather honest first movers may result in inflated reports by second or third movers to compensate for otherwise low rewards.

Finally, decision leaders in our game are exogenously determined. However, groups often determine their leader endogenously based on previous behavior, or leaders are appointed by other stakeholders of the organization, potentially because of their particular skills. For example, Fehrler et al. (2020) show that dishonest individuals over-proportionally self-select into leadership positions. We intentionally abstract from these potential confounds in order to highlight the pure impact of the decision structure itself.

To conclude, the large number of recent corporate scandals, potentially make it necessary to rethink how organizations structure their decision processes. We hope that future research illuminates which further situational variables can attenuatethe problem of collaborative corruption. Until this happens, our results suggest that firms should increase the visibility of a leader's action and be vigilant about simultaneous (more leaderless) decision structures, when dishonesty requires coordination.

Declaration of Competing Interest

None.

Acknowledgments

We would like to thank Kai Barron, Vojtech Bartos, and Lea Bitter, and Tommaso Reggiani for their helpful comments. We would like to thank participants in various research seminars and conferences at RWTH Aachen, the Burgundy Business School, CREED, the University of Tilburg, the University of Trier, the Academy of Management Conference 2018, the Colloquium of Personnel Economics 2018, the Early Career Behavioral Conference in Bergen (ECBE) 2018, the Maastricht Behavioral and Experimental Economics Symposium (M-BEES) 2018, and the Ratio-DMEP workshop in Jerusalem (2019). Financial support from the German Research Foundation through Germany's Excellence Strategy - EXC 2126/1 - 390838866 and the research units "Design and Behavior" (FOR 1371) and "Rationality and Competition" (CRC TRR 190), Ben-Gurion University, and the University of Amsterdam (ERC project ERC-StG-637915) is gratefully acknowledged. The data set of the experiment is available at https://osf.io/taybk/.

Appendix

Supplementary analysis

The die-rolling paradigm does not allow to trace lying behavior for each single instance (since die-rolls are truly private). For the purpose of analysis, we treat the most profitable reports as "suspicious", and players whose reports are consistently suspicious as "brazen". We classify first mover reports as suspicious if they equal 3. Such reports are honest with an ex-ante probability of 1/3 and dishonest with an ex-ante probability of 2/3.[28] A player is classified as a "brazen liar" if suspicious in all 30 periods. The ex-ante probability of observing a 30-period long sequence of suspicious outcomes is practically zero ($(\frac{1}{3})^{30} = 4.86 \cdot e^{-15}$).

First mover behavior In the one-shot interaction (Stage 2) the relative frequency of suspicious first mover reports across treatments is identical (90%).

In the repeated interaction (Stage 3) the fraction of brazen liars is substantially lower among first movers in the sequential treatments, relative to *Simultaneous*. In *Simultaneous*, 71% of all subjects report the highest number throughout all periods, but only 49% of first movers in *Partially Sequential* and 44% in *Fully Sequential* do so. Pairwise comparisons between *Simultaneous* and *Partially*/*Fully Sequential* show that the differences are significant ($p = 0.011$, $d = 0.41$ and $p = 0.002$, $d =$

[26] See e.g., Podsakoff and Podsakoff (2019) for a discussion on the usage of experiments in leadership research.

[28] Second mover reports are classified as suspicious if they equal the first mover's report, and reports of third movers are suspicious if they equal both the first and second movers' reports. The probability of honestly matching previous reports is 1/3, so the ex-ante probability that second/third movers' reports are honest is 1/3, and the probability that they are dishonest is 2/3.

Table A.1
Estimating reports in Stage 2.

	Dep. var.: Reported number r_i			
	OLS		Ordered logit	
Independent variables	(1)	(2)	(3)	(4)
Partially Sequential	−0.05	−0.04	−0.41	−0.40
	(0.06)	(0.06)	(0.42)	(0.43)
Fully Sequential	0.03	0.04	−0.04	0.12
	(0.05)	(0.05)	(0.43)	(0.45)
Ind. report < 3 (Stage 1)		−0.28***		−1.55***
		(0.07)		(0.36)
Constant	3.09***	3.14***		
	(0.14)	(0.14)		
Observations	369	369	369	369
(Pseudo) R2	0.052	0.117	0.065	0.124
Additional Controls	Yes	Yes	Yes	Yes

Notes: The table displays results from regression analyses. Models 1 and 2 show the results from the OLS regression with robust standard errors and models 3 to 4 from the ordered logistic regression. Standard errors are in parentheses. The reference group for treatment is *Simultaneous*. All reporting roles (first, second, and third movers) are included. In all regression models we additionally control for Age and Gender. Controlling for the reporting order does not change the results. Significance indicators: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table A.2
Estimating coordination in Stage 2.

