

Dimant, Eugen; Gesche, Tobias

Working Paper

Nudging Enforcers: How Norm Perceptions and Motives for Lying Shape Sanctions

CESifo Working Paper, No. 9385

Provided in Cooperation with:

Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

Suggested Citation: Dimant, Eugen; Gesche, Tobias (2021) : Nudging Enforcers: How Norm Perceptions and Motives for Lying Shape Sanctions, CESifo Working Paper, No. 9385, Center for Economic Studies and Ifo Institute (CESifo), Munich

This Version is available at:

<https://hdl.handle.net/10419/248930>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Nudging Enforcers: How Norm Perceptions and Motives for Lying Shape Sanctions

Eugen Dimant, Tobias Gesche

Impressum:

CESifo Working Papers

ISSN 2364-1428 (electronic version)

Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH

The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute

Poschingerstr. 5, 81679 Munich, Germany

Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de

Editor: Clemens Fuest

<https://www.cesifo.org/en/wp>

An electronic version of the paper may be downloaded

- from the SSRN website: www.SSRN.com
- from the RePEc website: www.RePEc.org
- from the CESifo website: <https://www.cesifo.org/en/wp>

Nudging Enforcers: How Norm Perceptions and Motives for Lying Shape Sanctions

Abstract

The enforcement of social norms is the fabric of a functioning society. Through the lens of multiple studies using different methodologies (a behavioral experiment and a vignette experiment in Study 1, as well as a norm elicitation experiment in Study 2), we examine how motives for lying and norm perceptions steer norm enforcement. Pursuing a pre-registered three-part data collection effort, our study investigates the extent to which norm breaches are sanctioned, how norm-nudges affect punishment behavior, and how enforcement links to norm perceptions. Using a representative sample of U.S. participants, we provide robust evidence that norm-enforcement is not only sensitive to the magnitude of the observed transgression (= size of the lie) but also to the consequence of the transgression (= whether the lie remedies or creates payoff inequalities). We also find that norm enforcers are sensitive to different norm-nudges that convey social information about actual lying behavior or its social disapproval. Importantly, these results hold both in the behavioral experiment and in an add-on vignette study that confirm the robustness of our findings in the context of whistleblowing. To explain the punishment patterns of the behavioral experiment in Study 1, we subsequently examine how norms are perceived across different transgressions and how norm-nudges change these perceptions. We find that social norm perceptions are malleable: norm-nudges are most effective when preexisting norms are vague. Importantly, we find that punishment patterns in the first experiment closely follow these norm perceptions. With that, our findings suggest that norm enforcement can be nudged successfully.

JEL-Codes: B410, D010, D900.

Keywords: lying, norm-nudges, nudging, punishment, social norms.

Eugen Dimant
University of Pennsylvania
Philadelphia / PA / USA
edimant@sas.upenn.edu

Tobias Gesche
Center for Law & Economics
ETH Zurich / Switzerland
tgesche@ethz.ch

This version: October 15, 2021

The most recent version of this working paper can always be downloaded by following this link:

<https://ssrn.com/abstract=3664995>

We would like to thank two anonymous reviewers, Loukas Balafoutas, Zvonimir Bašić, John Beshears, Giuseppe Danese, Mathias Ekström, Jana Freundt, Michael Hallsworth, Martin Kocher, Nils Köbis, Nikos Nikiforakis, Jan Schmitz, and Weiwei Tasch for their helpful comments. We would also like to thank Tanner Nichols for his help with the data collection for the second experiment. We acknowledge financial support from the German Research Foundation (DFG) under Germany's Excellence Strategy — EXC 2126/1– 390838866.

Introduction

Social norms are ubiquitous in human societies. They inform both individual behavior and interpersonal interactions in a variety of social and economic domains such as collective action, altruistic sharing, and deviance.¹ The consensus in the economic literature is that norm compliance can erode quickly, and enforcement is crucial to sustain social order (see, e.g., [Fehr and Gächter, 2000](#); [Balafoutas and Nikiforakis, 2012](#); [Balafoutas et al., 2014](#); [Khadjavi et al., 2021](#)).²

One particularly promising approach to enact behavioral change has emerged in the form of nudging ([Thaler and Sunstein, 2009, 2021](#)). The existing nudge literature typically focuses on interventions that aim at changing behavior of transgressors directly (e.g., [Benartzi et al., 2017](#); [Beshears and Kosowsky, 2020](#); [DellaVigna and Linos, 2020](#); [Dimant et al., 2021b](#); [Gelfand et al., 2021](#)). Recently, a growing body of literature has utilized so-called “norm-nudges”: nudges that attempt to change behavior by eliciting and changing existing social norms through the manipulation of social expectations.³ While some prior research points to instances where norm violators – and even those who fail to punish norm violators – are punished ([Winter and Zhang, 2018](#); [Martin et al., 2019](#); [Stamkou et al., 2019](#)), there is currently no systematic research accounting for how breaching different types of norms affects their enforcement. In addition, existing literature has been rather silent about the extent to which a punisher’s norm enforcement behavior can be nudged, especially with the help of norm-nudges. This is our core contribution: we show when and how norm enforcement patterns can be altered using norm-nudges.

More precisely, we address this gap in the literature by providing a complementary approach and utilize the power of norm-nudges to study their effects on those who *enforce* norm compliance (via punishment) of others. Across several decision scenarios that vary the motive behind and extent of the observed norm breach, we examine the effectiveness of interventions that use social information to steer norm enforcement. With that, our interventions aim at nudging individuals in positions of power who have the ability to enforce the transgressors’ adherence to social norms.

To accomplish this, we pursue our research agenda with two well-powered and pre-registered studies. Study 1 contains both a behavioral experiment and a vignette experiment examining the robustness of the behavioral findings. Here, we first use a behavioral experiment to investigate whether and how individuals differentiate punishment behavior based on different motives that

¹See, e.g., [Ostrom 2000](#); [Kimbrough and Vostroknutov 2016](#); [Albrecht et al. 2018](#); [Fehr and Schurtenberger 2018](#); [Bolton et al. 2021](#); [Choi et al. 2019](#); [Bott et al. 2020](#); [Bicchieri et al. 2020a](#); [Galeotti et al. 2021](#); [Dimant 2021](#).

²See also [Fehr and Fischbacher, 2004](#); [Herrmann et al., 2008](#); [Nikiforakis and Normann, 2008](#); [Sutter et al., 2010](#); [Fehr and Schurtenberger, 2018](#); [Bolton et al., 2021](#); [Bicchieri et al., 2021](#); [Brouwer et al., 2020](#). See also, punishment in (non-)monetary forms, such as shaming, is a powerful tool to increase compliance, trust, and cooperation (e.g., [Coleman, 1994](#); [Coricelli et al., 2010](#); [Dickinson et al., 2015](#); [Heffner and FeldmanHall, 2019](#)).

³For a theoretical conceptualization see [Bicchieri and Dimant \(2019\)](#). For applications see [Hallsworth et al. \(2017\)](#); [Bhanot \(2018\)](#); [Damgaard and Gravert \(2018\)](#); [Allcott and Kessler \(2019\)](#); [Bott et al. \(2020\)](#); [Bursztyn et al. \(2020a,b\)](#); [Dimant et al. \(2020\)](#).

lead to non-compliance with an existing norm. Our setting varies punishment opportunities in accordance with the participant’s motive: lying that achieves an unfair advantage over another person versus lying that restores equal chances. We capitalize on the concept of norm-nudging by investigating the extent to which punishment decisions are sensitive to norm information with respect to what others *do* (empirical information) or what people *approve of doing* (normative information). We borrow this approach directly from the social norms literature (e.g., Cialdini et al., 1990; Bicchieri, 2006; Tankard and Paluck, 2016; Young, 2016). Next, we verify the robustness of our core findings using another experimental method – a vignette – in which we elicit punishment patterns in the context of corrupt behavior in a company setting. In Study 2, we run a norm elicitation experiment to investigate the relationship between norm perceptions and punishment patterns in our behavioral experiment. We believe that this ‘modular approach’ of various experimental methods is best suited for achieving policy-relevant insights (Levitt and List, 2007; Falk and Heckman, 2009; List, 2011; Hallsworth and Kirkman, 2020).

In more detail, Study 1 consists of two separate data collection efforts: an incentive-compatible behavioral experiment that is the core of our investigation and a vignette experiment. In conjunction, these two experiments allow us to examine and establish the robustness of our core findings. In the behavioral experiment that studies the punisher’s behavior of Study 1 behavioral experiment, we consider a situation where two participants solve a task in order to qualify for having the chance to win an indivisible prize. One participant in each pair is then given the opportunity to lie. Here, larger lies increase a participant’s chance to win the fixed-size prize, thereby affecting concerns for (in-)equity.⁴ Using a representative sample of 1,240 U.S. participants acting as third-party punishers, we then study how such lying behavior is sanctioned in the behavioral experiment.⁵ We find that punishment increases in a linear fashion with the size of the lie. On top of that, the equity nature of the lie matters: if lying helps achieve equal chances of winning, then lying is punished less severely compared to when it would improve one’s chances of winning relative to the victim. Finally, we observe that punishers react to norm-nudges and punish more when lying behavior is in conflict with the information provided in the norm nudge. This is independent of whether the information points towards normative statements of others (disapproval of lying) or their actual behavior (that others did not lie).⁶ Thus, our results indicate that both types of norm-nudges affect punishment in similar ways.

⁴The context of lying is pertinent to our research question: existing literature has indicated that norm perceptions of lying can vary substantially and provide sufficient room for motivated reasoning (e.g., Bicchieri et al., 2020b; Dimant et al., 2020; Galeotti et al., 2020). Other attempts to shift norms in contexts where there is more ambiguity include the perception of migrants and female labor participation, among others (Bursztyn et al., 2020b,c; Grewenig et al., 2020; Haaland and Roth, 2020).

⁵Note, lying behavior data is collected solely for the purpose of being able to present real behavior that is sanctionable (and thus incentive-compatible) by other participants in the role of punishers.

⁶We base this information on two auxiliary experiments. In one, a majority of participants said that lying is not the right thing to do for participants in the behavioral experiment (=normative information). In another auxiliary experiment we verify that a majority of participants who could lie did not do so (=empirical information).

We also corroborate the external validity of these findings in a separate, pre-registered vignette experiment using a general population sample of $n=225$ participants. This data was collected in-person and outside the laboratory across various locations in the United States. In the vignette, exerting punishment was possible by “blowing the whistle” and reporting someone who accepted a bribe. Consistent with our design from the behavioral experiment – both the motives for deviance and the presence of norm information – were varied. The result from the vignette support those from the behavioral experiment and show that norm nudges can foster the intention to punish norm violators by blowing the whistle on them.

Study 2 then allows to get a better understanding of the mechanism behind norm enforcement. It examines whether the punishment patterns observed in the behavioral experiment have an analogous variation in the *perception* of social norms, using a separate set of 1,519 participants in. We use an incentive-compatible method to examine how the provision of norm information, the equity nature of the lie, and the interplay of these dimensions affect third parties’ perception of social norms (i.e., what one thinks others consider appropriate behavior). The elicited norm perceptions are consistent with the findings from Study 1: lying, which generates an advantage for the liar, relative to the other participant, is perceived to be less appropriate than lying that yields more equal chances for both participants. Also reflecting on the first experiment’s findings, we observe that providing norm information increases the perceived inappropriateness of lying, independent of the source of this information (normative statements that others perceive lying as inappropriate versus information that a majority of others did not lie when given the chance). We find that norm information works particularly well in situations where norms are perceived to be more lenient (i.e., when lying results in more equal chances).⁷

Taken together, our experiments not only allow us to investigate *when* and *how* norm breaches are sanctioned, but also show us that variations in norm enforcement are consistent with differences in the perception of social norms. Intuitively, this suggests that regulatory (top-down) interventions implemented to change behavior can be complemented by social (bottom-up) enforcement through informal norm-nudging, at least where social norms are clearly defined, transgressions are observable, and when they can be sanctioned. With that, we connect three literature streams: the study of transgressions in the context of lying, the enforcement of norms through punishment, and the perception of social norms. To the best of our knowledge, this approach is the first to investigate ways to nudge norm enforcement with social (rather than monetary) incentives via norm enforcers, as an alternative to traditional policy interventions that focus on nudging the behavior of transgressors directly (Gino et al., 2019).

⁷Our results also relate to existing research examining how equity concerns, social perception, and the size of a lie affect dishonest behavior (see, e.g., Gino and Pierce, 2010; John et al., 2014; Dufwenberg and Dufwenberg, 2018; Gneezy et al., 2018; Abeler et al., 2019). While this literature typically focuses on the factors determining the *intrinsic* costs of lying, we examine how these factors shape the *extrinsic* costs via punishment.

With respect to policy implications from our paper, it is important to note that existing research on nudging has often taken an *individualistic* perspective that focused on facilitating behavior change by directly targeting the individuals of interest. Existing research indicates that this can be a successful approach, although the immediate effectiveness and longevity of such interventions vary considerably (Brandon et al., 2017; Hummel and Maedche, 2019; DellaVigna and Linos, 2020). By targeting those who enforce behavior – rather than those whose behavior one wants to alter – we investigate the *collectivistic* perspective on behavior change. Peer mechanisms can successfully uphold norms such that deviations, even in the absence of formal rules and laws. However, this can be socially costly (Andreoni and Gee, 2012; Bicchieri et al., 2021). Our findings also provide an additional perspective to overcoming the challenges in nudging individuals directly (Sunstein, 2016; Hagmann et al., 2019) and in providing targeted tools for implementing theory of change successfully (Hallsworth and Kirkman, 2020). By complementing the arsenal of behavioral change techniques targeting individual decision-making (streamlining decision environments, defaults etc.), policy-makers can build momentum at the collective level by targeting those who can enforce norm adherence. As Bujold et al. (2020) argue, succeeding at this form of ‘meta-nudging’ is particularly important for policy-makers in various contexts, such as environmental preservation. We show that nudging can successfully steer the enforcers’ behavior and find that a collective perspective on nudging may harness behavioral insights. This perspective is not yet fully utilized in existing theoretical and applied work, and we hope that our findings can steer policy efforts in that direction.

Study 1: Norm Information and Norm Enforcement

Experimental design and procedures of the behavioral experiment

This experiment consists of two sub-experiments: a **Liar Experiment** and a **Punisher Experiment**. The former is necessary to elicit punishment behavior in the Punisher Experiment in an incentive-compatible manner. Figure 1 provides an overview of the main steps in these experiments, which were conducted in April/May 2019. A detailed description follows below:

Step 1 – Real effort task: The Liar Experiment involved $n=170$ participants, recruited via Amazon’s Mechanical Turk (MTurk); their demographics are presented in Table A.2 in the appendix.⁸ Participants were grouped in teams of two. Each team member then had to solve a real-effort task by correctly counting the 1’s in five matrices with numbers. This took about 7 minutes and earned each team member \$1.5 (added to a base payment of \$0.5).

⁸ To meet the criteria for robustness and generalizability of MTurk findings (e.g., Arechar et al., 2018; Snowberg and Yariv, 2021), we applied high quality restrictions on the sample: We utilized a combination of CAPTCHAs and comprehensive screening questions. Participants had to be in the US and have an approval rate >99%.

Step 2 – Bonus lottery and lying: Subjects in the Liar Experiment knew that in addition to their individual compensation of \$1.5, solving the real-effort task entitled them to win an additional bonus of \$2.5. However, this bonus could only be won by one person within each team. Specifically, the bonus was allocated by drawing a winning lottery ticket for each team where the number of tickets (and therefore chance of winning) differed by a subject’s role:

- *Victim:* One of the two subjects in each team was randomly allocated to this role and received a fixed amount of 3 lottery tickets.
- *Reporter:* The other subject in the team was entitled to a number of lottery tickets equal to the randomized outcome of a virtual 5-sided die that they rolled. However, subjects in this role were made aware that what they *reported* (rather than the actual outcome) would determine the number of lottery tickets they would receive.

The above setting provided an incentive to lie and overstate the outcome of one’s die roll to increase the chance of winning the bonus (similar to the “die under the cup”-paradigm as in, e.g., [Buccioli and Piovesan, 2011](#); [Fischbacher and Föllmi-Heusi, 2013](#), except that we have observable outcomes). Given the incentive to lie in this task, reported outcomes are indeed about 29% higher, on average, than the actual outcome of the die roll (Wilcoxon signed rank test: $p < 0.001$).⁹ Also, note that this situation is deliberately designed to be non-strategic: Reporters who lie cannot justify their behavior by forming (motivated) beliefs about their teammate’s behavior. In order to obtain a truthful measure of lying behavior while avoiding deception, reporters were told, after submitting their report, that a third participant in the role of a *punisher* would later observe – and potentially punish – their behavior (amounting to punishment of up to \$1.50). Participants in the role of reporters were then given the opportunity to revise their initial report, which was then used to determine their punishment in the behavioral experiment.

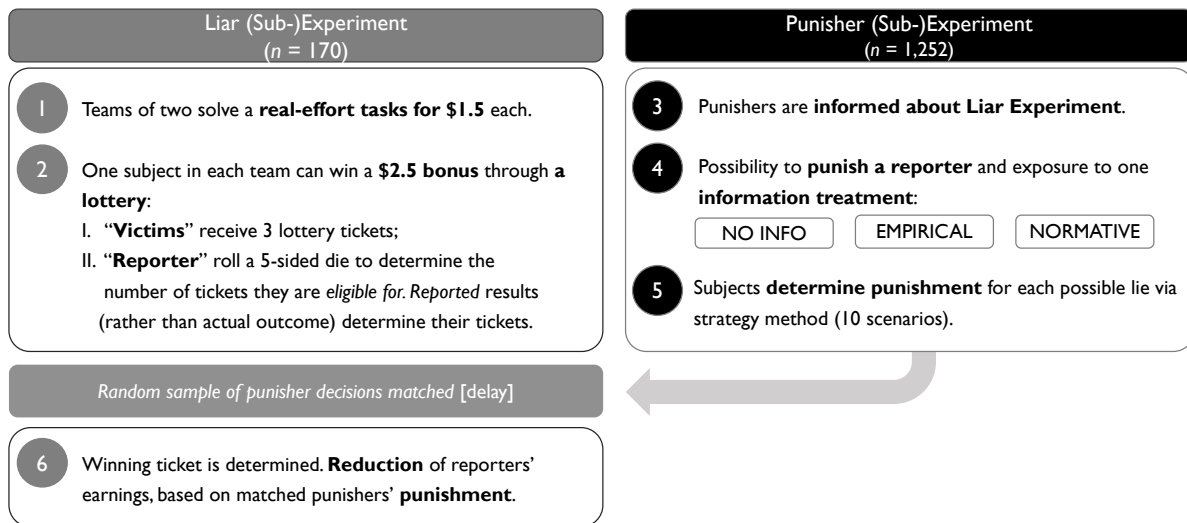
Step 3 – Punishers learn about the Liar Experiment: The Punisher Experiment featured $n=1,240$ subjects. They were recruited through a professional market research firm in order to get a representative sample of the US working-age population (for subjects’ demographics, see [Table A.3](#) in the appendix). This allows us to draw more generalizable inferences with respect to policy implications (see [Levitt and List, 2007](#); [Exadaktylos et al., 2013](#)).¹⁰

In the first part of the Punisher Experiment, the subjects read a description of the lying game. They were then assigned to the role of a punisher and learned that they could be matched with a reporter from the Liar Experiment whom they could punish for dishonest behavior. Punishment

⁹Since we elicited punishment using the strategy method, the actual lying behavior was irrelevant for our analysis of punishment decisions. However, we observe that the magnitude of punishment was calibrated well enough so that the *threat of* punishment works: The revised reports were, on average, lower by 0.44 than the the initial reports (signed rank test: $p < 0.001$). They were, however, still higher (by about 0.46) than the actual outcomes (signed rank test: $p=0.003$). [Figure A.1](#) in the appendix provides more details on the (revised) reporting pattern.

¹⁰Data collection was conducted through the market research firm Dynata (formerly “Research Now”) using a quota-based sample, with the aim of having representativeness for age and gender based on the US population.

Figure 1: Design of the Behavioral Experiment in Study 1)



would be implemented by decreasing the earnings of a reporter who lied about the die's outcome (after revision, see Step 2). Punishers had to pass some comprehension questions first, ensuring that they understood the context and impact of their decision.

Step 4 – Norm Information treatments: Right before punishers were asked to determine their punishment, they were presented with one randomly assigned norm information treatment. The two treatments EMPIRICAL and NORMATIVE provided norm information, whereas the baseline condition NO INFO contained no such information. The provided information was either descriptive (what other participants previously *did*; treatment EMPIRICAL) or normative (what other participants stated was the right thing to do; treatment NORMATIVE). More precisely, punishers in EMPIRICAL were told about a previous, auxiliary study in which a “Player A” was in the same decision scenario as reporters in the Liar experiment and what Player A’s *actually did*. Punishers in NORMATIVE were told about another auxiliary study in which participants said what a Player A’s *should do*.¹¹ One of the following two messages was then presented to punishers right before punishment decisions could be made:

Norm-nudge in EMPIRICAL – what Player A (decision as in Liar Experiment) *actually did*:

The majority of Player As in the previous study reported the number truthfully (i.e., reported exactly what the die showed).

Norm-nudge in NORMATIVE – what others stated Player A *should do*:

The majority of these people stated that the right thing to do for Player A is to report the number truthfully (i.e., report exactly what the die showed).

¹¹To ensure the validity of this information, we conducted the auxiliary experiment for EMPIRICAL with n=52 subjects who were in the same decision situation as reporters in the Liar Experiment, except that they could not revise their statement since there was no punishment. The majority of them (57.7%) reported truthfully. The auxiliary experiment for the NORMATIVE treatment involved n=51 subjects who were told about the Liar Game and the reporter’s decision scenario. They then had to report what they considered to be “the right thing to do”. The majority (68.6%) chose that reporting the actual outcome (i.e., to tell the truth) was the right thing to do.

Table 1: Assignment of punishment points (Example)

<i>I want to assign the following number of punishment points if Participant A...</i>						
	<i>...had a die outcome of "1"</i>					
	0	1	2	3	4	5
<i>...and reported "2"</i>	●	○	○	○	○	○
<i>...and reported "3"</i>	○	●	○	○	○	○
<i>...and reported "4"</i>	○	●	○	○	○	○
<i>...and reported "5"</i>	○	●	○	○	○	○
	<i>...had a die outcome of "2"</i>					
	0	1	2	3	4	5
<i>...and reported "3"</i>	○	○	●	○	○	○
<i>...and reported "4"</i>	○	○	●	○	○	○
<i>...and reported "5"</i>	○	○	●	○	○	○
	<i>...had a die outcome of "3"</i>					
	0	1	2	3	4	5
<i>...and reported "4"</i>	○	○	○	●	○	○
<i>...and reported "5"</i>	○	○	○	○	●	○
	<i>...had a die outcome of "4"</i>					
	0	1	2	3	4	5
<i>...and reported "5"</i>	○	○	○	○	○	●

Notes: The order determining whether punishment scenarios were presented by the die’s actual outcome in an increasing manner (shown here) or decreasing manner was randomized. For punishers, a reporter was referred to as “Participant A”. For the original screen, see Appendix D. The lying/punishment scenarios presented here will be referred to as p12, p13, p14, p15, p23, p24, p25, p34, p35, and p45, respectively.

Step 5 – Punishment: Punishment was elicited via strategy method: For each of the ten possible scenarios of dishonest reporting in the Liar Experiment, punishers could assign between 0 and 5 punishment points. Punishers knew that if they were matched with a liar, their decision would have direct payoff consequences for that liar. That is, the liar’s earnings would be reduced by \$0.3 for each punishment point that the punisher assigned in the particular scenario that corresponded to actual behavior of the liar.¹² Note that to limit the interpretation of our results to the main channels of interest, our experimental design to elicit punishment is non-strategic on purpose.¹³ Table 1 shows an example of how a punisher could assign punishment.

¹²For example, if such player rolled a “2” but reported a “4” and the matched punisher assigned $x \in \{0, \dots, 5\}$ punishment points for this scenario, that player’s earnings would be reduced by $x \times \$0.3$. Note that the relevant report for punishment was the second, potentially revised, report submitted by the active player. Also, while punishment affected that player’s earnings if matched with the punisher, it did not affect the lottery tickets and chances to win the bonus.

¹³For the purposes of our research question, a key advantage of our design over other games that have previously been used to study punishment behavior (e.g., public good games) is that it allows to study the effects of norm-nudges information in a setting where behavior change is solely driven by the main mechanisms of interest: the motive of the transgression and the provision of a norm-nudge. We deem this to be a suitable design choice compared to more complex, strategic settings in which the behavior of one’s peers is also relevant to the payoff.

Step 6 – Matching and payment: The experiment ended after punishers made their decisions and answered an exit questionnaire. Subsequently, a subset of punishers was randomly chosen to be matched with the reporters from the lying experiment and participants were then paid their respective earnings and bonuses.¹⁴

Further design aspects: Our norm information intervention (EMPIRICAL or NORMATIVE, see Step 4) follows a long and established tradition in the social norms literature (see, e.g., Cialdini and Goldstein, 2004; Bicchieri, 2006; Schultz et al., 2007; Goldstein et al., 2008). This literature has repeatedly validated the use of such “majority” messages – in which the majority of others’ behavior or approval is utilized – to point to social norms (see, e.g., Allcott, 2011; Ferraro et al., 2011; Bhanot, 2018). A common concept in this research is that norms are behavioral patterns embedded in a shared understanding of acceptable actions within a reference group (Ostrom, 2000) and that social norms have two distinct components (see Cialdini et al., 1990; Bicchieri, 2006): an empirical component (often referred to as a *descriptive norm*) and a normative component (often referred to as an *injunctive norm*). From this approach, norms can be understood as coordination games among the members of a reference group. Further, a shared signal can sustain norm adherence by facilitating coordination (Bicchieri, 2006; Young, 2016). Prior research that has leveraged these concepts has shown that both ingredients of a social norm affect and guide behavior, but their relative effectiveness can often differ (Schultz et al., 2007; Bicchieri and Xiao, 2009; Bhanot, 2018; Bicchieri et al., 2020b, 2021). This distinction is a central part of our experimental design which enables us to capture heterogeneous effects of norm nudges depending on what “component” of a norm is breached.

In order to carve out the effectiveness of norm-nudges in a methodologically reliable manner, we chose to use the strategy method and elicit punishment behavior for each potential lying scenario. We acknowledge that there are pros and cons of either choice, as discussed in Brandts and Charness (2011) who evaluate punishment studies and find that overall, the use of the strategy method leads to lower effects compared to the direct response approach. In this regard, our findings should indicate a lower bound effect of norm nudges. Note that due to the fact that we use the strategy method and the considered scenarios consistently for all subjects, any demand effect that may come with the use of this method does not correlate with our norm nudge treatments (for supporting discussion see Zizzo, 2010; De Quidt et al., 2018).

Another important feature of our design is punishment to be costless for the punisher. This is a necessary design choice in order to study the impact of the treatment variations mentioned above without having to consider the role of other factors (e.g., an additional monetary trade-off

¹⁴ Thus the matching ratio of punishers to active players was roughly 15:1. The instructions did not state or suggest any explicit value for this ratio. However, our exit questionnaire for punishers asked participants to provide an estimate for it. The average value (excluding one extreme outlier) implies a ratio of 1:0.87 – punishers therefore over-perceived the consequences of their decision. We also control for punishers’ estimated implementation ratio in our regression analysis but do not find a significant effect in any our specifications.

or risk assessment when there is the threat of counter-punishment; see [Coffman, 2011](#) or [Feess et al., 2018](#) for similar design choices).¹⁵ This reflects our goal to study the effectiveness of norm information in affecting punishment patterns absent incentives that counter the nudge intervention. Otherwise, the presence of incentives could mute observing the effects of norm nudges that would have otherwise been observable. However, we also acknowledge that while this paper, as a first step, isolates and examines the pure impact of norm-nudges in a controlled setting, punishment costs may be factor in future field research.

Behavioral Predictions

Our design of the die-roll and reporting task in the Liar Experiment allowed for different “punishment scenarios” in the Punisher Experiment. We denote them by a “p” and a number whose first digit indicates the actual outcome of the die roll and the second digit indicates the reported outcome. For example, the scenario when the active player rolled a “1” but reported a “2” will be referred to as “p12”. These ten scenarios differed on two main dimensions:¹⁶

1. The size of the lie (how much the reported outcome of the die roll exceeded the actual one).
 - *Lie size=1*: report larger by 1 than actual outcome (p12, p23, p34, and p45).
 - *Lie size=2*: report larger by 2 than actual outcome (p13, p24, p35).
 - *Lie size=3*: report larger by 3 than actual outcome (p14 and p25).
 - *Lie size=4*: report larger by 4 than actual outcome (p15).
2. The “equity nature” of a scenario (the chances of obtaining the bonus for the *active* player, relative to the *passive* player).
 - *Equity*: lying leads to more equity (reduces the gap) in the chances of winning the lottery (p12, p13, and p23).
 - *Inequity*: lying leads to (more) inequity in the chances of winning the lottery (p34, p35, and p45).
 - *Overclaiming*: starting from a situation with a disadvantaged active player, lying *reverts* inequality, leading to a now disadvantaged *passive* player (p14, p15, p24, and p25).

Combined with the norm information treatments, these punishment scenarios allow us to extend insights from the existing literature in order to explore the relationship between motives for norm breaching and punishment. This led to the following set of pre-registered hypotheses.¹⁷

¹⁵[Nikiforakis \(2008\)](#) shows that counter-punishment decreases enforcement of cooperation (see also [Balafoutas et al., 2016](#), for recent findings regarding how concerns for counter-punishment impede norm enforcement).

¹⁶Table [A.1](#) in Appendix A provides a summary of all punishment scenarios along these dimensions.

¹⁷See Appendix C for details on the pre-registration.

Size of the lie: We start by drawing on the extensive literature in the social sciences, which has established the determinants of lies and lying costs (Abeler et al., 2014; Dufwenberg and Dufwenberg, 2018; Gneezy et al., 2018; Abeler et al., 2019). Little is known, however, regarding whether and how these findings are reflected in the punishment of lies, especially not within the context of social norm breaching. Based on the existing theories and experimental evidence in regards to lying, we hypothesize that not only the occurrence of a lie but also its size matters:

HYPOTHESIS 1. *The amount of punishment assigned increases with the size of the lie, the reported minus actual outcome.*

Equity nature of the lie: A large literature emphasizes the importance of equity concerns, including in the context of deviant behavior (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000; Bolton et al., 2021). An unexplored question, however, is whether such motivations matter for the assessment of a norm breach and, consequently, affect the severity of punishment. Note that a mere aversion to unequal chances in getting the bonus is not able to explain the hypothesized effects of the norm nudges, as they do not change allocation. It also does not explain why larger lies should attract more punishment: this is because larger lies do, on average, increase inequity as much as they decrease it.¹⁸ However, for a given lie size, some punishment scenarios do increase this inequity while others decrease it. We hypothesize that breaching a norm in the form of over-reporting for the purposes of “getting ahead in unfair ways” is assessed differently from the purpose of leveling the playing field (see also Gino and Pierce, 2010; Bortolotti et al., 2017). While this logic also leads us to expect that lying in the Inequity-scenarios or Overclaim-scenarios will be punished harsher than in the Equity-scenarios, how punishments differ between the former two scenarios is an empirical question that we will investigate in our analysis.

HYPOTHESIS 2. *The equity nature of the lie matters. For a given size of the lie, the amount of punishment assigned*

- a) in Equity-scenarios is lower than in Inequity-scenarios,*
- b) in Equity-scenarios is lower than in Overclaim-scenarios.*

Effect of norm nudges: Finally, we derive hypotheses for our norm-information treatments. Existing research suggests that people are receptive to norm information and conform to both observed behavior and normative messaging (e.g., Goldstein et al., 2008; Bott et al., 2020; Dimaunt, 2019). We therefore hypothesize that norm nudges in EMPIRICAL and NORMATIVE

¹⁸Across the ten punishment scenarios, there is always either a scenario where, for a given lie size, an inequity increase in the chance to obtain the bonus is offset by a scenario where the same lie size decreases inequity (relative to an equal split of 3: p14/p25, p13/p35, p12/p45, and p23/p34) or where a lie only reverses inequity (p15 and p24) but does not change it. Note that the punishment can also not be used to decrease (expected) inequity via the punishment amount subtracted from the lying player’s earnings. The reason is that the size of the lie and inequity are uncorrelated. Higher punishment for larger lies would thus increase the (expected) inequity.

lead to more punishment compared to those in NO INFO. In addition, theoretical and experimental insights from [Bicchieri et al. \(2020b\)](#) suggest that while people interpret honest behavior, which is a costly action, as a strong indicator of normative disapproval of lying, the reverse may not be true: Merely saying what (not) to do does not necessarily have to be followed by the corresponding actions. Thus, we expect that empirical information may work as a stronger norm nudge in that acting in breach of what people actually do (“walking the talk”) can be a stronger signal than acting in breach of what people say one should do:

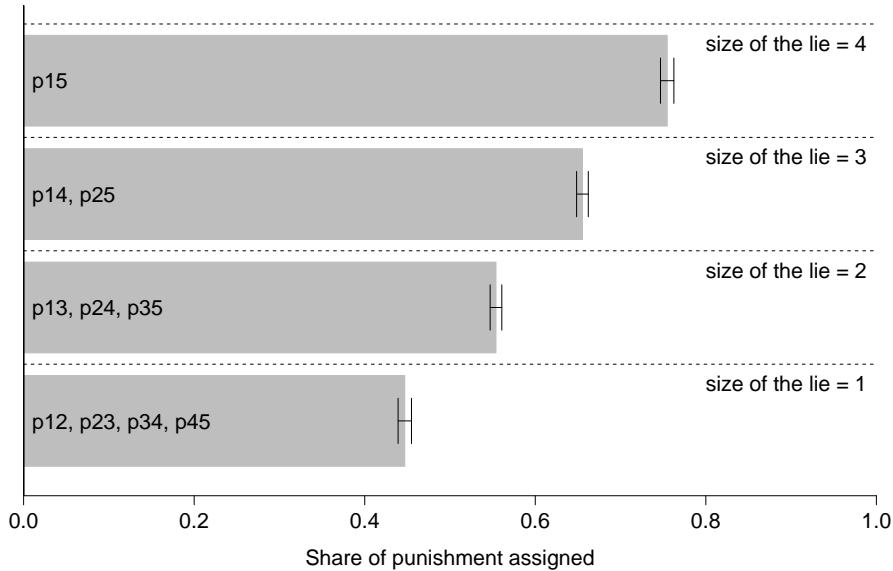
HYPOTHESIS 3. The amount of punishment assigned increases over our three treatments in the following order: NO INFO < NORMATIVE < EMPIRICAL

Results of the Punishment Experiment

In the following section, we report the punishers’ actions. In doing so, we follow the order of the hypotheses described above using non-parametric methods. In a last step, we corroborate our results using a regression framework.

Punishment and the size of the lie: To examine the relationship between the size of the lie and punishment, we calculate the mean of the punishment assigned across different sizes of lies.¹⁹ Figure 2 visualizes the results. Consistent with our intuition as formulated in Hypothesis 1, punishment increases significantly with the size of the lie.

Figure 2: Punishment by size of the lie



Notes: Punishment assigned as a share of total punishment points available in the punishment scenarios for a given size of the lie (associated punishment scenarios are displayed in each bar). Error bars denote SEM.

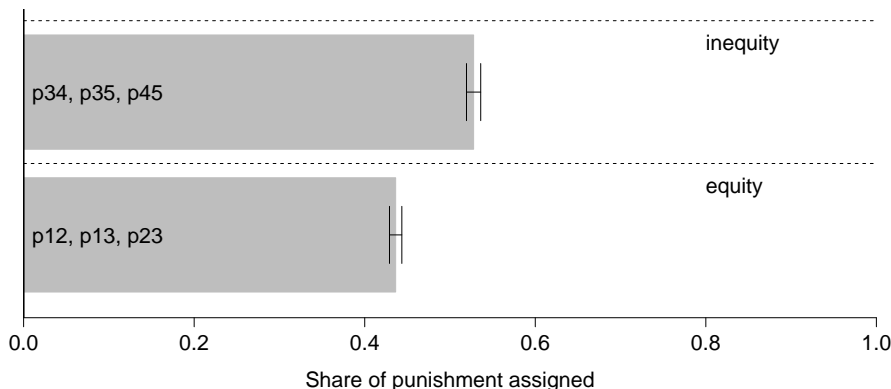
¹⁹For example, if the size of the lie = 2, it groups the share of punishment assigned in the scenarios p13, p24, and p35 (Figure A.2 in the Appendix shows the ungrouped results for every single punishment scenario).

Specifically, we observe that the smallest possible lie (size = 1) is only assigned a share of punishment of 44.7% of the maximum possible punishment. With each larger lie, the share of punishment increases by about 10 percentage points in a linear manner. A Wilcoxon signed rank test also shows that the shares of punishment for different sizes of the lies always differ significantly ($p < 0.001$ for all six pairwise comparisons). These results provide compelling evidence that individuals do not only punish norm breaches per se, but also take into consideration the extent to which norms are breached.

Punishment and the equity nature of the lie: Hypothesis 2 posits that while individuals care about the extent of a norm breach, not all norm breaches are created equally and, thus, not punished equally. That is, lying to correct an initial unfair situation might be judged – and punished – differently than lying to exacerbate an already unfair situation. Consequently, we examine Equity-, Inequity-, and Overclaim-Lying separately.

Figure 3 displays the means for punishment in the scenarios that feature Equity- or Inequity-Lying.²⁰ In both cases, there are three associated scenarios with the same lie sizes ($2 \times \text{Lie size} = 1$ and $1 \times \text{Lie size} = 2$). While punishment for lies that achieve equity amounts to 43.6%, the share of punishment for lies that achieve inequity is 9.2 percentage points higher, at a level of 52.8%. We also find that the punishment choices across these two equity norms are significantly different (Wilcoxon signed rank test: $p < 0.001$). This confirms Hypothesis 2a.²¹

Figure 3: Punishment by equity norm



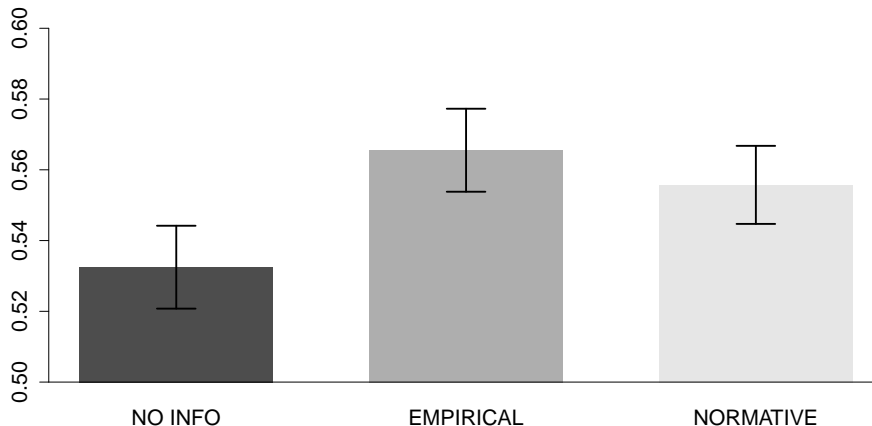
Notes: Punishment assigned as a share of total punishment points available in the punishment scenarios for a given equity norm (associated punishment scenarios are displayed in each bar). Error bars denote SEM.

²⁰Figure A.3 in Appendix A displays the average punishment level for each individual punishment scenario.

²¹Note that the figure does not display punishment for Overclaim-Lying. The reasons are that there are four associated punishment scenarios, rather than the three scenarios for Equity- and Inequity-Lying and that they also have different lie sizes ($1 \times \text{Lie size} = 2$, $2 \times \text{Lie size} = 3$, and $1 \times \text{Lie size} = 4$). This makes it visually challenging to compare Overclaim-Lying with the other scenarios. We provide this comparison as a part of the regression analysis further below, which allows us to account for these differences and answer Hypothesis 2b.