	Dep. var.: Coordination (0/1)			
	Subject level		Group level	
Independent variables	(1)	(2)	(3)	(4)
Partially Sequential	0.07	0.10*	0.07	0.09**
	(0.07)	(0.06)	(0.06)	(0.04)
Fully Sequential	0.20***	0.21***	0.20***	0.15***
	(0.06)	(0.06)	(0.06)	(0.04)
Ind. report < 3 (Stage 1)		0.31***		0.38***
		(0.07)		(0.14)
Observations	369	369	123	123
(Pseudo) R2	0.087	0.239	0.144	0.519
Additional Controls	Yes	Yes	Yes	Yes

Notes: The table displays marginal effects from the probit regression analyses. The dependent variable is the dummy equal to 1 if coordination is achieved, and 0 otherwise. Models 1 to 2 are based on individual data clustered on the group level. Models 3 to 4 are based on the group-level data. Here the control variables represent the group average. Robust standard errors are in parentheses. The reference group for treatment is *Simultaneous*. In all regression models we additionally control for Age and Gender. Significance indicators: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table A.3
Estimating period t reports in Stage 3.

	Dep. var.: Reported number $r_{i,t}$		
Independent variables	(1)	(2)	(3)
Partially Sequential	0.19***	0.19***	0.18***
	(0.06)	(0.05)	(0.04)
Fully Sequential	0.25***	0.27***	0.26***
	(0.06)	(0.05)	(0.05)
Ind. report < 3 (Stage 1)		0.18***	0.11***
		(0.04)	(0.04)
Triple 3 (Stage 2)		0.26***	0.24***
		(0.05)	(0.05)
Triple 3_{t-1} (Stage 3)		0.07*	0.04
		(0.04)	(0.04)
Period		0.00**	0.00***
		(0.00)	(0.00)
Constant	2.92***	2.67***	2.79***
	(0.02)	(0.06)	(0.13)
Observations	11,070	10,701	10,701
R^2	0.039	0.143	0.179
Additional controls	No	No	Yes

Notes: GLS regression with random effects. Robust standard errors clustered on the group level are in parentheses. The reference group is treatment *Simultaneous*. Control variables are “Ind. report <3” (1 if the subject did not report a 3 in Stage 1 of the experiment), “Period”, “Triple $3_{(t-1)}$” (1 if group coordinated on a triple of 3 in Stage 2)” “Triple $3_{(t-1)}$” (1 if the group coordinated on a triple of 3 in the previous period). All additional control variables not shown in the table are session dummies, age, and female. Model 1 contains data from periods 1 to 30, while Models 2 and 3 contain data from periods 2 to 30. Significance indicators: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table B.1
Estimating period t second mover behavior in Stage 2.

	Dep. var.: Reported number r_i	
Independent variables	(1)	(2)
Fully Sequential	0.12	0.11*
	(0.08)	(0.06)
First mover r		0.57***
		(0.10)
Constant	2.78***	1.13***
	(0.06)	(0.30)
Observations	123	123
R-squared	0.01	0.19

Notes: OLS regression robust standard errors in parentheses. The reference group is *Partially Sequential*. "First mover r" (reported number of respective first mover). Significance indicators: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table B.2
Estimating period t second and third mover behavior in Stage 3.

	Dep. var.: Reported number $r_{i;t}$ of... Second mover		Third mover
Independent variables	(1)	(2)	(3)
Fully Sequential	−0.05	0.03	
	(0.08)	(0.04)	
Period		0.00	-0.00
		(0.00)	(0.00)
Triple 3 (Stage 2)		0.23***	−0.09
		(0.07)	(0.08)
Triple 3_{t-1} (Stage 3)		0.00	0.05
		(0.03)	(0.05)
Ind. report <3 (Stage 1)		0.15**	0.05
		(0.07)	(0.06)
First mover r_t		0.68***	0.32***
		(0.06)	(0.07)
Second mover r_t			0.39***
			(0.07)
Constant	2.70***	0.67***	0.79***
	(0.05)	(0.16)	(0.21)
Observations	3690	3567	1189
R^2	0.001	0.463	0.423