Punishment and norm-nudges: To examine Hypothesis 3 we start by looking at the aggregate impact of norm nudges on punishment. To test this, we compute the share of the possible punishment that subjects assigned across the 10 punishment scenarios shown to them.²² We then compare the mean of the share of punishment assigned by punishers in our three treatments. Figure 4 displays the means and their associated standard errors.

Figure 4: Punishment in the different norm information treatments



Notes: Punishment assigned as a share of total punishment points available in the 10 punishment scenarios, by norm-information treatments. Error bars denote SEM.

Following Hypothesis 3, we expect behavior that is in conflict with others' behavior (EMPIRICAL treatment) should be punished at least as harshly as behavior that is in conflict with what is deemed appropriate by others (NORMATIVE treatment). We find conclusive evidence for our hypothesis in that both information treatments affect the extent of inflicted punishment. In particular, we observe the highest punishment in the EMPIRICAL treatment, which is significantly different from punishment in the NO INFO treatment (Wilcoxon rank-sum test: $p=0.038$). Punishment in EMPIRICAL is also directionally – but not significantly – larger than punishment in the NORMATIVE condition (Wilcoxon rank-sum test: $p=0.734$). Our results also show that punishment in the NORMATIVE condition is larger than punishment in the NO INFO condition (Wilcoxon rank-sum test: $p=0.058$). In summary, we find evidence that is consistent with Hypothesis 3 such that observing lying behavior in combination with norm-related information leads to higher punishment.

Regression analysis: Our regression results come from fitting the regression equation

$$p_{is} = \alpha + \beta_1 \text{EMPIRICAL}_i + \beta_2 \text{NORMATIVE}_i + \beta_3 \text{LieSize}_s + \beta_4 \text{Inequity}_s + \beta_5 \text{Overclaiming}_s + \gamma' \text{Controls}_i + \epsilon_{is}. \quad (1)$$

²²For example, consider a punisher who assigned 2 punishment points (=40% of available punishment) in five scenarios and 3 punishment points (=60% of available punishment) in the remaining five scenarios. That punisher's share of punishment assigned would then be 50%.

Table 2: Determinants of punishment – regression results

	Share of punishment assigned in %			
	(1)	(2)	(3)	(4)
Size of the Lie	10.321*** (0.295)			10.110*** (0.324)
Inequity		9.140*** (0.532)		9.140*** (0.532)
Overclaiming		21.948*** (0.523)		5.098*** (0.402)
EMPIRICAL			3.513** (1.651)	3.513** (1.651)
NORMATIVE			2.734* (1.610)	2.734* (1.610)
Constant	35.412*** (6.008)	44.533*** (5.981)	53.810*** (6.130)	28.810*** (6.165)
<i>N</i>	12,400	12,400	12,400	12,400
<i>R</i> ²	0.123	0.100	0.009	0.140

F-test: coeff. for Size of the Lie; $p < 0.001$ (all pairwise comparisons in Col. 1 or 2)

F-test: coeff. for Inequity = Overclaiming; $p < 0.001/0.001$ (Col. 3/4)

F-test: coeff. for EMPIRICAL = NORMATIVE; $p = 0.629/0.629$ (Col. 1/4)

Notes: OLS results regressing the share of punishment assigned on the size of the lie, the equity nature of the lie (inequity, overclaiming), and indicators for the norm information treatment (EMPIRICAL or NORMATIVE); the baseline category is therefore a subject in treatment NO INFO and an equity-based punishment scenario with lie size=1. Additional control variables include age, gender, education, and controls for how punishment scenarios were presented (increasing/decreasing) and the estimated implementation probability). 10 punishment scenarios per punisher; standard errors are clustered at the punisher level. */**/***: $p < 0.10/0.05/0.01$.

In the regression above, the dependent variable p_{is} is the share of punishment by subject i in scenario s , expressed as a number between 0 and 100. Consequently, the coefficients for the independent variables can be interpreted as percentage point shifts. The independent variables in (1) are the following: EMPIRICAL_i and NORMATIVE_i are dummies indicating whether subject i was in one of the norm nudge treatments (or in NO INFO, the baseline category), LieSize_s measures the size of the lie in punishment scenario s , while Inequity_s and Overclaiming_s are dummies indicating the equity norm of scenario s (with an Equity-scenario being the baseline). Age, gender, and education are collected in the **Controls** $_i$ -vector for each punisher. This vector also contains a dummy to control for the order in which scenarios were presented (counterbalanced over treatments), and punisher i 's estimate for the ratio of punishers to active players in Part 1 (i.e., the implementation probability). We then fit the above model using OLS using standard errors clustered on the subject level i .

Table 2 reports the regression results. The first three columns repeat our previous statistical analysis parametrically. In Column (1) we replicate the finding that larger lies lead to significantly more punishment by about 10 percentage points for each unit increase in the size of the lie. Column (2) shows that relative to Equity-based punishment scenarios, Inequity- and Overclaim-scenarios lead to significant increases in the share of punishment assigned by about 9 and 22 percentage points, respectively. We also find that the implied difference of 13 (=22-9) percentage points between the Inequity- and Overclaim-scenarios is statistically significant (F-test: $p < 0.001$). Column (3) shows that providing empirical and normative information respectively leads to a (marginally) significant increase in punishment of about 2.7 and 3.5 percentage points (these coefficients do not differ significantly from each other; F-test: $p = 0.629$)

The full model in Column (4) allows us to measure the effect of different equity scenarios while controlling for differing lie sizes in the associated scenarios where the underlying lies differ (i.e., in the Overclaiming-scenarios). This is particularly important in order to check the robustness of the results concerning Equity-norms in Column (2) as these results could also be due to the differing sizes of lies in the Inequity- and Overclaim-scenarios, respectively. Consistent with this caveat, we find that indeed the coefficient for the Overclaim-scenarios shrinks by about three quarters. In fact, the change by 17.1 punishment points in the Overclaim-coefficient captures the change in average size of the lie when controlling for the size of the lie (=22.2-5.1, comparing columns 3 and 4). It also corresponds very closely to the 1.7-increase in the average size of the lie (3.0 in Overclaim vs. 1.3 in the Equity- and Inequity scenarios) multiplied with the ‘‘Size of the Lie’’-coefficient of about 10. However, even when controlling for these larger lie sizes, we find that the Overclaim-scenarios yield a significant increase of 5.1 percentage points relative to the baseline Equity-scenarios. Thus, while smaller than initially suggested, we find support for Hypothesis 2b: (In)Equity concerns do not only matter when lying creates inequality but

also when it reverts pre-existing inequality that benefits a liar. Overall, the regression results (re-)confirm Hypotheses 1, 2, and 3.²³

Corroborating the behavioral experiment using a vignette experiment

To provide additional insights regarding the robustness and external validity of our main findings in the Punisher Experiment (for the Behavioral Experiment in Study 1), we conducted a vignette study with multiple treatments and random assignment. For our purposes, we chose the setting of the vignette such that it is closer to a real-life situation while simultaneously retaining the key elements of the behavioral experiment’s setting: accepting a bribe (\approx cheating) and whistleblowing (\approx punishing) in a corporate context. Our working assumption is that testing our main insights (that punishers are sensitive to both the motive of deviant behavior and the presence of norm information) in a different context and using another methodology is a first step for achieving external validity (for a related discussion, see [List, 2020](#)). It featured a general population sample of $n=225$ that was collected in person by various research assistants across 10 U.S. states. Subjects were randomly presented with one scenario in which they observed a co-worker taking a bribe and given the opportunity to blow the whistle on their co-worker by submitting incriminating evidence about the bribe-taking behavior. Subjects were able to make a selection between different forms of behavior (submitting varying amounts of evidence upon blowing the whistle) and thereby affecting the probability of the co-worker being punished (details are described in Appendix B).

In mirroring the treatment dimensions of the preceding Punisher Experiment (for the Behavioral Experiment in Study 1), the vignette experiment featured a 2(within) \times 2(between)-design. The within-subject conditions varied in a randomly determined order, in addition to the inequity nature of the co-worker taking the bribe: in one case it was creating EQUITY relative to the subject’s role (the co-worker had not, for reasons outside his/her responsibility, received an end-of-year bonus which the subject’s role had been given), in the other case, it was creating INEQUITY (since the co-worker did get the same bonus as the subject). In addition, there was a between-subject variation altering whether subjects were provided with a norm-nudge of the same kind as in the behavioral experiment’s NORMATIVE-condition: information regarding the fact that the majority of participants in a previous study objected to accepting the bribe. In the NO INFO-condition, this information was not provided. Appendix B features further details on the design of the vignette experiment and its result, which we report in summary below:

The results of the vignette experiment suggest that the core insights from a more stylized decision environment of the behavioral results from the preceding behavioral experiment carry over to a more applied environment. We find that providing norm information increases the

²³ In the appendix, we also provide an explanatory analysis of how the treatment effect from providing norm information interacts with the equity nature of the lie. We find that it is comparatively weaker in the Overclaiming-scenarios (see Table A.4 in Appendix A and the text preceding it).

average chosen probability of punishment over equity-scenarios overall (75.0% vs. 65.7% in INFO and NO INFO, respectively; $p \leq 0.044$ in regression and ranksum-tests). This effect of providing norm information appears in both, the INEQUITY-scenario (82.8% vs. 73.3%; $p < 0.005$) and the EQUITY scenario (67.2% vs. 57.9%; $p = 0.074$). Likewise, we also repeat our observation from the behavioral experiment in that, overall, the chosen average punishment for opportunistic behavior is higher if it creates INEQUITY rather than EQUITY (77.8% vs. 62.3%; $p < 0.001$). Together, these results confirm the findings of the behavioral experiment and show that the punishment patterns observed in this experiment are also observed in applied situations, such as whistle-blowing in the workplace.

Study 2: Norm Information and Norm Perception

The results from Study 1 suggest that people punish norm violations differently depending on the norm information that they receive and the equity nature of the lie. To better understand the mechanism of *why* norm enforcement varies, we examine whether the observed punishment patterns have an analogous variation in the perception of social norms. Investigating this is important: To achieve prolonged norm adherence, enforcement needs to reflect a shared (perceived) social norm and will be less effective if the norms are in conflict with formal rules or comprised of idiosyncratic judgments (Acemoglu and Jackson, 2017; Bicchieri et al., 2021).

However, prior research has shown that the provision of norm information can – but does not necessarily have to – change the perception of social norms (for a recent example where this approach succeeded see Bursztyjn et al., 2020b; and for an example of where it didn't, see Dimant et al., 2020). Thus, examining whether changes in punishment patterns are consistent with changes in social norm perceptions is in our context – an empirical question. We therefore devised Study 2 in which we vary, as in Study 1, the type of norm information and equity nature of lies. However, instead of eliciting punishment decisions, we elicit how the perception of social norms differs across these dimensions.²⁴

Design

In November 2019, we recruited a new set of $n=1,519$ subjects via MTurk to elicit their norm perceptions regarding the lying behavior as observed in the behavioral experiment of Study 1. While this follow-up experiment reflects key features of that experiment, we did not elicit punishment behavior here. Rather, we elicit perceptions regarding the normative appropriateness of lying for those scenarios using the incentive-compatible procedure by Krupka and Weber (2013).

Prior to the elicitation of their norm perceptions, participants were informed about the original lying task in the same way that it was explained to punishers in the behavioral experiment

²⁴As is customary in the norms literature, we elicit norm-related beliefs from participants who have not participated in Study 1 (this avoids, among others, priming- or demand effects and the generation of post-hoc justifications by subjects; see d'Adda et al., 2016 for a discussion).

of Study 1 (i.e., about the structure of Part 1 and Part 2; see Figure 1). Subsequently, each subject was presented with one of several lying situations that varied along two main dimensions. The first dimension was whether no norm information (NO INFO), normative information (NORMATIVE) or empirical information (EMPIRICAL) was provided on top of the observed lying behavior. The second dimension was the equity nature of the lie, which corresponds to three lying scenarios reflecting Equity-Lying (p13), Inequity-Lying (p35), and Overclaim-Lying (p24). Note that for comparability, the size of the lie was constant (at a size of 2).

This yields a between-subjects design that varies norm information and equity nature of a lie in a fully factorial manner over 3 (norm information) \times 3 (equity lying scenarios) treatments to which subjects were randomly assigned.²⁵ We measured our dependent variable of interest by asking participants to rate the extent to which other subjects deemed the observed lying behavior socially (in)appropriate. They did so using a 4-point Likert scale ranging from “very socially inappropriate” (*VSI*), “somewhat socially inappropriate” (*SSI*), “somewhat socially appropriate” (*SSA*), and “very socially appropriate” (*VSA*). In each treatment, participants were given a monetary incentive to guess the modal answer, thus allowing for incentive-compatible elicitation of norm perceptions.

Results

Figure 5 illustrates the distributions of social (in)appropriateness ratings, split by whether norm information is provided (columns) and by the equity nature of the lie (rows). We first examine the role of the (in)equity-scenarios, then the role of norm information, and lastly their interaction.

Equity nature of the lie: Aggregated over the norm-nudge information treatments (i.e., row-wise comparisons in Figure 5), the average norm appropriateness rating in the Equity-scenario is higher compared to ratings for the Inequity- and Overclaiming-scenarios. This pattern is also marginally significant, according to pairwise ranksum tests (Equity vs Inequity $p=0.063$, Equity vs Overclaiming $p<0.066$, Inequity vs Overclaiming $p=0.983$; all ranksum test reported here and in the following are two-sided). Thus, our results show that the perception of social norms across different equity settings reflect the punishment patterns we observed in the behavioral experiment of Study 1.

Norm information: Likewise, we find a similar effect for the norm nudge as in the behavioral experiment. In particular, we do not see a difference in responses between the two norm information treatments NORMATIVE and EMPIRICAL (ranksum test: $p=0.218$). Consequently,

²⁵For a breakdown of observations per treatment see Table A.5 in the appendix where we also describe subject characteristics and provide randomization checks. The average duration of the experiment was about four minutes. Participants were paid a show-up fee of \$0.20 with an opportunity to receive a \$0.20 bonus based on their answers and thus paid well above average MTurk pay (Hara et al., 2018). We employed the same participant pool restrictions used in the behavioral experiment of Study 1 (see Footnote 8).

we pool these two scenarios for the remainder of the analysis.²⁶ When we compare the different pooled equity-scenarios over the norm nudges (i.e., compare the two columns in Figure 5), we find that providing norm information (INFO) leads to a lower average social appropriateness rating of lying (ranksum-test: NO INFO vs. INFO, $p < 0.001$). Also, note that due to the between-subjects nature of the experiment, the norm responses in the three NO INFO treatments should also reflect the subjects’ beliefs in the INFO-treatments before they were given norm information (i.e., their ex-ante beliefs).

Interaction of norm information and equity nature: In NO INFO, one can observe that the perceived normative appropriateness is relatively heterogeneous across the various equity-scenarios. For the Equity-scenario, displayed in the first row, the perceived social norm towards lying is relatively forgiving, with *VSI*, *SSI*, and *SSA* each obtaining about one-third of the total ratings. In contrast, most subjects seem to be aware that lying in the Inequity-scenario (second row) is “Very socially inappropriate” (*VSI*). For the Overclaim-scenario (third row), the pattern is similar, even though less pronounced. In line with these observations, responses in the Inequity- and Overclaiming-scenarios do not differ significantly (ranksum test: $p = 0.261$). In contrast, responses in the Equity-scenario are (marginally) significantly different from responses in the Inequity- and Overclaim-scenarios (ranksum tests: $p = 0.003$ and $p = 0.061$, respectively).

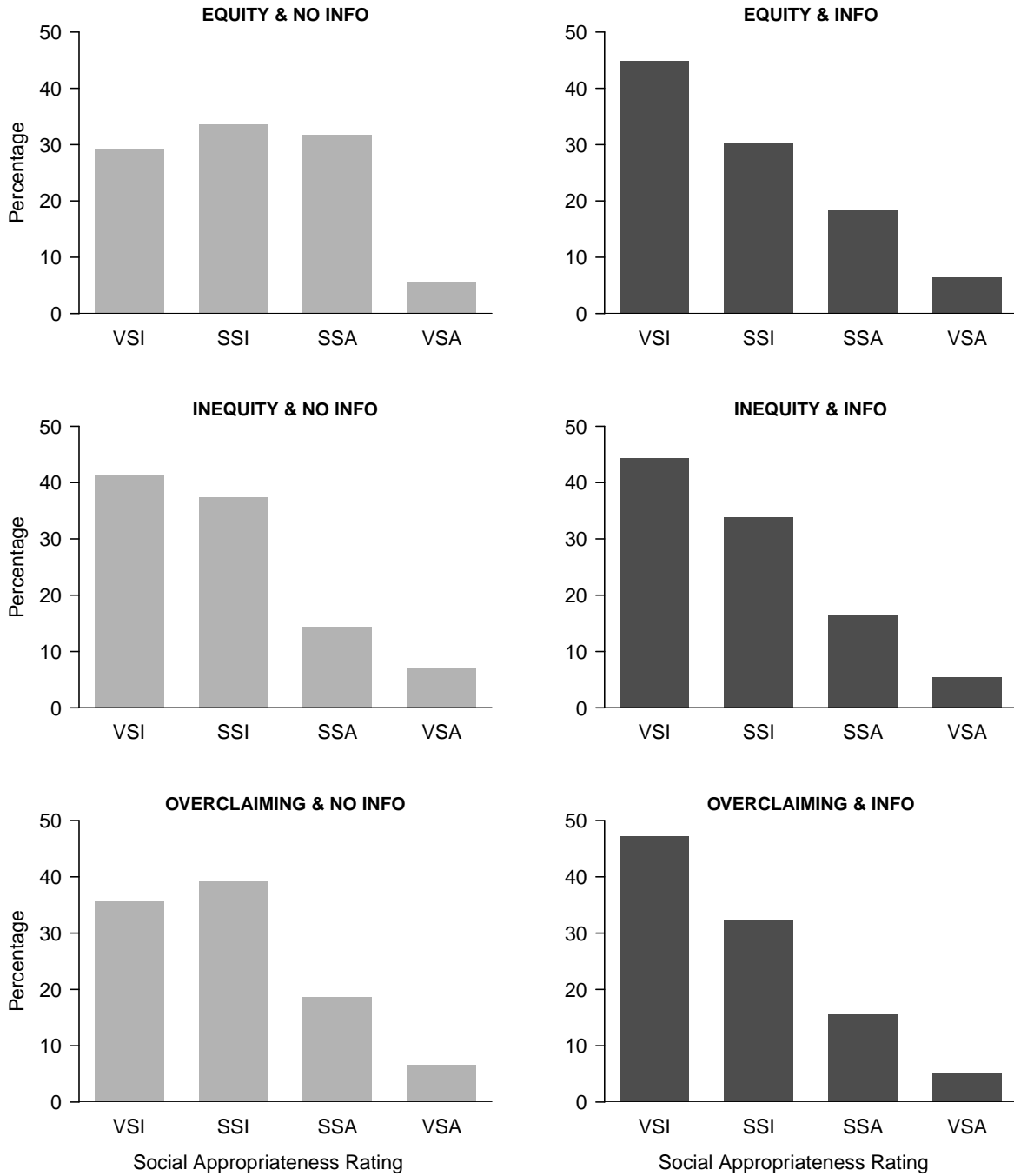
These patterns are different when norm-related information is provided (right column of Figure 5). Here, the distribution of perceived social appropriateness is the same, irrespective of the lie’s equity nature: Across all scenarios, “Very Socially Inappropriate” is the modal answer, whereas almost no one considers lying to be appropriate. This homogeneous pattern induced by norm information across equity conditions is also reflected in ranksum tests which do not indicate any significant differences ($p \geq 0.311$ for all pairwise comparisons of the three INFO-treatments).

Noteworthy, the provision of norm information helps people to coordinate: in the INFO treatments, the frequency of the modal category (here, *VSI*) is always substantial. In the NO INFO treatments, such coordination on the modal category is, overall, less pronounced. However, if we examine this effect of norm information conditional on the equity scenario, we see that norm nudges are effective when there is a conflicting social norm, such as when the liar is initially disadvantaged by chance. This manifests in the observation that the effect of norm nudges leads lying to be perceived as less acceptable than without norm information for the Equity-scenario (ranksum test: $p < 0.001$) and the Overclaiming-scenario (ranksum test: $p = 0.020$). In contrast, we do not find a significant difference in the case of Inequity-based lying, where the pre-existing norm against lying was already relatively strong (ranksum-test: $p = 0.645$).²⁷

²⁶Figure A.4 in Appendix A shows when the distributions of responses when NORMATIVE and EMPIRICAL are separately displayed; the results are largely the same. Specifically, we find that appropriateness ratings in EMPIRICAL and NORMATIVE are significantly higher than ratings in NO INFO (ranksum-tests: EMPIRICAL vs. NO INFO, $p = 0.012$; NORMATIVE vs. NO INFO, $p < 0.001$).

²⁷These findings are also supported by a regression analysis in Table A.6 in Appendix A. It shows a concentration of the effect of norm info on appropriateness ratings in the Equity-scenario and no effects in the Inequity-

Figure 5: Social appropriateness of lying over different norm information and equity situations



Notes: Norm information varies by column, and the equity nature of the lie varies by rows. The social norm is measured via a 4-items Likert scale ranging over “Very Socially Inappropriate” (*VSI*), “Somewhat Socially Inappropriate” (*SSI*), “Somewhat Socially Appropriate” (*SSA*), “Very Socially Appropriate” (*VSA*). INFO pools treatments the norm information treatments NORMATIVE and EMPIRICAL.

Discussion

In Study 1, we investigate the drivers of norm enforcement in the context of lying. Across different norm information settings, participants observe the behavior of liars who differ in how much they lie (as measured by the difference between stated and actual outcomes) and the equity-consequences of the lie. We find that punishment is higher for larger lies and lies that increase inequity for the liar. In contrast, punishment is less severe when the lie serves the benefit of offsetting an ex-ante imbalance. These results are also confirmed in a vignette study which looks at punishment of bribe-taking behavior.

Study 2 investigates the reasons for *why* norm enforcement varies across the settings that subjects considered in the first experiment, using a separate online sample to elicit social norm perceptions across these lying settings. We observe that the norm perception resembles the punishment patterns in the first experiment. That is, our results show that inequity-based lies are perceived to be less acceptable than equity-based lies. Study 1 also shows that providing norm information leads to increased punishment, independently of whether this information is empirically grounded in what other actually do or what they say *should* be done. Again, this is also reflected in the elicited norm perceptions in Study 2, showing that there is a close link between variations in punishment and norm perception.²⁸ The results show that providing norm information adjust the perceived acceptability of lying downwards, independently of the nature of the norm information. Furthermore, the results in this experiment indicate that the downward-shift in the perceived acceptability of lying works primarily through a change in the acceptability of equity-based lies. These results are consistent with the notion of 'tightness' and 'looseness' of norms (Gelfand et al., 2011; Dimant et al., 2021a) in that the perception of how socially (in)appropriate lying is in the 'equity & info' condition is more concentrated (and thus exhibits a clearer and tighter norm) than in the 'equity & no info' condition (that more closely resembles the initial norm perception that participants bring into our experiment). What we show, and to the best knowledge for the first time, is that nudging in a 'looser' norm environment yields stronger shifts of norm perceptions than in 'tighter' norm environments.

Taken together, the main insights derived from the results of our experiments are that norm-nudges such as providing norm information do, in principle, foster norm enforcement. A deeper analysis of our findings suggests that the effect of such information occurs through shifting norm perceptions. This applies in particular for equity-based lies, for which norm perceptions were the most dispersed without norm information. Together, this suggests that

and Overclaiming-scenarios. This is consistent with the result from the interaction analysis for the behavioral experiment of Study 1 where punishment was found to be particularly weak for the Overclaiming-scenarios (but not weaker in the Inequity-scenarios, see Table A.4 and Footnote 23).

²⁸Recall that the behavioral experiment in Study 1 is a within-subjects design with respect to the different lying scenarios, whereas Study 2 varies all treatment dimensions between subjects. The fact that results are nevertheless very consistent across these independent experiments suggests that our results are robust to both of these design choices (see also Clifford et al., 2020, for a recent discussion on the robustness of within-designs).

norm nudges also work by reducing normative uncertainty for punishers. Conversely, we show in Study 2 that where clear norms exist but are not necessarily honoured, norm-nudges via norm information are less effective. This finding is consistent with existing research such as the seminal study by Goldstein et al. (2008) investigating the impact of norm information provision on environmentally friendly behavior (towel reuse in hotels) in the United States. Importantly, however, a conceptual replication by Bohner and Schlüter (2014) failed to replicate this in Germany. The authors argue that this is likely due to the already existing strong norm of environmental preservation. Additional norm information was uninformative and thus yielded no behavior change (see Dimant et al. 2020 for related findings in the context of lying).

It is important to note that our experimental design and the results coming from it can capture key considerations and consequences for norm-enforcers and violators in applied settings. In practice, social norms are often not very clear-cut and in conflict with other norms, similar to lying in our inequity-scenarios. Furthermore, even though punishment in natural contexts is often non-pecuniary (e.g. in the form of disapproving gazes and shaming; see Coricelli et al., 2010; Eriksson et al., 2021), it often comes with reputational and pecuniary costs (e.g., from becoming socially ostracized or being excluded from cooperative interactions; see Bolton et al., 2005). In fact, Masclet et al. (2003) show that monetary and non-monetary enforcement are comparable in their effectiveness for sustaining cooperation. Given that our key experimental results in the context of the stylized framing replicated in the arguably more applied framing of the vignette, we are therefore optimistic that the findings from our experimental setup can yield valuable insights for the design and evaluation of “real-world”-policies and norm nudges.

Conclusion

Enforcing social norms is central to the fabric of a functioning society, and this foundation is undermined when transgression goes unpunished (Bicchieri, 2006). While individually costly, the society’s collective gains can be substantial through facilitating coordination and norm adherence (see Xiao, 2018, for a recent review). While much of the existing literature has studied the *effect* of punishment in social interactions and its ability to uphold social norms, less is known about the *drivers* of such punishment. It is therefore imperative to understand the circumstances for which an observed norm-transgression is punished and the role that the motives behind the transgression play. Our contribution is to highlight key aspects of norm enforcement: its multi-layered nature, the impact of norm-nudges, and the link to perception of norms.

We analyze the extent to which norm enforcement is sensitive to the consequences of the observed transgression (e.g., achieving equity versus achieving an unfair advantage), as well as the type of norm-nudges (empirical versus normative). We examine questions through the lens of three experiments: The first two experiments aim to measure norm-*enforcement* behavior, whereas the third experiment aims to measure the associated norm-*perceptions* of the observed lies. This approach allows us not only to understand how norm transgressions are punished,

but also the extent to which variations in norm enforcement align with variations in norm perceptions. Indeed, we find that both map onto each other very well.

From a policy perspective, our results emphasize the necessity of capturing varying social motives of both transgressors and norm enforcers, especially when “soft” interventions such as nudges are used to achieve behavioral change. Existing work has shown that norm-based interventions such as norm-nudges need to be meticulously mapped onto the social environment in which they are implemented. Taking this into account, our results also inform the ongoing scholarly debate regarding reasons why prior studies that used such soft norm-nudge interventions have had mixed success (Fellner et al., 2013; Hallsworth et al., 2017; Dimant et al., 2020; Gelfand et al., 2021). For example, our findings suggest that norm-nudges exhibit the largest impact when pre-existing norm perceptions are rather inconclusive (that is, where our norm-elicitation measure did not identify a uniquely prevailing norm, as in Study 2’s Equity & NoInfo setting). This is in-line with research showing that – even if individuals are aware of a norm – moral wiggle room in norm perceptions can weaken norm compliance because it can be exploited in self-serving ways (Di Tella et al., 2015; Bicchieri et al., 2020b; Dimant et al., 2020). From an applied perspective – and in line with the libertarian paternalism approach (Thaler and Sunstein, 2021) –, the nudge interventions evaluated in our paper could be implemented by those who intervene (e.g., government bodies) by strategically targeting those with norm enforcement power with the relevant social information. Given the observed responsiveness of norm enforcers to such interventions, our study adds an additional layer to what choice architects can do to achieve behavior change (Ambuehl et al., 2021). This refers to, for example, environments in which norm enforcers are feed information about factually true descriptive or injunctive norms.

We help to advance this scholarly debate by pointing towards what works while also highlighting the reasons for unsuccessful nudging. On the one hand, our results suggest that norm enforcers unhesitatingly utilize punishment in a way that is sensitive to the motives underlying a norm breach. Thus, a “hands-off” approach can be warranted as long as norm following is self-enforcing through peer punishment. On the other hand, we find that using simple norm-nudges does not necessarily change the norm perception of the transgression – and the extent to which it is punished – if a norm already exists firmly. The implication of our findings is that norms in peer groups should be strengthened to sustain their enforcement. We also show that this can happen via simple messages that raise awareness for the applicable norm. Together with the opportunity to punish – as in our experimental setting – this can then be a promising approach to facilitate enforcement of norm compliance, especially when conformity within a peer group alone is not enough. However, we also show that in order to achieve behavioral change and foster norm enforcement, one first needs to better understand the context in which the intended behavior should occur and the existing norms. One may then be able to rely on gentle norm-nudges, as studied here; stronger, nudge-adjacent interventions such as *shoves* and *boosts*; or even explicit economic incentives (Grüne-Yanoff and Hertwig, 2016; Gino et al., 2019).

References

- Abeler, J., Becker, A., and Falk, A. (2014). Representative evidence on lying costs. *Journal of Public Economics*, 113:96–104.
- Abeler, J., Nosenzo, D., and Raymond, C. (2019). Preferences for truth-telling. *Econometrica*.
- Acemoglu, D. and Jackson, M. O. (2017). Social norms and the enforcement of laws. *Journal of the European Economic Association*, 15(2):245–295.
- Albrecht, F., Kube, S., and Traxler, C. (2018). Cooperation and norm enforcement—the individual-level perspective. *Journal of Public Economics*, 165:1–16.
- Allcott, H. (2011). Social norms and energy conservation. *Journal of Public Economics*, 95(9-10):1082–1095.
- Allcott, H. and Kessler, J. B. (2019). The welfare effects of nudges: A case study of energy use social comparisons. *American Economic Journal: Applied Economics*, 11(1):236–76.
- Ambuehl, S., Bernheim, B. D., and Ockenfels, A. (2021). What motivates paternalism? an experimental study. *American Economic Review*, 111(3):787–830.
- Andreoni, J. and Gee, L. K. (2012). Gun for hire: Delegated enforcement and peer punishment in public goods provision. *Journal of Public Economics*, 96(11-12):1036–1046.
- Arechar, A. A., Gächter, S., and Molleman, L. (2018). Conducting interactive experiments online. *Experimental Economics*, 21(1):99–131.
- Balafoutas, L. and Nikiforakis, N. (2012). Norm enforcement in the city: A natural field experiment. *European Economic Review*, 56(8):1773–1785.
- Balafoutas, L., Nikiforakis, N., and Rockenbach, B. (2014). Direct and indirect punishment among strangers in the field. *Proceedings of the National Academy of Sciences*, 111(45):15924–15927.
- Balafoutas, L., Nikiforakis, N., and Rockenbach, B. (2016). Altruistic punishment does not increase with the severity of norm violations in the field. *Nature Communications*, 7.
- Benartzi, S., Beshears, J., Milkman, K. L., Sunstein, C. R., Thaler, R. H., Shankar, M., Tucker-Ray, W., Congdon, W. J., and Galing, S. (2017). Should governments invest more in nudging? *Psychological Science*, 28(8):1041–1055.
- Beshears, J. and Kosowsky, H. (2020). Nudging: Progress to date and future directions. *Organizational Behavior and Human Decision Processes*, 161:3–19.
- Bhanot, S. P. (2018). Isolating the effect of injunctive norms on conservation behavior: New evidence from a field experiment in california. *Organizational Behavior and Human Decision Processes*.
- Bicchieri, C. (2006). *The grammar of society: The nature and dynamics of social norms*. Cambridge University Press.
- Bicchieri, C. and Dimant, E. (2019). Nudging with care: The risks and benefits of social information. *Public Choice*, pages 1–22.
- Bicchieri, C., Dimant, E., Gaechter, S., and Nosenzo, D. (2020a). Observability, social proximity, and the erosion of norm compliance. Working Paper Available at SSRN: <https://dx.doi.org/10.2139/ssrn.3355028>.
- Bicchieri, C., Dimant, E., and Sonderegger, S. (2020b). It’s not a lie if you believe the norm does not apply: Conditional norm-following with strategic beliefs. Working Paper Available at SSRN: <https://dx.doi.org/10.2139/ssrn.3326146>.

- Bicchieri, C., Dimant, E., and Xiao, E. (2021). Deviant or wrong? the effects of norm information on the efficacy of punishment. *Journal of Economic Behavior & Organization*, 188:209–235.
- Bicchieri, C. and Xiao, E. (2009). Do the right thing: But only if others do so. *Journal of Behavioral Decision Making*, 22(2):191–208.
- Bohner, G. and Schlüter, L. E. (2014). A room with a viewpoint revisited: descriptive norms and hotel guests’ towel reuse behavior. *PloS one*, 9(8):e104086.
- Bolton, G., Dimant, E., and Schmidt, U. (2021). Observability and social image: On the robustness and fragility of reciprocity. Working Paper Available at SSRN: <https://dx.doi.org/10.2139/ssrn.3294375>.
- Bolton, G. E., Katok, E., and Ockenfels, A. (2005). Cooperation among strangers with limited information about reputation. *Journal of Public Economics*, 89(8):1457–1468.
- Bolton, G. E. and Ockenfels, A. (2000). Erc: A theory of equity, reciprocity, and competition. *American Economic Review*, 90(1):166–193.
- Bortolotti, S., Soraperra, I., Sutter, M., and Zoller, C. (2017). Too lucky to be true-fairness views under the shadow of cheating. Working Paper Available at SSRN: <https://ssrn.com/abstract=3014734>.
- Bott, K. M., Cappelen, A. W., Sorensen, E., and Tungodden, B. (2020). You’ve got mail: A randomised field experiment on tax evasion. *Management Science*, 66:2801–2819.
- Brandon, A., Ferraro, P. J., List, J. A., Metcalfe, R. D., Price, M. K., and Rundhammer, F. (2017). Do the effects of social nudges persist? theory and evidence from 38 natural field experiments. National Bureau of Economic Research Working Paper.
- Brandts, J. and Charness, G. (2011). The strategy versus the direct-response method: a first survey of experimental comparisons. *Experimental Economics*, 14(3):375–398.
- Brouwer, T., Galeotti, F., and Villeval, M. C. (2020). Teaching norms in the streets. Mimeo.
- Buccioli, A. and Piovesan, M. (2011). Luck or cheating? a field experiment on honesty with children. *Journal of Economic Psychology*, 32(1):73–78.
- Bujold, P. M., Williamson, K., and Thulin, E. (2020). The science of changing behavior for environmental outcomes: A literature review. Rare Center for Behavior the Environment and the Scientific and Technical Advisory Panel to the Global Environment Facility.
- Bursztyn, L., Egorov, G., and Fiorin, S. (2020a). From extreme to mainstream: the erosion of social norms unravel. *American Economic Review*, forthcoming.
- Bursztyn, L., González, A. L., and Yanagizawa-Drott, D. (2020b). Misperceived social norms: Women working outside the home in saudi arabia. *American Economic Review*, 110(10):2997–3029.
- Bursztyn, L., Haaland, I., Rao, A., and Roth, C. (2020c). Disguising prejudice: Popular rationales as excuses for intolerant expression. *University of Chicago, Becker Friedman Institute for Economics Working Paper*, (2020-73).
- Choi, D. D., Poertner, M., and Sambanis, N. (2019). Parochialism, social norms, and discrimination against immigrants. *Proceedings of the National Academy of Sciences*, 116(33):16274–16279.
- Cialdini, R. B. and Goldstein, N. J. (2004). Social influence: Compliance and conformity. *Annual Review of Psychology*, 55:591–621.
- Cialdini, R. B., Reno, R. R., and Kallgren, C. A. (1990). A focus theory of normative conduct: recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology*, 58(6):1015–1026.
- Clifford, S., Sheagley, G., and Piston, S. (2020). Increasing precision in survey experiments without introducing bias. *APSA Preprints*. doi: 10.33774/apsa-2020-tbd3c.

- Coffman, L. C. (2011). Intermediation reduces punishment (and reward). *American Economic Journal: Microeconomics*, 3(4):77–106.
- Coleman, J. S. (1994). *Foundations of social theory*. Harvard University Press.
- Coricelli, G., Montmarquette, C., and Villeval, M. C. (2010). Cheating, emotions, and rationality: an experiment on tax evasion. *Experimental Economics*, 11(3):226–247.
- d’Adda, G., Drouvelis, M., and Nosenzo, D. (2016). Norm elicitation in within-subject designs: Testing for order effects. *Journal of Behavioral and Experimental Economics*, 62:1–7.
- Damgaard, M. T. and Gravert, C. (2018). The hidden costs of nudging: Experimental evidence from reminders in fundraising. *Journal of Public Economics*, 157:15–26.
- De Quidt, J., Haushofer, J., and Roth, C. (2018). Measuring and bounding experimenter demand. *American Economic Review*, 108(11):3266–3302.
- DellaVigna, S. and Linos, E. (2020). Rcts to scale: Comprehensive evidence from two nudge units. Technical report, Working Paper, UC Berkeley.
- Di Tella, R., Perez-Truglia, R., Babino, A., and Sigman, M. (2015). Conveniently upset: Avoiding altruism by distorting beliefs about others’ altruism. *American Economic Review*, 105(11):3416–42.
- Dickinson, D. L., Masclet, D., and Villeval, M. C. (2015). Norm enforcement in social dilemmas: An experiment with police commissioners. *Journal of Public Economics*, 126:74 – 85.
- Dimant, E. (2019). Contagion of pro-and anti-social behavior among peers and the role of social proximity. *Journal of Economic Psychology*, 73:66–88.
- Dimant, E. (2021). Hate trumps love: The impact of political polarization on social preferences. Working Paper Available at SSRN: <https://dx.doi.org/10.2139/ssrn.3680871>.
- Dimant, E., Gelfand, M., Hochleitner, A., and Sonderegger, S. (2021a). Strategic behavior with tight and loose norms. Mimeo.
- Dimant, E., Gerben, A. v. K., and Shalvi, S. (2020). Requiem for a nudge: Framing effects in nudging honest. *Journal of Economic Behavior & Organization*, 172:247–266.
- Dimant, E., Pieper, D., Clemente, E. G., Dreber, A., and Gelfand, M. J. (2021b). Politicizing mask-wearing: Predicting the success of behavioral interventions among republicans and democrats. Working Paper Available at SSRN: <https://ssrn.com/abstract=3915256>.
- Dufwenberg, M. and Dufwenberg, M. A. (2018). Lies in disguise—a theoretical analysis of cheating. *Journal of Economic Theory*, 175:248–264.
- Eriksson, K., Strimling, P., Gelfand, M., Van Lange, P., de Barra, M., et al. (2021). Perceptions of the appropriate response to norm violation in 57 societies.
- Exadaktylos, F., Espín, A. M., and Branas-Garza, P. (2013). Experimental subjects are not different. *Scientific Reports*, 3(1):1–6.
- Falk, A. and Heckman, J. J. (2009). Lab experiments are a major source of knowledge in the social sciences. *science*, 326(5952):535–538.
- Feess, E., Schildberg-Hörisch, H., Schramm, M., and Wohlschlegel, A. (2018). The impact of fine size and uncertainty on punishment and deterrence: Theory and evidence from the laboratory. *Journal of Economic Behavior & Organization*, 149:58–73.
- Fehr, E. and Fischbacher, U. (2004). Third-party punishment and social norms. *Evolution and Human Behavior*, 25(2):63–87.