Notes: GLS regression with random effects. Robust standard errors clustered on the group level are in parentheses. The reference group for model 1 and 2 is treatment *Partially Sequential*. Control variables are "Ind. report <3" (1 if the subject did not report a 3 in Stage 1 of the experiment), "Period", "Triple $3_{(t-1)}$" (1 if group coordinated on a triple of 3 in Stage 2), "Triple $3_{(t-1)}$" (1 if the group coordinated on a triple of 3 in the previous period), "Fist mover r_t" (reported number of respective first mover), "Second mover r_t" (reported number of respective second mover). Model 1 and 2 include behavior of second movers, model 3 includes behavior of third mover in *Fully Sequential*. Significance indicators: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

0.49, respectively, χ^2-test). Again, there is no significant difference between *Partially Sequential* and *Fully Sequential* ($p = 0.658$, $d = 0.07$).

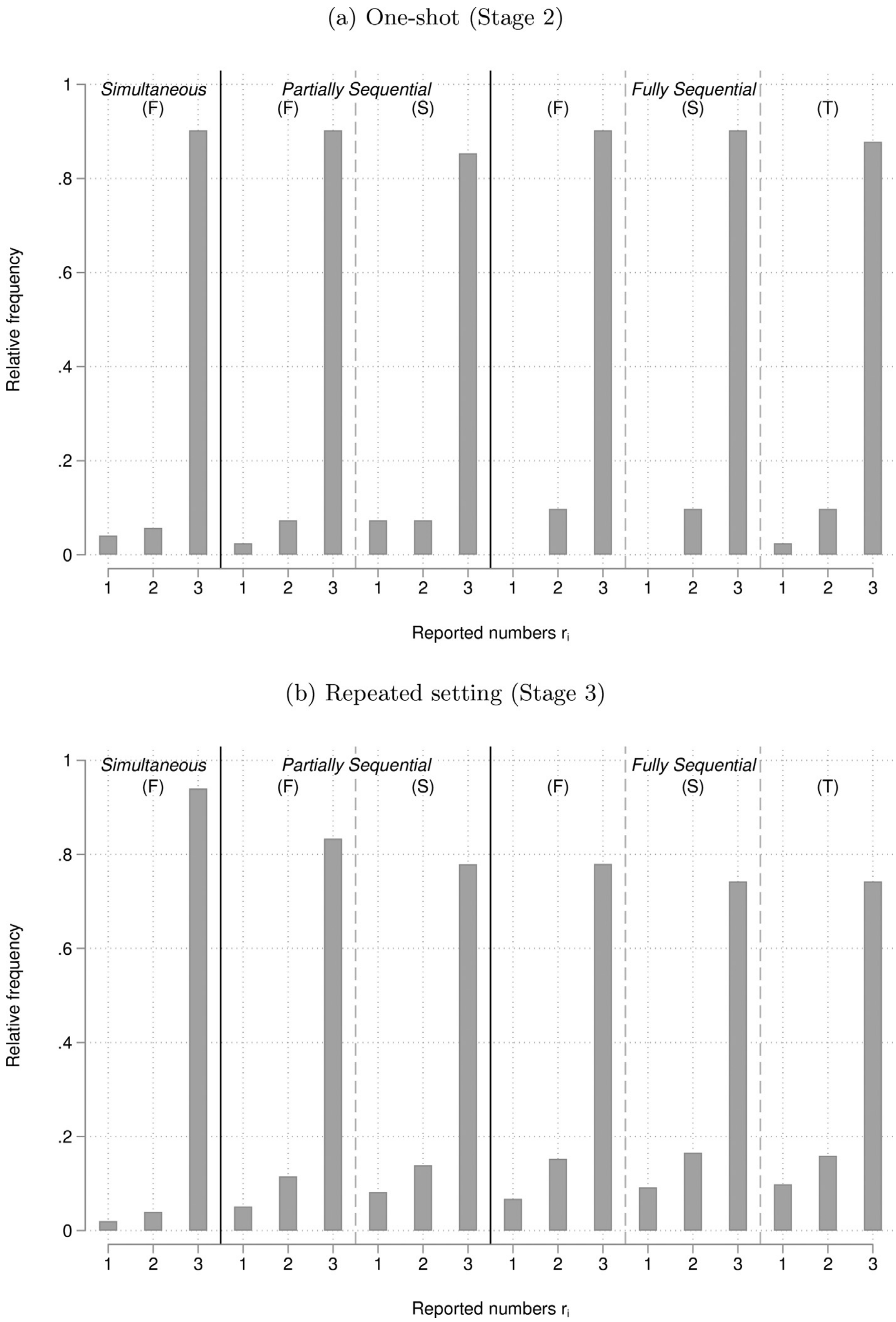
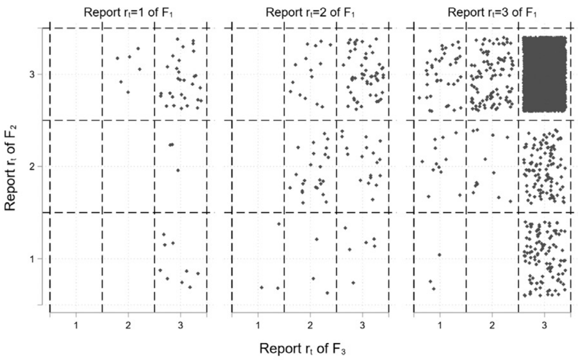