- Fehr, E. and Gächter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review*, 90(4):980–994.
- Fehr, E. and Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, 114(3):817–868.
- Fehr, E. and Schurtenberger, I. (2018). Normative foundations of human cooperation. *Nature Human Behaviour*, 2(7):458.
- Fellner, G., Sausgruber, R., and Traxler, C. (2013). Testing enforcement strategies in the field: Threat, moral appeal and social information. *Journal of the European Economic Association*, 11(3):634–660.
- Ferraro, P. J., Miranda, J. J., and Price, M. K. (2011). The persistence of treatment effects with norm-based policy instruments: evidence from a randomized environmental policy experiment. *American Economic Review*, 101(3):318–22.
- Fischbacher, U. and Föllmi-Heusi, F. (2013). Lies in disguise—an experimental study on cheating. *Journal of the European Economic Association*, 11(3):525–547.
- Galeotti, F., Maggian, V., and Villeval, M. C. (2021). Fraud deterrence institutions reduce intrinsic honesty. *The Economic Journal*, 131(638):2508–2528.
- Galeotti, F., Saucet, C., and Villeval, M. C. (2020). Unethical amnesia responds more to instrumental than to hedonic motives. *Proceedings of the National Academy of Sciences*, 117(41):25423–25428.
- Gelfand, M., Li, R., Stankou, E., Pieper, D., Denison, E., Fernandez, J., Choi, V. K., Chatman, J., Jackson, J. C., and Dimant, E. (2021). Persuading republicans and democrats to comply with mask wearing: An intervention tournament. Working Paper Available at SSRN: <https://dx.doi.org/10.31234/osf.io/6gjh8>.
- Gelfand, M. J., Raver, J. L., Nishii, L., Leslie, L. M., Lun, J., Lim, B. C., Duan, L., Almaliach, A., Ang, S., Arnadottir, J., et al. (2011). Differences between tight and loose cultures: A 33-nation study. *science*, 332(6033):1100–1104.
- Gino, F., Hauser, O. P., and Norton, M. I. (2019). Budgeting beliefs, nudging behaviour. *Mind & Society*, pages 1–12.
- Gino, F. and Pierce, L. (2010). Lying to level the playing field: Why people may dishonestly help or hurt others to create equity. *Journal of Business Ethics*, 95(1):89–103.
- Gneezy, U., Kajackaite, A., and Sobel, J. (2018). Lying aversion and the size of the lie. *American Economic Review*, 108(2):419–53.
- Goldstein, N. J., Cialdini, R. B., and Griskevicius, V. (2008). A room with a viewpoint: Using social norms to motivate environmental conservation in hotels. *Journal of Consumer Research*, 35(3):472–482.
- Grewenig, E., Lergetporer, P., and Werner, K. (2020). Gender norms and labor-supply expectations: Experimental evidence from adolescents. Technical report, CESifo Working Paper.
- Grüne-Yanoff, T. and Hertwig, R. (2016). Nudge versus boost: How coherent are policy and theory? *Minds and Machines*, 26(1-2):149–183.
- Haaland, I. and Roth, C. (2020). Labor market concerns and support for immigration. *Journal of Public Economics*, 191:104256.
- Hagmann, D., Ho, E. H., and Loewenstein, G. (2019). Nudging out support for a carbon tax. *Nature Climate Change*, 9(6):484–489.
- Hallsworth, M. and Kirkman, E. (2020). *Behavioral insights*. MIT Press.
- Hallsworth, M., List, J. A., Metcalfe, R. D., and Vlaev, I. (2017). The behavioralist as tax collector: Using natural field experiments to enhance tax compliance. *Journal of Public Economics*, 148:14–31.

- Hara, K., Adams, A., Milland, K., Savage, S., Callison-Burch, C., and Bigham, J. P. (2018). A data-driven analysis of workers’ earnings on amazon mechanical turk. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, page 449. ACM.
- Heffner, J. and FeldmanHall, O. (2019). Why we don’t always punish: Preferences for non-punitive responses to moral violations. *Scientific Reports*, 9(1):1–13.
- Herrmann, B., Thöni, C., and Gächter, S. (2008). Antisocial punishment across societies. *Science*, 319(5868):1362–1367.
- Hummel, D. and Maedche, A. (2019). How effective is nudging? a quantitative review on the effect sizes and limits of empirical nudging studies. *Journal of Behavioral and Experimental Economics*, 80:47–58.
- John, L. K., Loewenstein, G., and Rick, S. I. (2014). Cheating more for less: Upward social comparisons motivate the poorly compensated to cheat. *Organizational Behavior and Human Decision Processes*, 123(2):101–109.
- Khadjavi, M., Sipangule, K., and Thiele, R. (2021). Social capital and large-scale agricultural investments: An experimental investigation. *The Economic Journal*, 131(633):420–449.
- Kimbrough, E. O. and Vostroknutov, A. (2016). Norms make preferences social. *Journal of the European Economic Association*, 14(3):608–638.
- Krupka, E. L. and 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.
- Levitt, S. D. and List, J. A. (2007). What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic Perspectives*, 21(2):153–174.
- List, J. A. (2011). Why economists should conduct field experiments and 14 tips for pulling one off. *Journal of Economic perspectives*, 25(3):3–16.
- List, J. A. (2020). Non est disputandum de generalizability? a glimpse into the external validity trial. Technical report, National Bureau of Economic Research.
- Martin, J. W., Jordan, J. J., Rand, D. G., and Cushman, F. (2019). When do we punish people who don’t? *Cognition*, 193:104040.
- Masclot, D., Noussair, C., Tucker, S., and Villeval, M.-C. (2003). Monetary and nonmonetary punishment in the voluntary contributions mechanism. *American Economic Review*, 93(1):366–380.
- Maxwell, S. E., Kelley, K., and Rausch, J. R. (2008). Sample size planning for statistical power and accuracy in parameter estimation. *Annual Review of Psychology*, 59.
- Nikiforakis, N. (2008). Punishment and counter-punishment in public good games: Can we really govern ourselves? *Journal of Public Economics*, 92(1-2):91–112.
- Nikiforakis, N. and Normann, H.-T. (2008). A comparative statics analysis of punishment in public-good experiments. *Experimental Economics*, 11(4):358–369.
- Ostrom, E. (2000). Collective action and the evolution of social norms. *Journal of Economic Perspectives*, 14(3):137–158.
- Schultz, P. W., Nolan, J. M., Cialdini, R. B., Goldstein, N. J., and Griskevicius, V. (2007). The constructive, destructive, and reconstructive power of social norms. *Psychological Science*, 18(5):429–434.
- Snowberg, E. and Yariv, L. (2021). Testing the waters: Behavior across participant pools. *American Economic Review*, 111(2):687–719.

- Stankou, E., van Kleef, G. A., Homan, A. C., Gelfand, M. J., van de Vijver, F. J., van Egmond, M. C., Boer, D., Phiri, N., Ayub, N., Kinias, Z., et al. (2019). Cultural collectivism and tightness moderate responses to norm violators: Effects on power perception, moral emotions, and leader support. *Personality and Social Psychology Bulletin*, 45(6):947–964.
- Sunstein, C. R. (2016). Do people like nudges. *Admin. L. Rev.*, 68:177.
- Sutter, M., Haigner, S., and Kocher, M. G. (2010). Choosing the carrot or the stick? Endogenous institutional choice in social dilemma situations. *The Review of Economic Studies*, 77(4):1540–1566.
- Tankard, M. E. and Paluck, E. L. (2016). Norm perception as a vehicle for social change. *Social Issues and Policy Review*, 10(1):181–211.
- Thaler, R. H. and Sunstein, C. R. (2009). *Nudge: Improving decisions about health, wealth, and happiness*. Penguin.
- Thaler, R. H. and Sunstein, C. R. (2021). *Nudge: The final edition*. Penguin.
- Winter, F. and Zhang, N. (2018). Social norm enforcement in ethnically diverse communities. *Proceedings of the National Academy of Sciences*, 115(11):2722–2727.
- Xiao, E. (2018). Punishment, social norms, and cooperation. In *Research Handbook on Behavioral Law and Economics*. Edward Elgar Publishing.
- Young, H. P. (2016). *Social Norms*, pages 1–7. Palgrave Macmillan UK, London.
- Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. *Experimental Economics*, 13(1):75–98.

Norm Makers and Norm Breakers:
How Motives for Deviance Shape Punishment of Norm Violations

Online Appendix

Contents:

A: Additional figures and data analysis

B: Details on the Vignette Experiment in Study 1

C: Information regarding pre-registration

D: Instructions

1. Instructions for the Liar Experiment in Study 1
2. Instructions for the Behavioral Experiment (Punishers) in Study 1
3. Instructions for the Vignette Experiment in Study 1
4. Instructions for the Norm Elicitation Experiment in Study 2

Appendix A: Additional figures, tables, and data analysis

Additional analyses for the Behavioral Experiment in Study 1

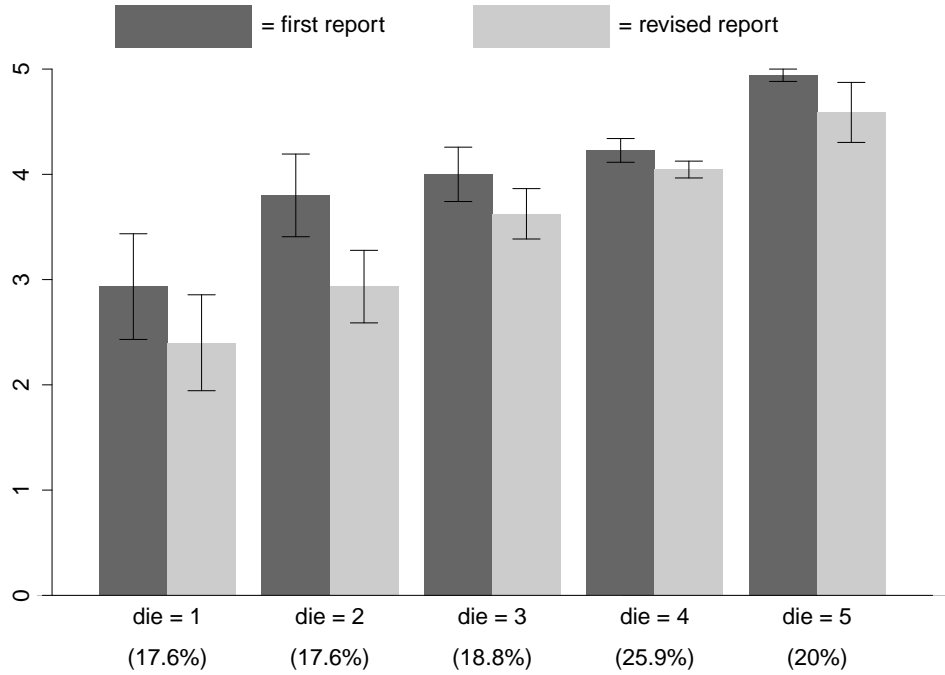
Table A.1 below provides a systematic breakdown of the various scenarios studied in our paper and clarifies the relationship between punishment scenario, actual outcome, reported outcome, size of the lie, and the implied equity nature of the lie. This helps to put our behavioral predictions and results in context.

Table A.1: Dimensions of punishment scenarios

Punishment scenario	Actual outcome	Reported outcome	Size of the lie	equity nature
p12	1	2	1	equity
p13	1	3	2	equity
p14	1	4	3	overclaiming
p15	1	5	4	overclaiming
p23	2	3	1	equity
p24	2	4	2	overclaiming
p25	2	5	3	overclaiming
p34	3	4	1	inequity
p35	3	5	2	inequity
p45	4	5	1	inequity

Figure A.1 provides a systematic breakdown of a **liar's** behavior (both first report and revised report) conditional on the random outcome of the die toss. We can observe a consistent revision pattern in that liars are less likely to misreport the number after being made aware of the punishment risk.

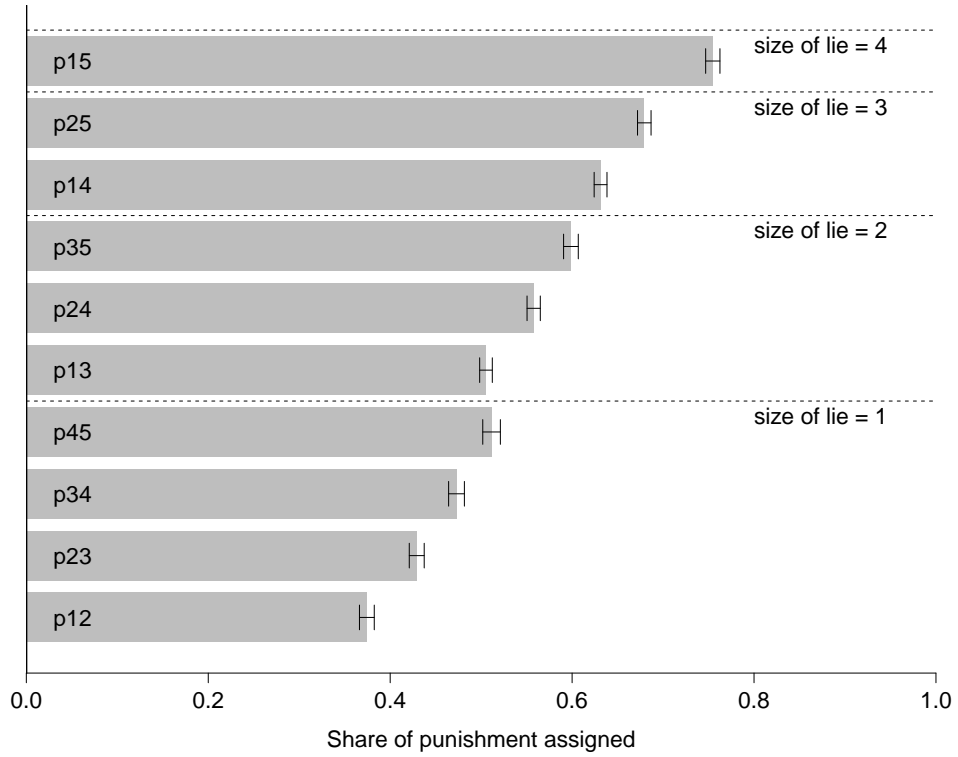
Figure A.1: Liars' first and revised reports by actual result of the die



Notes: Means of first and revised reports for the outcome of the 5-sided die toss, grouped by the actual outcome of the die toss (frequency for each outcome in parentheses). Error bars denote SEM.

Figure A.2 a systematic breakdown of a **punisher's** behavior conditional on the punishment scenario and the size of the lie, as illustrated in Figure A.1. We observe a clear pattern in that the share of assigned punishment increases with the size of the lie.

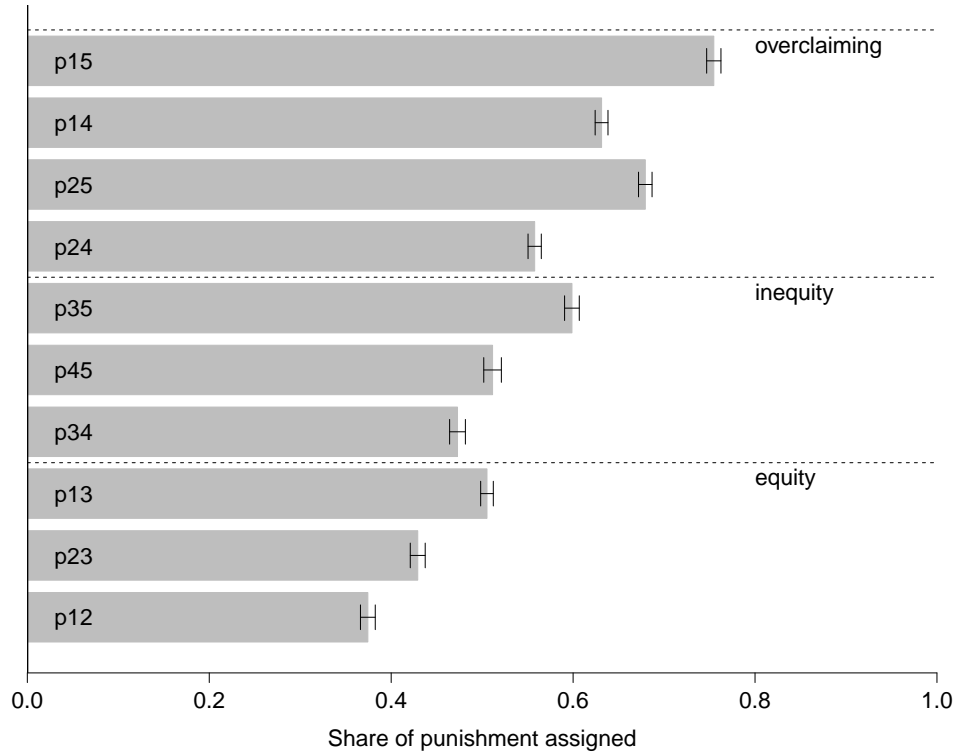
Figure A.2: Punishment for each punishment scenario, ordered by size of the lie



Notes: Punishment assigned as a share of total punishment points available in each punishment scenarios, ordered by the size of the lie. Error bars denote SEM.

Figure A.3 illustrates the results from Figure A.2 in a different way. In particular, the figure below breaks down the share of assigned punishment conditional on the equity nature of the lie (as defined in Table A.1). The results suggest that lying to overclaim (achieve equality) is punished the most (least), which is consistent with our results from Study 2.

Figure A.3: Punishment for each punishment scenario, ordered by equity norm



Notes: Punishment assigned as a share of total punishment points available in each punishment scenarios, ordered by equity nature of the lie. Error bars denote SEM.

Table A.2: Descriptive statistics for active players

	Mean	S.D.
Age	36.835	10.194
Male	0.682	0.468
Edu.: some high school	0.012	0.108
Edu.: finished high school	0.118	0.324
Edu.: some college	0.341	0.477
Edu.: finished college	0.447	0.500
Edu.: higher degree	0.082	0.277
observations	N=	85

Notes: Personal characteristics of active players in the pre-experiment (mean and standard deviation).

Table A.3: Descriptive Statistics for punishers by norm information treatments

	NO INFO		NORMATIVE		EMPIRICAL	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
Age	43.545	13.478	42.552	14.122	43.303	14.089
Male	0.468	0.500	0.477	0.500	0.482	0.500
Edu.: some high school	0.037	0.189	0.039	0.193	0.024	0.152
Edu.: finished high school	0.163	0.370	0.245	0.430	0.194	0.396
Edu.: some college	0.280	0.449	0.257	0.437	0.277	0.448
Edu.: finished college	0.300	0.459	0.293	0.456	0.314	0.465
Edu.: higher degree	0.208	0.406	0.160	0.367	0.175	0.380
observations	N=	404	N=	413	N=	423

Notes: Demographic characteristics of punishers in the Punisher Experiment (for the Behavioral Experiment in Study 1), by norm-information treatment (mean and standard deviation). We do not observe statistically significant differences between the treatments for age (Kruskal-Wallis test: $p=0.548$), gender (χ^2 -test: $p=0.916$), or education (χ^2 -test: $p=0.190$).

Table A.4 (next page) follows regression equation (1). Instead of using the $EMPIRICAL_i$ - and $NORMATIVE_i$ -dummies, however, it uses a an $INFO_i$ -dummy, with different meanings across three specifications covering different data:

1. Only data from treatments the baseline NO INFO ($INFO_i = 0$) and EMPIRICAL ($INFO_i = 1$).
2. Only data from treatments the baseline NO INFO ($INFO_i = 0$) and NORMATIVE ($INFO_i = 1$).
3. Data from treatments NO INFO ($INFO_i = 0$) and both NORMATIVE and EMPIRICAL ($INFO_i = 1$); the latter two treatment are therefore pooled.

In addition, it also features interactions of the $INFO_i$ -dummy and the $Inequity_i$ - and $Overclaiming_i$ -dummies. The three columns in Table A.4 correspond to the three specifications above.

Table A.4: Differential effect of norm info on punishment – regression results

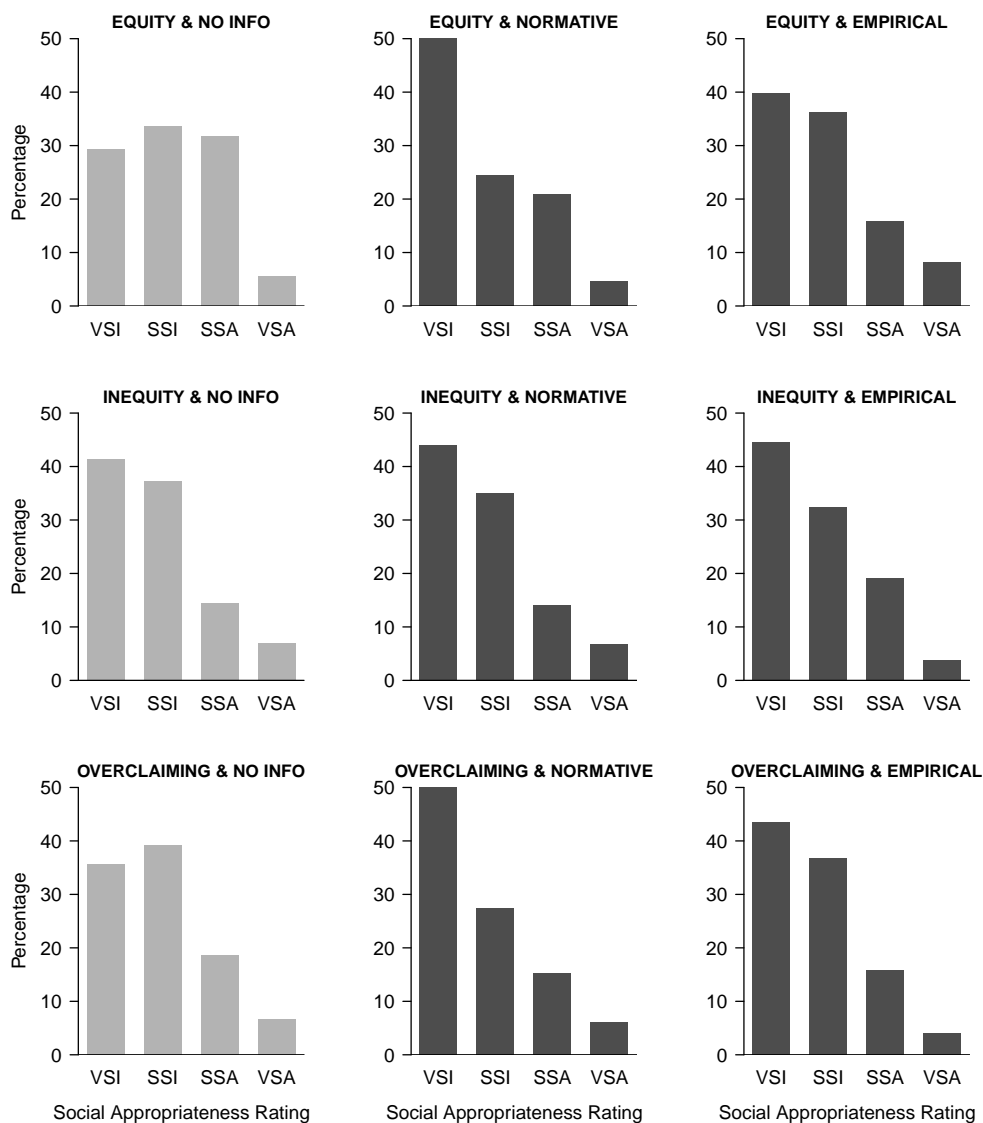
	Share of punishment assigned in %		
	(1)	(2)	(3)
INFO	4.146** (1.780)	2.915* (1.753)	3.583** (1.526)
Inequity	8.168*** (1.146)	8.168*** (0.867)	8.168*** (0.867)
Overclaim	6.247 (0.791)	6.344*** (0.783)	6.802*** (0.811)
INFO × Inequity	1.067 (1.062)	1.824 (1.283)	1.441 (1.094)
INFO × Overclaiming	-2.749* (1.498)	-2.299* (1.254)	-2.527** (1.073)
Lie	10.443 (0.396)	10.384*** (0.393)	10.110*** (0.324)
Constant	17.728*** (1.739)	20.552*** (3.694)	19.640*** (3.089)
INFO refers to	EMPIRICAL ONLY	NORMATIVE ONLY	EMPIRICAL & NORMATIVE
N	8,270	8,170	12,400
R^2	0.145	0.148	0.139

Notes: OLS results regressing the share of punishment assigned on the size of the lie, the equity nature of the lie (inequity, overclaiming), a dummy for whether norm information was provided (INFO) and interactions of the latter with the equity nature; the baseline category is therefore a subject in treatment NO INFO and an equity-based punishment scenario with lie size=1. INFO in column 1/2 designates treatment EMPIRICAL/NORMATIVE, respectively as opposed to NO INFO (no data from treatment NORMATIVE/EMPIRICAL used, respectively); in column 3 INFO designates treatments EMPIRICAL and NORMATIVE (all data used). Additional control variables include age, gender, education, and controls for how punishment scenarios were presented (increasing/decreasing) and the estimated implementation probability Robust standard errors in parentheses. */**/**: $p < 0.10/0.05/0.01$.

Additional Analyses for Study 2

Figure A.4 is a more fine-grained breakdown of Figure 5 in that the norm conditions are also broken down by normative and empirical information. Because both norm conditions yield consistent results, we collapsed them under 'Information' in Figure 5 in the main text.

Figure A.4: Social appropriateness of lying over different norm information and equity treatment



Notes: Norm information varies by column, the equity nature of the lie by rows. The social norm is measured via a 4-item Likert scale ranging over “Very Socially Inappropriate” (*VSI*), “Somewhat Socially Inappropriate” (*SSI*), “Somewhat Socially Appropriate (*SSA*)”, “Very Socially Appropriate (*VSA*)”.

Table A.5: Descriptive Statistics for subjects in Study 2

	Panel a) NO INFO					
	Equity		Inequity		Overclaiming	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
Age	37.862	13.791	39.529	13.686	39.843	12.835
Male	0.469	0.501	0.483	0.501	0.506	0.501
Educ.: less than high school degree	0.000	0.000	0	0.000	0.000	0.000
Educ.: high school graduate	0.094	0.292	0.080	0.273	0.133	0.340
Educ.: some college but no degree	0.150	0.358	0.161	0.369	0.193	0.396
Educ.: associate degree in college	0.094	0.292	0.092	0.290	0.090	0.288
Educ.: bachelor degree in college	0.481	0.501	0.477	0.501	0.398	0.491
Educ.: master degree	0.131	0.339	0.172	0.379	0.157	0.365
Educ.: doctoral degree	0.025	0.157	0.006	0.076	0.006	0.078
Educ.: professional degree (JD, MD)	0.025	0.157	0.011	0.107	0.024	0.154
observations	N=	160	N=	174	N=	166
	Panel b) INFO					
	Equity		Inequity		Overclaiming	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
Age	38.437	12.483	38.508	12.715	38.303	12.696
Male	0.469	0.500	0.456	0.499	0.435	0.497
Educ.: less than high school degree	0.012	0.108	0.000	0.000	0.003	0.054
Educ.: high school graduate	0.076	0.265	0.048	0.214	0.094	0.292
Educ.: some college but no degree	0.178	0.383	0.213	0.410	0.171	0.377
Educ.: associate degree in college	0.105	0.307	0.111	0.315	0.118	0.323
Educ.: bachelor degree in college	0.466	0.500	0.411	0.493	0.432	0.496
Educ.: master degree	0.128	0.335	0.180	0.385	0.144	0.352
Educ.: doctoral degree	0.012	0.108	0.012	0.109	0.026	0.161
Educ.: professional degree (JD, MD)	0.023	0.151	0.024	0.153	0.012	0.108
observations	N=	343	N=	333	N=	340

Notes: Demographic characteristics of subjects in Study 2 (mean and standard deviation), split by the different treatments (Equity norm: Equity, Inequity, and Overclaiming times NO INFO and INFO). We do not observe statistically significant differences between the six treatments for age (Kruskal-Wallis test: $p=0.502$), gender (χ^2 -test: $p=0.749$), or education (χ^2 -test: $p=0.197$). NO INFO pools the observations from treatments NORMATIVE and EMPIRICAL. If taken individually, the number of observations in NORMATIVE are $n=172$, $n=176$, and $n=163$ for the Equity-, Inequity-, and Overclaim-treatments, respectively; the corresponding number of observations in EMPIRICAL are $n=171$, $n=157$, and $n=177$.

Table A.6: Social appropriateness ratings – regression results

	Social Appropriateness score (1 – 4)		
	(1)	(2)	(3)
INFO	-0.154*** (0.050)		-0.261*** (0.088)
Inequity		-0.109* (0.058)	-0.268*** (0.100)
Overclaim		-0.109* (0.058)	-0.163 (0.101)
INFO × Inequity			0.232* (0.122)
INFO × Overclaiming			0.078 (0.123)
Constant	1.348*** (0.186)	1.215*** (0.176)	1.472*** (0.195)
<i>N</i>	1,516	1,516	1,516

F-test: coeff. for INFO + INFO × Inequity = Inequity; p=0.140

F-test: coeff. for INFO + INFO × Overclaiming = Overclaiming; p=0.908

Notes: OLS results regressing the social appropriateness score (coded as “Very Socially Inappropriate”=1, “Somewhat Socially Inappropriate”=2, “Somewhat Socially Appropriate”=3, “Very Socially Appropriate”=4) on indicators for the norm information treatments (INFO – pools EMPIRICAL and NORMATIVE) and the equity nature of the lie (inequity, overclaim); the baseline category is therefore a subject in treatment NO INFO and an equity-based punishment scenario with lie size=1. Additional control variables include age, gender, education. Robust standard errors in parentheses. */**/***: p<0.10/0.05/0.01.

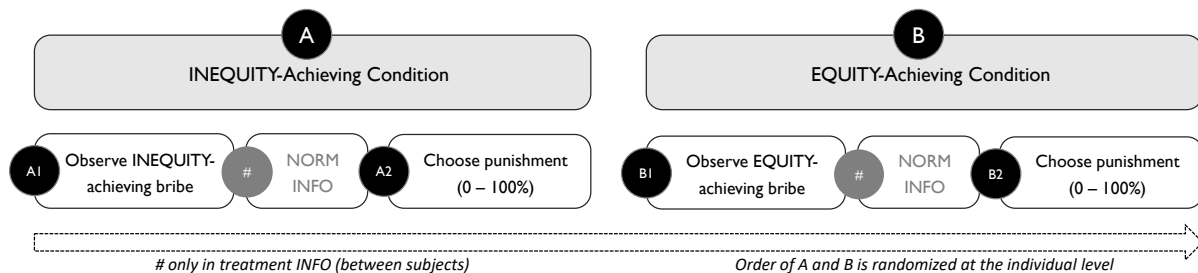
Appendix B: Details on the Vignette Experiment to corroborate the robustness of the results in the Behavioral Experiment of Study 1

Design

Our vignette featured a scenario where subjects imagined themselves in the role of an employee who observed a co-worker (Alex) in the same company taking a bribe amounting to \$10,000 in order to favor one of the company’s suppliers. The subjects could then blow the whistle and inflict punishment on Alex. Participants were able to choose from a range of potential whistleblowing actions (from not reporting at all, to reporting the case and also submitting tangible evidence). This then determined the probability of Alex being punished (from 0% to 100%, in 25 ppt. increments). Besides this feature, which reflects the staggered punishment in the Punisher Experiment (for the Behavioral Experiment in Study 1), two further crucial features of that behavioral experiment are also captured in this vignette experiment: i) whether Alex accepting the bribe creates EQUITY (bribe was accepted to make up for not having received a bonus payment outside of Alex’s control) or INEQUITY (bribe was accepted on top of a bonus payment), and ii) whether injunctive norm information (previous participants indicating that accepting a bribe is not the right thing to do) was present or not. We elicited the probability of punishing the co-worker twice for each subject in two *within-subject* treatments that were presented in random order:

- EQUITY-creating bribe treatment: Alex had been held accountable for a failed project tender. Even though Alex was not personally responsible for the failure, Alex was reprimanded by the company by being deprived of a \$10,000 end-of-the-year bonus. We also established a payoff difference in order to capture the equity-aspect of the behavioral experiment. For this, the subject playing the vignette was told that, she herself/he himself received the very same end-of-the-year bonus. By virtue of this setting, accepting the bribe – which also has a value of \$10,000 – creates equity for Alex when compared with the co-worker’s (i.e., the subject’s) payoff.

Figure B.1: Design of the Vignette Experiment in Study 1



- **INEQUITY-creating bribe treatment:** The setup was the same as in EQUITY, except that the failed tender did *not* result in Alex being reprimanded for it. Thus, both Alex and the subject received the \$10,000 end-of-the-year bonus. Consequently, compared to the subject, accepting the bribe therefore creates an (advantageous) inequity for Alex.

In addition to the above, each subject was in one of two *between-subject* conditions:

- **NORM INFO:** In this treatment, subjects were presented with an empirical norm nudge similar to the one used in the behavioral experiment.¹ It informed them that in a previous study, the majority of participants agreed with the statement that “*taking a bribe is highly unethical and not the right thing to do.*”²
- **NO INFO:** This treatment did not feature the information described above.

Based on the information above, this vignette experiment features a 2(within)×2(between) design, as illustrated in Figure B.1. Besides randomizing allocation to the INFO and NO INFO-between conditions, we also randomized whether subjects first saw the EQUITY and then the INEQUITY condition or vice versa. In addition, we also randomize the order in which the punishment actions were displayed (from 0% to 100% punishment probability in either increasing or decreasing order). The experiment concluded with a demographic questionnaire and participants were then entered into a lottery with the chance to win \$100 for participating.

Procedures

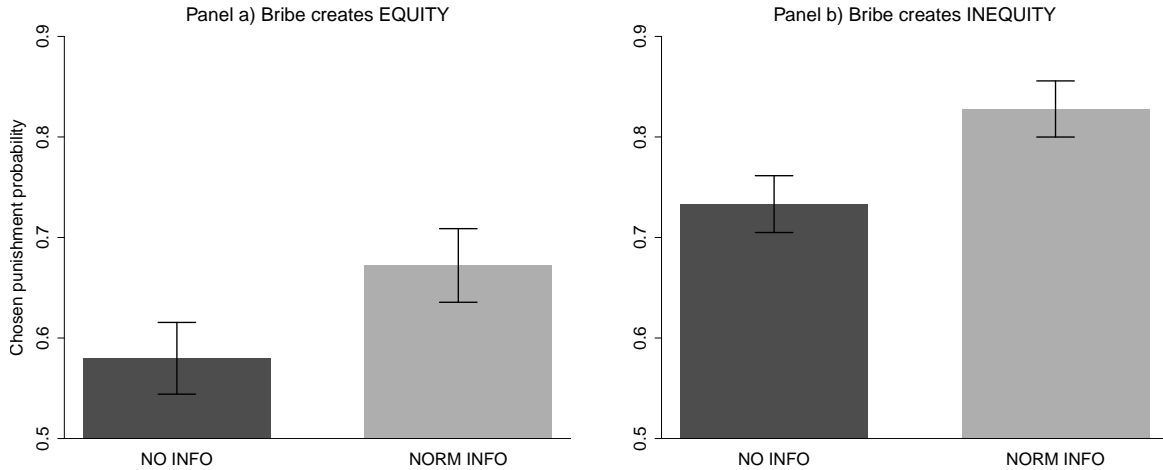
With the help of several research assistants, we collected data in person from a diverse population in May and June of 2021. Data was collected in various cities across 10 U.S. states, yielding a total of n=225 observations.³ Our study was pre-registered at (see Appendix C for details) and our analyses follow this pre-registration.

¹We opted for testing only one norm nudge, and the **NORMATIVE** condition in particular, for the following two reasons: first, we wanted the norm information in the vignette to be based on truthful information. For practical reasons, we were able to verify this normatively but not empirically, given that it is a fictitious setting. The truthfulness of the normative message was obtained based on a sample (n=60) at the University of Pennsylvania. Participants were first shown the vignettes (using the exact wording) and then asked to indicate whether they believe that accepting a bribe is the right thing to do. A majority said that accepting a bribe is not the right thing to do. Second, as our behavioral results above indicate, the normative nudge leads directionally to lower punishment than the empirical nudge. For this reason, our normative nudge in the vignette may be understood as the lower-bound among the two nudges.

²The truthfulness of this message was obtained based on a student sample (n=60) at the University of Pennsylvania. Participants were first shown the vignettes (using the exact wording) and then asked to indicate whether they believe that accepting a bribe is the right thing to do. A majority of participants indicated that accepting a bribe is *not* the right thing to do. Details are available upon request.

³Various research assistants approached participants in the vicinity of city centers at random and explained that they could participate in a scientific experiment in return for which they would have the chance to win one of three \$100 Amazon vouchers. Upon agreeing to participate, a QR code was presented that the participants were asked to scan on their phone. To ensure anonymity, the research assistants then stepped away and approached other potential participants. To be properly powered, we pre-registered to collect at least 200 observations.

Figure B.2: Punishment probability chosen for accepting a bribe



Notes: Punishment probability by whether NORM INFO or NO INFO is provided (between subject design). Panels a) and b) shows this for the situation when taking the bribe led to INEQUITY and EQUITY respectively (the two within-subject conditions). Error bars denote SEM.

Results

Because our goal is to examine the robustness of the Punisher Experiment (for the Behavioral Experiment in Study 1), both our hypotheses and analyses follow from it. To achieve comparability across settings, we are mainly interested in understanding two behavioral aspects: first, is enforcement sensitive to the provision of norm information? Second, do punishers enforce norms differently depending on whether Alex’s acceptance of the bribe achieves equity versus inequity?

We first analyze the results using non-parametric tests. They mirror those from the behavioral experiment and are presented in Figure B.2. For one, we observe that providing norm information significantly increases punishment for inequity-creating corruption (0.733 vs. 0.828, Wilcoxon rank-sum test: $p=0.005$) and, to a lesser extent, equity-creating corruption (0.672 vs. 0.580, $p=0.074$).⁴ We also observe an increase when comparing the effect of norm information when examining the punishment formed as averages across the two different inequity scenarios (0.657 vs. 0.750, $p=0.003$). Consistent with the Punisher Experiment (for the Behavioral Experiment in Study 1), motives matter for the norm enforcement decisions across norm information treatments: the punishment for accepting an equity-creating bribe is always lower than that of an inequity-creating bribe (over both norm-INFO conditions 0.623 vs. 0.778, Wilcoxon signed-rank test: $p<0.001$). This also holds when we look at the (within-) effect of the bribe’s equity-effect separately by whether norm information was provided to the subject (0.672 vs. 0.828, $p<0.001$) or not (0.580 vs. 0.733, $p=0.002$).

⁴We also find a positive, significant effect of providing norm information when we only look at the first responses by subjects, before they experienced their respective within variation (0.645 vs. 0.762, $p=0.010$).

We also estimate the following regression model:

$$p_{is} = \alpha + \beta_1 \text{INFO}_i + \beta_2 \text{EQUITY-creating}_s + \boldsymbol{\gamma}' \mathbf{Controls}_i + \epsilon_{is}. \quad (2)$$

In the above, the dependent variable p_{is} is the chosen punishment probability of subject i in (equity-) scenario s (in %). INFO_i indicates whether the subject was in a treatment where norm information was provided and EQUITY-creating_s denotes whether accepting the bribe in scenario s created equity (as opposed to inequity). Finally, the vector $\mathbf{Controls}_i$ captures subjects' age and gender. It also contains a dummy to control for the order in which the two equity-scenarios were presented for each subject and whether the punishment options were presented in an increasing or decreasing order.

The results from estimating the above model by OLS (clustering standard errors on the subject level) are presented in Table B.1. Column 1 shows the significant effects of providing norm-INFO and being in the EQUITY-creating scenario (8.1 ppt.-increase and 15.4 ppt.-decreases in punishment, respectively). Column 2 adds the interaction term of the two treatment dummies to the above model. The unchanged estimates for the non-interacted coefficients and the effectively zero coefficient for the interaction term show that the effect of providing the norm-INFO separately does not differ substantially across the equity-conditions. Overall, these findings corroborate our findings from the behavioral experiment in the context of whistle-blowing. Across different specifications, we find that the punishment for accepting a bribe is stronger when NORM-Info is provided and weaker when accepting the bribe creates equity.