Fig. A.1. Histograms of reported numbers. *Notes:* For illustration purposes, we label first movers with F, second movers with S, and third movers with T.

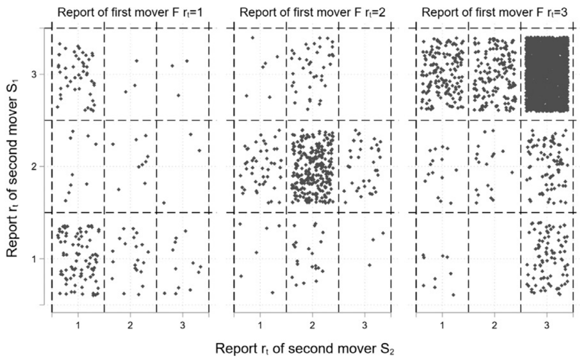
Follower behavior

In the repeated interaction (Stage 3), compared to the fraction of brazen liars in *Simultaneous* (71%) second movers in *Partially Sequential* (59%) and *Fully Sequential* (59%) tend to be less dishonest ($p = 0.064$, $d = 0.33$; $p = 0.643$, $d = 0.28$). However, no significant difference can be observed compared to third movers in *Fully Sequential* (73%, $p = 0.20$, $d = 0.06$).

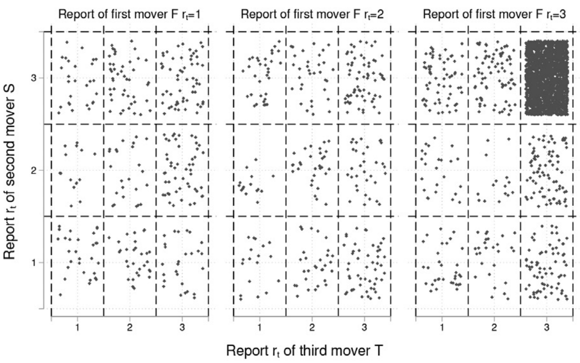
While in *Partially Sequential* 48% of all first movers brazenly report a three, 58% of all second movers match the first movers report brazenly across all third periods ($p = 0.383$, $d = 0.22$). In *Fully Sequential* 44% of all first movers brazenly report a three, 59% of second movers ($p = 0.269$, $d = 0.29$) and 73% of third movers match brazenly ($p = 0.013$, $d = 0.61$).



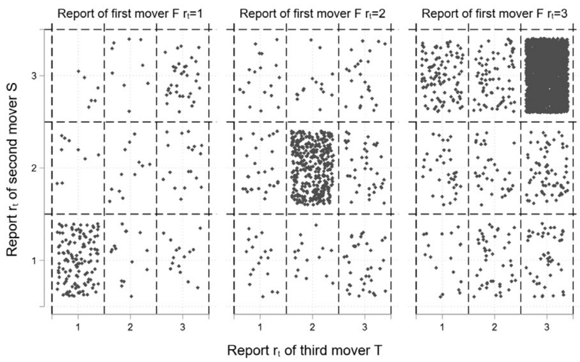
(a) *Simultaneous*



(b) *Partially Sequential*



(c) *Fully Sequential 3*



(d) *Fully Sequential*

Fig. A.2. Scatterplots of reported numbers in Stage 3. *Notes*: Scatterplots show observations from all periods of Stage 3 of the experiment. Each dot represents the combination of the reported numbers by a respective group of three. For reasons of comparability, we put *Fully Sequential 3* below *Simultaneous* and next to *Fully Sequential*.