Table B.1: Punishment in the vignette experiment – regression results

	Punishment probability in %	
	(1)	(2)
INFO	8.120** (4.027)	8.235** (3.927)
EQUITY	-15.444*** (1.922)	-15.336*** (2.725)
INFO×EQUITY		-0.230 (6.263)
Constant	64.159*** (7.458)	64.105*** (7.439)
Controls	yes	yes
<i>N</i>	450	450
<i>R</i> ²	0.137	0.137

Notes: OLS results regressing the chosen punishment probability on a dummy whether a subject is in the norm-INFO treatment and whether the punishment refers to the situation with a EQUITY-creating bribe plus the interaction of the two; the baseline category is therefore a subject in treatment NO INFO and an inequity-creating bribe. Additional control variables include age, gender and controls for how punishment scenarios were presented (increasing/decreasing) and whether the EQUITY-creating was shown first. Two punishment scenarios per punisher; standard errors are clustered at the punisher level. */**/**: $p < 0.10/0.05/0.01$.

Appendix C: Information regarding pre-registration

Both experiments were pre-registered on <http://aspredicted.org>. Below are the links to the anonymized pre-registration files that are uploaded to <https://osf.io/nfkyc>.

Links:

- Study 1 – Behavioral Experiment: <https://osf.io/a3nzt>
- Study 1 – Vignette Experiment: <https://osf.io/4h7xk>
- Study 2: <https://osf.io/jnc65>

Notes regarding data collection for the behavioral experiment of Study 1:

- Data for the behavioral experiment was collected in April, May, and November 2019.
- A first batch of 106 observations was collected in April 2016 to test the functionality of the experimental interface with subjects from the representative pool provided by Dynata. Based on these observations, we pre-registered the study and started to collect additional observations in May.
- Due to miscommunication with the data collection company, we obtained almost twice as many observations (987 instead of the targeted 500) for the punishers by May.
- Upon inspecting the data, we found that the quota-based sample was representative of the US population across our defined age and gender-bins but not for the cross product of those bins (e.g., 50-59 years old females). We then approached the survey firm to rectify this issue. By a courtesy agreement, the firm sampled additional observations in November in order to have a representative sample also along these cross bins. This yielded our final sample with a total of 1,240 observations.
- Given this windfall in statistical power, the resulting improvement in the precision of our estimates (see, e.g., [Maxwell et al., 2008](#)), and to avoid a waste of resources, we decided to utilize the full sample.
- As a robustness check, we also repeated our analysis while excluding the 106 observations collected in April before we pre-registered the behavioral experiment. All results are similar in term of significance and magnitude to those stated in the main text.
- For all punisher observations, we collected observations for the potentially punished active players according to our stated 15:1-ratio-rule (see Footnote 14).

Appendix D: Instructions

In the following, we display screenshots for all of our experiments. Specifically, we present them in the following order:

1. Instructions for the Liar (Sub-)Experiment in Study 1
2. Instructions for the Punisher (Sub-)Experiment in Study 1
3. Instructions for the Vignette Experiment in Study 1
4. Instructions for the experiment in Study 2

Notes:

- In order to facilitate a comprehensive display, we omit repeating elements such as "continue"-buttons.
- If a screen does not fit a single page in this appendix, it is split over two pages and the screen number gets an "a" or "b"-suffix for the first and second part of the screen (e.g., "Screen 7a).

D.1: Instructions for the Liar (Sub-)Experiment in Study 1

The following pages display the screens for the Liar Experiment in Study 1 (see Figure 1). Note that subjects had different roles in the experiment: "liars" and "victims". We present the screens in the order in which they were presented to subjects. Some screens were the same, independent of a subject's role. If not otherwise noted in the respective captions, screens are shown to subjects in both roles.

Figure C.1.1: Liar Experiment (for the Behavioral Experiment in Study 1), Screen #1a

WELCOME!

In the following, you will be asked to answer questions and make some decisions.

Note: Sometimes, it will take a moment until you can click the "continue"-button at the bottom.

Confidentiality

The data collected in this study do not include any personally identifiable information about you. Your data will be kept separate from your MTurk ID. By participating, you understand and agree that the research data gathered during this study will be used by ETH Zurich and aggregated results will be published.

Contact Information

This scientific study is conducted by the Swiss Federal Institute of Technology (ETH) in Zurich, Switzerland. If you have any questions concerning this study, please write to descil@ethz.ch.

We recommend participating in this study only from a computer, not a touchscreen device.

Please click each of the boxes below to indicate that you have understood the rules and accept them.

You will not be able to return to previous pages.

You have to finish the entire study to get your additional payment.

The last page of this study will provide you with a verification code. Please provide the verification code on the MTurk-page where you accepted the assignment. Without this code, we cannot approve your submission and cannot pay you for it.

Please read the instructions carefully. There will be simple control questions which will check your understanding. If you do not answer these questions correctly you will not be able to finish the entire study.

I am a US resident and 18 years or older.

I would like to participate in this study and agree to the above rules.

In this study you can earn earnings in addition to the announced wage (payed as a bonus). However, the transfer of this additional payment will take up to a week.

Figure C.1.2: Liar Experiment (for the Behavioral Experiment in Study 1), Screen #1b

RULES IN THIS STUDY

You have received \$0.50 for participating in this study and you can earn an additional bonus.

In this study, you will perform a task and make decisions, which can influence your bonus payment. Any payment will only occur upon completion of the study.

Please read the rules carefully to maximize your earnings:

Your decisions

- The task is to count numbers. You are requested to do so for about seven minutes. You will receive a bonus of \$1.50 for its completion.
- Upon completion, you will be paired with another participant who also finished the task.
- One of you will be chosen at random to make an additional decision. The details about this decision will be given to you in time.
- Both of you will enter a lottery with the chance to win an additional bonus of \$2.50 (in addition to the \$1.50 you can earn for completing the task).
- Everyone who decides to participate has equal chances to win this bonus.
- The experiment ends with some demographic questions.

Your payment

- \$0.50 for participation.
- A bonus payment of up to \$4.00 (up to \$1.50 for completion and \$2.50 as an additional bonus), depending on the outcome of the study.
- The final bonus is subject to decisions made by you and others in throughout the experiment as well as the outcome of the lottery.

If you do not agree with the above, please close this window.

By clicking the button below I agree that I have understood the above rules.

Figure C.1.3: Liar Experiment (for the Behavioral Experiment in Study 1), Screen #1b

RULES IN THIS STUDY

You have received \$0.50 for participating in this study and you can earn an additional bonus.

In this study, you will perform a task and make decisions, which can influence your bonus payment. Any payment will only occur upon completion of the study.

Please read the rules carefully to maximize your earnings:

Your decisions

- The task is to count numbers. You are requested to do so for about seven minutes. You will receive a bonus of \$1.50 for its completion.
- Upon completion, you will be paired with another participant who also finished the task.
- One of you will be chosen at random to make an additional decision. The details about this decision will be given to you in time.
- Both of you will enter a lottery with the chance to win an additional bonus of \$2.50 (in addition to the \$1.50 you can earn for completing the task).
- Everyone who decides to participate has equal chances to win this bonus.
- The experiment ends with some demographic questions.

Your payment

- \$0.50 for participation.
- A bonus payment of up to \$4.00 (up to \$1.50 for completion and \$2.50 as an additional bonus), depending on the outcome of the study.
- The final bonus is subject to decisions made by you and others in throughout the experiment as well as the outcome of the lottery.

If you do not agree with the above, please close this window.

By clicking the button below I agree that I have understood the above rules.

Figure C.1.4: Liar Experiment (for the Behavioral Experiment in Study 1), Screen #3

Counting Task - Explanation

In the following, you will have to repeatedly perform a counting task. For this, you have to count how often either a "1" appears in a sequence of numbers.

Note: If the number you entered is wrong by more than one from the actual number, you have to re-count and enter it again. Before you can submit an entry, you always have to wait 45 seconds.

In the example, you have to count how often a "1" occurs.

(In this example, we have colored this number. This will not be the case for the actual tasks.)

For this example, we tell you the number: It is 46 times that a "1" occurs.

0	1	0	0	0	1	0	0	0	0	1	1
0	0	1	0	0	1	0	1	0	1	0	1
0	1	1	1	1	0	1	0	0	0	1	0
0	0	1	0	0	1	0	1	1	0	0	1
1	1	1	0	0	0	0	0	1	1	0	1
1	1	0	1	0	0	0	1	0	1	0	0
0	1	0	0	0	0	1	1	1	1	1	0
0	0	0	1	1	0	1	1	0	0	0	0
0	1	1	0	1	0	0	1	0	1	0	0

Please enter how often "1" occurs in the above.

Figure C.1.5: Liar Experiment (for the Behavioral Experiment in Study 1), Screen #4

Counting Task - Explanation

You can now start with the actual task clicking the button below. You will have to do this task **5 times**.

Figure C.1.6: Liar Experiment (for the Behavioral Experiment in Study 1), Screen #5 (five such screens, "Counting Task 1" – "Counting Task 5")

Counting Task 1

1	1	0	0	0	0	0	1	0	1	0	1
0	0	0	1	1	0	1	1	1	0	1	0
0	1	0	0	0	1	1	1	1	0	1	1
1	1	0	0	0	1	1	1	0	1	0	1
1	0	0	0	0	1	1	0	1	0	0	1
0	0	1	0	0	1	1	1	0	0	1	0
1	0	1	1	0	1	0	0	1	1	1	1
1	0	1	0	1	0	0	0	1	1	0	0
1	0	1	1	0	1	0	1	0	1	0	0

Please enter how often "1" occurs in the above.

Figure C.1.7: Liar Experiment (for the Behavioral Experiment in Study 1), Screen #6a ("liars" only)

Earning the additional bonus - Explanation

You will now be paired with another participant who also finished the task. Both of you are given the chance to win an additional bonus. However, only one of you will actually win it. Who will win it is determined by chance.

How it works:

- You have been randomly chosen to toss a fair, five-sided die (with numbers "1", "2", "3", "4", "5"). You will do so on the next page.
- You are eligible for as many lottery tickets as the die shows.
- We will ask you to report what the die shows.
- You will receive the number of lottery tickets you report.
- The other participant with whom you are paired has been randomly chosen to not toss a die. Instead, this participant gets a fixed amount of 3 lottery tickets.
- Your and the other participant's lottery tickets will be put into a (virtual) raffle. One ticket will be randomly drawn. The person to whom this ticket belongs earns the bonus of \$2.50.

Figure C.1.8: Liar Experiment (for the Behavioral Experiment in Study 1), Screen #6b ("liars" only)

Rules of the game:

- Report the number that the die shows.
- You will receive the number of tickets you report.
- If you report a higher number you receive more tickets.
- If you report a lower number you receive less tickets.

Below, we provide you with a few examples.

Please familiarize yourself with them and the above-described procedure and check all the examples to confirm that you have read and understood them. There will be some questions which will check your understanding.

Example: the other participant already has **3** tickets. You roll a **3** and report a **1**. You receive **1** ticket. The chance for you to win the bonus is $1/4 = 25\%$.

Example: the other participant already has **3** tickets. You roll a **3** and report a **6**. You receive **6** tickets. The chance for you to win the bonus is $6/9 = \text{about } 67\%$.

Example: the other participant already has **3** tickets. You roll a **3** and report a **3**. You receive **3** tickets as well. The chance for you to win the bonus is $3/6 = 50\%$.

Figure C.1.9: Liar Experiment (for the Behavioral Experiment in Study 1), Screen #7 ("liars" only)

Earning the additional bonus - Check Question

Consider the following situation: *Your die toss shows a "2". You report "4".*

How many tickets will you get? (enter a number)

How many tickets will there be in total, that is, for you and the other player together? (enter a number)

Click to write the question text

(please click here if you want to see the explanations and examples shown to you before)

Figure C.1.10: Liar Experiment (for the Behavioral Experiment in Study 1), Screen #8 ("liars" only)

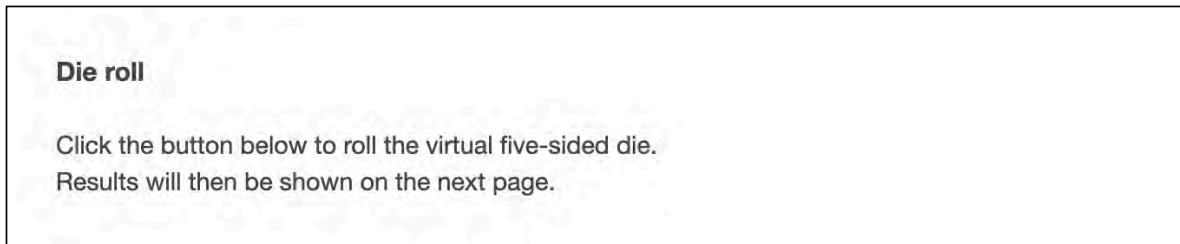


Figure C.1.11: Liar Experiment (for the Behavioral Experiment in Study 1), Screen #9 ("liars" only)

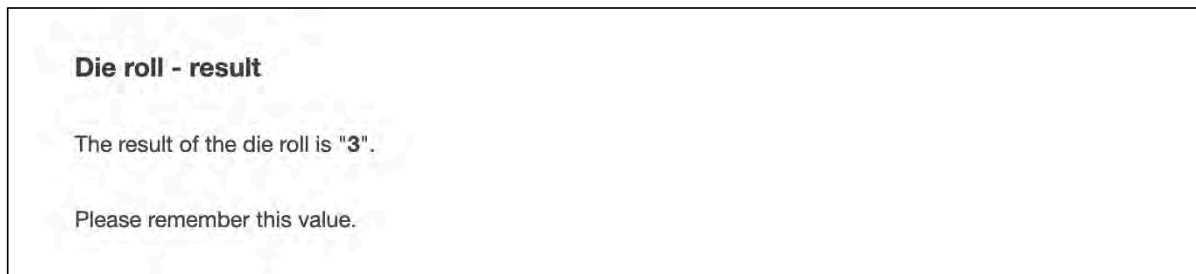


Figure C.1.12: Liar Experiment (for the Behavioral Experiment in Study 1), Screen #10a ("liars" only)

Your chance to win the additional bonus

We now ask you to report the number shown on the die. The number you report will determine how many lottery tickets you get.

For your reference, you find the explanation of how exactly this works repeated below (the same information that was displayed to you before).

Earning the additional bonus - Explanation

You will now be paired with another participant who also finished the task. Both of you are now given the chance to win an additional bonus. However, only one of you will actually win it. Who wins it is determined by chance.

How it works:

- You have been randomly chosen to toss a fair, five-sided die (with numbers "1", "2", "3", "4", "5"). You will do so on the next page.
- You are eligible for as many lottery tickets as the die shows.
- We will ask you to report what the die shows.
- You will receive the number of lottery tickets you report.
- The other participant with whom you are paired has been randomly chosen to not toss a die. Instead, this participant gets a fixed amount of 3 lottery tickets.
- Your and the other participant's lottery tickets will be put into a (virtual) raffle. One ticket will be randomly drawn. The person to whom this ticket belongs earns the bonus of \$2.50.

Figure C.1.13: Liar Experiment (for the Behavioral Experiment in Study 1), Screen #10b ("liars" only)

Rules of the game:

- Report the number that the die shows.
- You will receive the number of tickets you report.
- If you report a higher number you receive more tickets.
- If you report a lower number you receive less tickets.

Below, we provide you with a few examples.
Please familiarize yourself with them and the above-described procedure and check all the boxes to confirm that you have read and understood them.
You can then continue.

Example: the other participant already has **3** tickets. You roll a **3** and report a **1**. You receive **1** ticket. The chance for you to win the bonus is $1/4 = 25\%$.

Example: the other participant already has **3** tickets. You roll a **3** and report a **3**. You receive **3** tickets as well. The chance for you to win the bonus is $3/6 = 50\%$.

Example: the other participant already has **3** tickets. You roll a **3** and report a **6**. You receive **6** tickets. The chance for you to win the bonus is $6/9 = \text{about } 67\%$.

Please enter the number you want to report (a number between 1 and 5).
You will receive the corresponding number of tickets.

Figure C.1.14: Liar Experiment (for the Behavioral Experiment in Study 1), Screen #11 ("liars" only)

Possibility to revise your report

On the previous page, you reported the following number:

This is what your die actually showed:

Punishment:

- The information above will be used to let other participants in another study decide whether they want to punish you.
- These other participants do not play the game you have just played but they know its rules.
- They also know that the rule was to report the number which the die showed.
- They can punish you only if you reported a higher number than you actually rolled.
- Punishment will be in the form of "punishment points". Each punishment point is worth minus \$0.30 and you can receive up to five punishment points.
- The sum of your punishment points will be subtracted from the \$1.50 you receive as an individual bonus for having successfully completed the counting task (not from the \$2.50-bonus which either you or the other participant you have been paired with can receive).

Example: You receive two punishment points. You will earn an individual bonus of $1.50 - 2 \times 0.30 = \$0.90$.

Example: You receive five punishment points. You will earn an individual bonus of $1.50 - 5 \times 0.30 = \$0.00$.

Revision:

You can now revise the number you reported. If you do so, this revised number - not your initially reported number - will be used to determine the number of lottery tickets for winning the additional \$2.50-bonus (the other player with whom you are paired still receives three tickets).

Also, your punishment will be based only on the revised number (the other participants who can punish you will never learn your originally reported number).

If you want to revise your report, please do so below by entering a new number.

If you do not want to revise your report, you re-enter your initial number.

Figure C.1.15: Liar Experiment (for the Behavioral Experiment in Study 1), Screen #12 ("victims" only)

Earning the additional bonus - Explanation

You will now be paired with another participant who also finished the task. Both of you are given the chance to win an additional bonus. However, only one of you will actually win it. Who will win it is determined by chance.

How it works:

- One participant with in your pair, called Participant A, was randomly chosen to toss a fair, five-sided, virtual die (with numbers "1", "2", "3", "4", "5").
- Participant A is eligible for as many lottery tickets as the die showed.
- Participant A is asked to report what the die shows and received the corresponding number of lottery tickets.
- The other participant, called participant B, gets a fixed amount of 3 lottery tickets.
- Participant A and B's lottery tickets are put into a (virtual) raffle from which one ticket was randomly drawn.
- The participant to whom this ticket belongs earned the bonus of \$2.50.

Rules of the game for Participant A:

- Report the number that the die shows.
- You will receive the number of tickets you report.
- If you report a higher number you receive more tickets.
- If you report a lower number you receive less tickets.

Below, we provide you with a few examples.
Please familiarize yourself with them and the above-described procedure and check all the boxes to confirm that you have read and understood them. There will be some questions which will check your understanding.

Example: Participant B has **3** tickets. Participant A rolls a **3** and reports a **6**. Participant A receives **6** tickets. The chance for Participant A to win the bonus is $6/9 = \text{about } 67\%$.

Example: Participant B has **3** tickets. Participants A rolls a **3** and reports a **3**. Participant A receives **3** tickets. The chance for Participant A to win the bonus is $3/6 = 50\%$.

Example: Participant B has **3** tickets. Participant A rolls a **3** and reports a **1**. Participant A receives **1** ticket. The chance for Participant A to win the bonus is $1/4 = 25\%$.

Figure C.1.16: Liar Experiment (for the Behavioral Experiment in Study 1), Screen #13a ("victims" only)

Earning the additional bonus - Check Question

Consider the following situation: *Participant A's die toss showed a "2". Participant A reported "4".*

How many tickets will Participant 1 get? (enter a number)

How many tickets will there be in total, that is, for Participant 1 and 2 together? (enter a number)

For your reference, here are the explanations and examples provided to you on the previous page:

Earning the additional bonus - Explanation

You will now be paired with another participant who also finished the task. Both of you are given the chance to win an additional bonus. However, only one of you will actually win it. Who will win it is determined by chance.