Table B.3
Second and third mover behavior between stages.

	Dep. var.: Reported number $r_{i;t}$ of.. ... second mover			... third mover	
Independent variables	(1)	(2)	(3)	(4)	(5)
Fully Sequential	−0.04	0.12	0.11*		
	(0.08)	(0.08)	(0.06)		
Fully Sequential x Stage 3		−0.17**	−0.10		
		(0.08)	(0.07)		
Stage 3	−0.14***	−0.08	−0.02	−0.21***	−0.06
	(0.04)	(0.05)	(0.05)	(0.07)	(0.06)
First mover r_t			0.68***		0.31***
			(0.05)		(0.07)
Second mover r_t					0.38***
					(0.07)
Constant	2.84***	2.78***	0.81***	2.85***	0.86***
	(0.06)	(0.06)	(0.16)	(0.07)	(0.22)
Observations	3,813	3,813	3,813	1,271	1,271
R^2	0.003	0.003	0.42	0.003	0.412

Notes: GLS regression with random effects. Robust standard errors clustered on the group level are in parentheses. The reference group for model 1–2 is treatment *Partially Sequential*. “Fist mover r_t” (reported number of respective first mover), ”Second mover r_t” (reported number of respective second mover). Model 1–3 include behavior of second movers, model 4 and 5 include behavior of third movers in *Fully Sequential*. Significance indicators: $^{***}p < 0.01, ^{**}p < 0.05, ^{*}p < 0.1$.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jebo.2021.05.007.

References

Abeler, J., Becker, A., Falk, A., 2014. Representative evidence on lying costs. J. Public Econ. 113, 96–104.
Abeler, J., Nosenzo, D., Raymond, C., 2019. Preferences for truth-telling. Econometrica 87, 1115–1153.
Ahmad, M.G., Loch, C., 2020. What do the followers want? The core functions of leadership. Leadersh. Q. 31 (2), 101293.
Barr, A., Michailidou, G., 2017. Complicity without connection or communication. J. Econ. Behav. Organ. 142, 1–10.
Bartling, B., Fischbacher, U., 2012. Shifting the blame: on delegation and responsibility. Rev. Econ. Stud. 79 (1), 67–87.
Bendahan, S., Zehnder, C., Pralong, F.P., Antonakis, J., 2015. Leader corruption depends on power and testosterone. Leadersh. Q. 26 (2), 101–122.
Bonferroni, C.E., 1935. Il Calcolo Delle Assicurazioni su Gruppi di Teste. Tipografia del Senato, Rome.
Cappelen, A.W., Reme, B.-A., Sørensen, E.O., Tungodden, B., 2016. Leadership and incentives. Manag. Sci. 62 (7), 1944–1953.
Conrads, J., Irlenbusch, B., Rilke, R.M., Walkowitz, G., 2013. Lying and team incentives. J. Econ. Psychol. 34, 1–7.
D'Adda, G., Darai, D., Pavanini, N., Weber, R.A., 2017. Do leaders affect ethical conduct? J. Eur. Econ. Assoc. 15 (6), 1177–1213.
Dai, Z., Galeotti, F., Villeval, M.C., 2018. Cheating in the lab predicts fraud in the field: an experiment in public transportation. Manag. Sci. 64 (3), 1081–1100.
Danilov, A., Biemann, T., Kring, T., Sliwka, D., 2013. The dark side of team incentives: experimental evidence on advice quality from financial service professionals. J. Econ. Behav. Organ. 93, 266–272.
Dannenberg, A., 2015. Leading by example versus leading by words in voluntary contribution experiments. Soc. Choice Welf. 44 (1), 71–85.
Eichenseer, M., 2019. Leading by example in public goods experiments: what do we know? SSRN Electron. J 1, 1–35.
Eisenkopf, G., 2020. Words and deeds – experimental evidence on leading-by-example. Leadersh. Q. 31 (4), 101383.
EY, 2016. Corporate Misconduct: Individual Consequences. Technical Report.
Feess, E., Kerzenmacher, F., & Muehlheusser, G. (2020). Moral transgressions by groups: what drives individual voting behavior?IZA Discussion Paper No. 13383.
Falk, A., Neuber, T., Szech, N., 2020. Diffusion of being pivotal and immoral outcomes. Rev Econ Stud 87 (5), 2205–2229.
Fehrler, S., Fischbacher, U., Schneider, M.T., 2020. Honesty and self-selection into cheap talk. Econ. J.
Fischbacher, U., Föllmi-Heusi, F., 2013. Lies in disguise-an experimental study on cheating. J. Eur. Econ. Assoc. 11 (3), 525–547.
Gächter, S., Nosenzo, D., Renner, E., Sefton, M., 2012. Who makes a good leader? Cooperativeness, optimism, and leading-by-example. Econ. Inq. 50 (4), 953–967.
Gibson, R., Tanner, C., Wagner, A.F., 2013. Preferences for truthfulness: heterogeneity among and within individuals. Am. Econ. Rev. 103 (1), 532–548.
Glöckner, A., Irlenbusch, B., Kube, S., Nicklisch, A., Normann, H.T., 2011. Leading with(out) sacrifice? A public-goods experiment with a privileged player. Econ. Inq. 49 (2), 591–597.
Gneezy, U., Kajackaite, A., Sobel, J., 2018. Lying aversion and the size of the lie. Am. Econ. Rev. 108 (2), 419–453.
Greiner, B., 2015. Subject pool recruitment procedures: organizing experiments with ORSEE. J. Econ. Sci. Assoc. 1 (1), 114–125.
Gross, J., Leib, M., Offerman, T., Shalvi, S., 2018. Ethical free riding: when honest people find dishonest partners. Psychol. Sci. 29 (12), 1956–1968.
Gürerk, O., Irlenbusch, B., Rockenbach, B., 2009. Motivating teammates: the leader's choice between positive and negative incentives. J. Econ. Psychol. 30 (4), 591–607.
Güth, W., Levati, M.V., Sutter, M., van der Heijden, E., 2007. Leading by example with and without exclusion power in voluntary contribution experiments. J. Public Econ. 91 (5–6), 1023–1042.
Harsanyi, J.C., Selten, R., 1988. A General Theory of Equilibirum Selection in Games. The MIT Press.
Khalmetski, K., Sliwka, D., 2019. Disguising lies - image concerns and partial lying in cheating games. Am. Econ. J.: Microecon. 11 (4), 79–110.
Kocher, M.G., Schudy, S., Spantig, L., 2018. I lie? We lie! Why? Experimental evidence on a dishonesty shift in groups. Manag. Sci. 64 (9), 3995–4008.
Kröll, M., & Rustagi, D. (2016). Shades of dishonesty and cheating in informal milk markets in India. SAFE Working Paper, 134.
Nosenzo, D., Quercia, S., Sefton, M., 2015. Cooperation in small groups: The effect of group size. Exp. Econ. 18 (1), 4–14.
Podsakoff, P.M., Podsakoff, N.P., 2019. Experimental designs in management and leadership research: Strengths, limitations, and recommendations for improving publishability. Leadersh. Q. 30 (1), 11–33.
Potters, J., Stoop, J., 2016. Do cheaters in the lab also cheat in the field? Eur. Econ. Rev. 87, 26–33.
Rajan, R.G., Wulf, J., 2006. The flattening firm: evidence from panel data on the changing nature of corporate hierarchies. Rev. Econ. Stat. 88 (4), 759–773.
Rivas, M.F., Sutter, M., 2011. The benefits of voluntary leadership in experimental public goods games. Econ. Lett. 112 (2), 176–178.
Soraperra, I., Weisel, O., Zultan, R., Kochavi, S., Leib, M., Shalev, H., Shalvi, S., 2017. The bad consequences of teamwork. Econ. Lett. 160, 12–15.
Weisel, O., Shalvi, S., 2015. The collaborative roots of corruption. Proc. Natl. Acad. Sci. 112 (34), 10651–10656.
Wiltermuth, S.S., 2011. Cheating more when the spoils are split. Organ. Behav. Hum. Decis. Process. 115 (2), 157–168.