Figure C.1.17: Liar Experiment (for the Behavioral Experiment in Study 1), Screen #13b ("victims" only)

How it works:

- One participant with in your pair, called Participant A, was randomly chosen to toss a fair, five-sided, virtual die (with numbers "1", "2", "3", "4", "5").
- Participant A is eligible for as many lottery tickets as the die showed.
- Participant A is asked to report what the die shows and received the corresponding number of lottery tickets.
- The other participant, called participant B, gets a fixed amount of 3 lottery tickets.
- Participant A and B's lottery tickets are put into a (virtual) raffle from which one ticket was randomly drawn.
- The participant to whom this ticket belongs earned the bonus of \$2.50.

Rules of the game for Participant A:

- Report the number that the die shows.
- You will receive the number of tickets you report.
- If you report a higher number you receive more tickets.
- If you report a lower number you receive less tickets.

Example: Participant 2 has **3** tickets. Participant 1 rolls a **3** and reports a **3**. Participant 1 receives **3** tickets. The chance for Participant 1 to win the bonus is $3/6 = 50\%$.

Example: Participant B has **3** tickets. Participant A rolls a **3** and reports a **6**. Participant A receives **6** tickets. The chance for Participant A to win the bonus is $6/9 = \text{about } 67\%$.

Example: Participant 2 has **3** tickets. Participant 1 rolls a **3** and reports a **1**. Participant 1 receives **1** ticket. The chance for Participant 1 to win the bonus is $1/4 = 25\%$.

Figure C.1.18: Liar Experiment (for the Behavioral Experiment in Study 1), Screen #14 ("victims" only)

Your opinion

Before we tell you whether you are Player A (and can roll the die) or Player B, we ask what is the right thing to do.

Note: Player B always gets 3 tickets, Player A the reported number (1-5).

Based on the scenario described above, what do you think is the right thing for Player A to do?

I think the right thing to do for PlayerA if he/she

	<i>...is to report a "1".</i>	<i>...is to report a "2".</i>	<i>...is to report a "3".</i>	<i>...is to report a "4".</i>	<i>...is to report a "5".</i>
<i>... rolled a "4"</i>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<i>... rolled a "3"</i>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<i>... rolled a "1"</i>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<i>... rolled a "5"</i>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<i>... rolled a "2"</i>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Figure C.1.19: Liar Experiment (for the Behavioral Experiment in Study 1), Screen #14 ("victims" only)

Your opinion

Before we tell you whether you are Player A (and can roll the die) or Player B, we ask what is the right thing to do.

Note: Player B always gets 3 tickets, Player A the reported number (1-5).

Based on the scenario described above, what do you think is the right thing for Player A to do?

I think the right thing to do for PlayerA if he/she

	<i>...is to report a "1".</i>	<i>...is to report a "2".</i>	<i>...is to report a "3".</i>	<i>...is to report a "4".</i>	<i>...is to report a "5".</i>
<i>... rolled a "4"</i>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<i>... rolled a "3"</i>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<i>... rolled a "1"</i>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<i>... rolled a "5"</i>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<i>... rolled a "2"</i>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Figure C.1.20: Liar Experiment (for the Behavioral Experiment in Study 1), Screen #15 ("victims" only)

Your role

You have been randomly chosen to be
Player B.

Figure C.1.21: Liar Experiment (for the Behavioral Experiment in Study 1), Screen #16 ("victims" only)

Earning the additional bonus - Explanation

You will now be paired with another participant who also finished the task. Both of you are given the chance to win an additional bonus. However, only one of you will actually win it. Who will win it is determined by chance.

How it works:

- The other participant has been randomly chosen to toss a fair, five-sided die (with numbers "1", "2", "3", "4", "5").
- The other participant is eligible for as many lottery tickets as the die shows.
- The other participant will be asked to report what the die shows.
- The other participant will receive the number of lottery tickets he/she reports.
- You have been randomly chosen to not toss a die. Instead, you get a fixed amount of 3 lottery tickets.
- Your and the other participant's lottery tickets will be put into a (virtual) raffle. One ticket will be randomly drawn. The person to whom this ticket belongs earns the team bonus of \$2.50.

Rules of the game:

- The other participant is supposed to report the number that the die shows.
- The other participant receives the number of tickets he/she reports.
- If the other participant reports a higher number he/she receives more tickets.
- If the other participant reports a lower number he/she receives less tickets.

The draw will take place when the study is over. You will then receive your bonus payment after a few days.

Figure C.1.22: Liar Experiment (for the Behavioral Experiment in Study 1), Screen #17

Some final questions

The study is now over. Before its end, we ask you to provide us with some information about yourself.

What is your year of birth?

Figure C.1.23: Liar Experiment (for the Behavioral Experiment in Study 1), Screen #18

What is the gender you identify yourself with the most?

Female

Male

D.2: Instructions for the Punisher (Sub-)Experiment in Study 1

The following pages display the screens for the Punisher Experiment in Study 1 (see Figure 1). All subjects were in the role of punishers but were assigned to different treatments, displaying different norm information (NO INFO, NORMATIVE, and EMPIRICAL). This norm information is displayed on Screen #8, here displaced in Figure C.2.9. This screenshot shows the norm information displayed in treatment NORMATIVE. For EMPIRICAL, the text was adjusted to the one reproduced in the main text; for NO INFO, no information was shown.

Figure C.2.1: Punisher Experiment (for the Behavioral Experiment in Study 1), Screen #1

WELCOME!

In the following, you will be asked to answer questions and make some decisions.

Note: Sometimes, it will take a moment until you can click the button on the bottom to continue.

Confidentiality

The data collected in this study do not include any personally identifiable information about you. By participating, you understand and agree that the research data gathered during this study will be used by ETH Zurich and aggregated results will be published.

Contact Information

This scientific study is conducted by the Swiss Federal Institute of Technology (ETH) in Zurich, Switzerland. If you have any questions concerning this study, please write to descil@ethz.ch.

We recommend to participate in this study only from a computer, not a touchscreen device.

"Continue"-Buttons will appear after some time.

Please click each of the boxes below to indicate that you understand and accept the rules.

You will not be able to return to previous pages.

You have to finish the entire study.

Please read the instructions carefully. There will be simple control questions which will check understanding. If you do not answer these questions correctly you will not be able to finish the entire study.

I am a US resident and 18 years or older.

I would like to participate in this study and agree to the above rules.

Figure C.2.2: Punisher Experiment (for the Behavioral Experiment in Study 1), Screen #2

Purpose of this study

In the current study (henceforth called "**Study C**"), you will be asked to judge how participants in another, the previous study have behaved. In this previous study (henceforth called "**Study P**"), participants could earn real money, potentially through cheating. Your judgment of the observed behavior will consist of deciding whether and how much to punish such behavior. Your inflicted punishment will have real monetary consequences for participants in Study P.

We will explain to you the situation in two parts:

Explanation Part a) - Situation of other participants in Study P: an explanation of what participants in the other study had to do, how they could cheat and earn money.

Explanation Part b) - Punishment by you: an explanation of how you can punish the other participants.

Note: please read the following information carefully. Participant in Study P have already participated but will not receive their payment until the current Study C is over. Your punishment will have real consequences on how much money participants in Study P can earn.

Figure C.2.3: Punisher Experiment (for the Behavioral Experiment in Study 1), Screen #3

Explanation A - Situation for other participants

In **Study P**, every participant was able to earn \$2 plus an additional bonus of \$2.5 for completing a counting task. The task was to count either how many zeros or ones appeared in a picture similar to the one shown below.

1	1	0	0	0	0	0	1	0	1	0	1
0	0	0	1	1	0	1	1	1	0	1	0
0	1	0	0	0	1	1	1	1	0	1	1
1	1	0	0	0	1	1	1	0	1	0	1
1	0	0	0	0	1	1	0	1	0	0	1
0	0	1	0	0	1	1	1	0	0	1	0
1	0	1	1	0	1	0	0	1	1	1	1
1	0	1	0	1	0	0	0	1	1	0	0
1	0	1	1	0	1	0	1	0	1	0	0

In total, participants in **Study P** had to do complete five such counting tasks.

(After a while, the button to proceed will appear below. If you have read the above, you can click it.)

Figure C.2.4: Punisher Experiment (for the Behavioral Experiment in Study 1), Screen #4

Explanation Part a) - Situation for other participants

After receiving \$2 for completing the counting task, participants in **Study P** were paired with one other participant (who also finished the task). They were then given the opportunity to earn an **additional bonus of \$2.5**. The rules were as follows:

How it works:

- One participant within a pair, called Participant A, was randomly chosen to toss a fair, five-sided, virtual die (with numbers "1", "2", "3", "4", "5").
- Participant A was then eligible for as many lottery tickets as the die showed.
- Participant A was asked to report what the die shows and received the corresponding number of lottery tickets.
- The other participant of the pair, called Participant B, always received a fixed amount of 3 lottery tickets without rolling the die.
- Participant A and B's lottery tickets were put into a (virtual) raffle from which one ticket was randomly drawn.
- The participant to whom this ticket belongs earned the bonus of \$2.50.

The rules of the game given to Participants A was that they should report the number shown on the die. However, the participants could cheat and report any number between 1 and 5. Thus, reporting a higher or lower number yielded more or fewer tickets, respectively.

Below are three examples:

(please read and then click each of them in order to proceed)

Example: A Participant B has **3** tickets. Participants A rolls a **3** and reports a **3**. Participant A receives **3** tickets. The chance for Participant A to win the bonus is $3/6 = 50\%$.

Example: A Participant B has **3** tickets. Participant A rolls a **3** and reports a **1**. Participant A receives **1** ticket. The chance for Participant A to win the bonus is $1/4 = 25\%$.

Example: A Participant B has **3** tickets. Participant A rolls a **3** and reports a **5**. Participant A receives **5** tickets. The chance for Participant A to win the bonus is $5/8 = \text{about } 63\%$.

Figure C.2.5: Punisher Experiment (for the Behavioral Experiment in Study 1), Screen #5

Earning the additional bonus - Check Question

Consider the following situation: *Participant A's die toss showed a "2". Participant A reported "4".*

The paired Participant B gets 3 tickets. How many tickets will Participant A get? (enter a number)

You can only proceed if your answer is correct. Please make sure that you understood the instructions before answering to avoid delays.

[\(Click here if you want to see again the explanation and examples how punishment works which were shown to you on the previous page\)](#)

Figure C.2.6: Punisher Experiment (for the Behavioral Experiment in Study 1), Screen #6

Explanation Part b) - Your punishment decision

You will now have the opportunity to punish the cheating behavior of Participant As, which will inflict real monetary consequences.

How your punishment works:

- You will be presented with different scenarios of how Participant As could cheat (e.g., "die shows 2, Participant A reports 4").
- You can assign up to five punishment points for each scenario.
- Each punishment point is worth (minus) \$0.30, which will be subtracted from Participant A's earnings. Thus, the total value of punishment point can be between \$0 ($=0 \times \0.3) and \$1.50 ($=5 \times 0.3$).
- Every Participant A will be assigned to one participant of this study.
(If there are more Participant As than participants in the current study, some randomly determined participants of the current study will not be paired with a Participant A.)
- The participant who is paired will lose the amount which corresponds to the punishment determined by you.

Below, are two examples of how punishment works:

(please read and then click each of them in order to proceed)

Example: Suppose you determined to punish a given type of Participant A's behavior with **5 points**. As a consequence, Participant A who is assigned to you then earns $2.00 - 5 \times 0.3 = \$0.5$ if he/she exhibited such behavior.

Example: Suppose you determined to punish a given type of Participant A's behavior with **2 points**. As a consequence, Participant A who is assigned to you then earns $2.00 - 2 \times 0.3 = \$1.4$ if he/she exhibited such behavior.

Figure C.2.7: Punisher Experiment (for the Behavioral Experiment in Study 1), Screen #7

Your punishment decision - Comprehension Check Question

Consider the following hypothetical situation:

- You have been paired with a Participant A in the other study for whom the result of the die toss was "2".
- That Participant A reported "4".

Suppose you decided to assign 3 punishment points for such behavior.

What is the total punishment which Participant A receives?

(enter the \$-amount which will be subtracted from Participant A's earnings)

You can only proceed if your answer is correct. Please make sure that you understood the instructions before answering to avoid delays.

(Click here if you want to see again the explanation and examples how punishment works which were shown to you on the previous page)

Figure C.2.8: Punisher Experiment (for the Behavioral Experiment in Study 1), Screen #8a

Your punishment decision

We now ask you to make punishment decisions for 10 different scenarios in which cheating could occur.

To repeat:

- If matched with you, you decide the punishment for Participant A's cheating behavior.
- Each punishment point will lead to a loss of \$0.3 for the punished participant in the role of Participant A.

please click if you have read the above

Please note the following:

In a previous study, we asked people what they consider to be the right thing to do for participants in the role of Player A. Those people could not play the game themselves.

The majority of these people stated that the right thing to do for Player A is to report the number truthfully (i.e., report exactly what the die showed).

please click if you have read the above

Please click below how many punishment points (0-5) you want to assign for cheating behavior in each of 10 possible scenarios (you have to enter a value for each scenario).

I want to assign the following number of punishment points if Participant A...

Figure C.2.9: Punisher Experiment (for the Behavioral Experiment in Study 1), Screen #8b

<i>...had a die-outcome of "1"</i>						
	0	1	2	3	4	5
<i>...and reported "2":</i>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<i>...and reported "3":</i>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<i>...and reported "4":</i>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<i>...and reported "5":</i>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

<i>...had a die-outcome of "2"</i>						
	0	1	2	3	4	5
<i>...and reported "3":</i>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<i>...and reported "4":</i>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<i>...and reported "5":</i>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

<i>...had a die-outcome of "3"</i>						
	0	1	2	3	4	5
<i>...and reported "4":</i>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<i>...and reported "5":</i>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

<i>...had a die-outcome of "4"</i>						
	0	1	2	3	4	5
<i>...and reported "5":</i>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Figure C.2.10: Punisher Experiment (for the Behavioral Experiment in Study 1), Screen #9

Some final questions

The study is now over. Before its end, we ask you to provide us with some information about yourself.

So far, you were not explicitly told how many Participant As took part in this study.

Please tell us your best guess:

For every 100 participants in the current study (Study C), how many Participant As do you think are there in the previous study (Study P)?

(enter an integer number)

Examples:

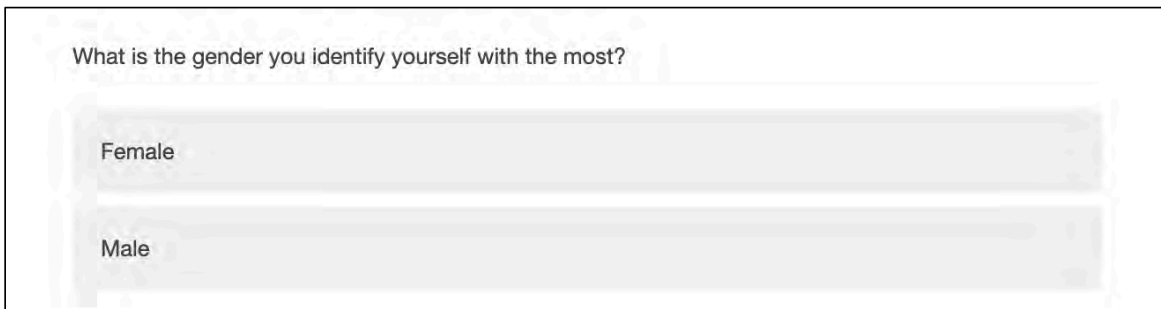
- If you enter "50", this means that you expect 50 Participant As in **Study P** for every 100 participants in **Study C** (e.g., every second participant of **Study C** determines the punishment for one Participant A in **Study P**).
- If you enter "100", this means that you expect 100 Participant As in **Study P** for every 100 participants in **Study C** (e.g., every participant of **Study C** determines the punishment for one Participant A in **Study P**).
- If you enter "200", this means that you expect 200 Participant As in **Study P** for every 100 participants in **Study C** (e.g., every participant of **Study C** determines the punishment for two Participant As in **Study P**).

Figure C.2.11: Punisher Experiment (for the Behavioral Experiment in Study 1), Screen #10



What is your year of birth?

Figure C.2.12: Punisher Experiment (for the Behavioral Experiment in Study 1), Screen #11

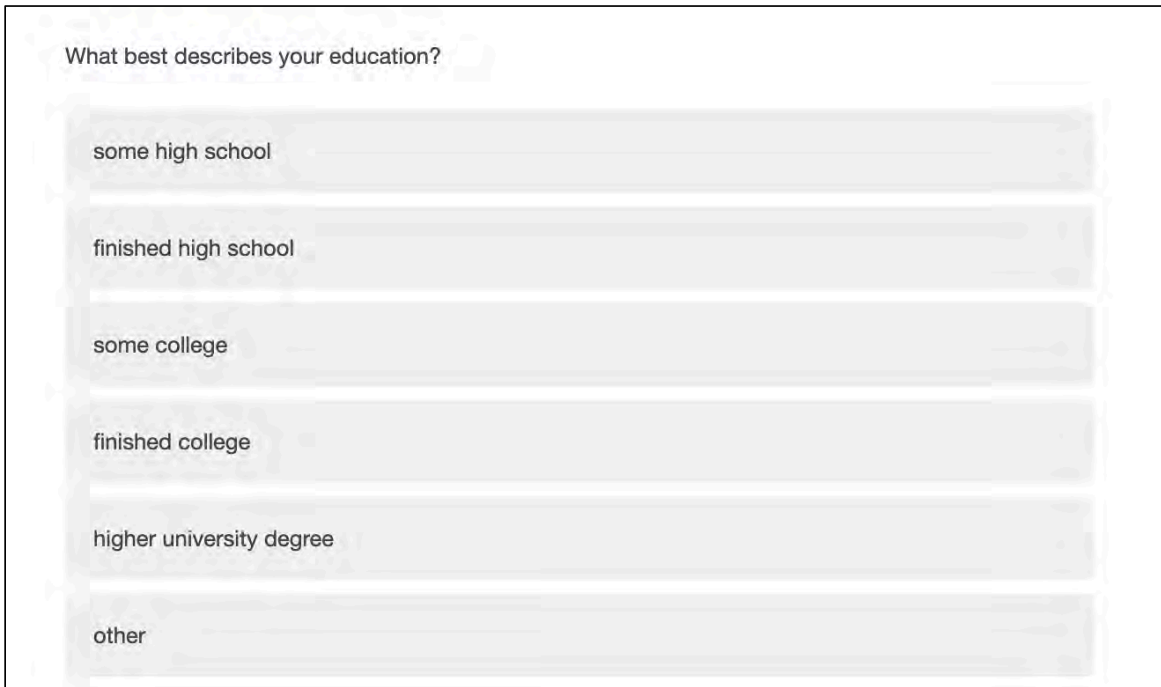


What is the gender you identify yourself with the most?

Female

Male

Figure C.2.13: Punisher Experiment (for the Behavioral Experiment in Study 1), Screen #12



What best describes your education?

some high school

finished high school

some college

finished college

higher university degree

other

D.3: Instructions for the Vignette Experiment

The following pages display the screens for the Vignette Experiment. All subjects were in the same role as a co-worker of Alex. The provided norm information and lying scenario differed between subjects. The version shown here are from the treatment INFO. In treatment NO INFO, the norm information (blue boxes in figures C.3.7 and figure C.3.12) was not shown. Each subject was shown two equity condition. Here, the EQUITY-condition is shown first (see Figure C.3.3) and the INEQUITY-condition afterwards (see Figure C.3.9); for other subjects the order was reversed.

Figure C.3.1: Vignette Experiment in Study 1, Screen #1

Informed Consent

The purpose of this research study is to understand human decision-making. You will answer a series of questions regarding beliefs and behaviors. We will also ask you to provide demographic information at the end. We will not ask for your name or any personal information. Overall, this study will take approximately 5 minutes. All decisions you make during this experiment will remain anonymous. The risks of participating are no greater than those encountered in everyday life. Your participation in this study is completely voluntary, and you may refuse to participate.

At the end of the experiment, you will have the opportunity to enter a raffle lottery (with your e-mail address) and win **\$100** (in form of 1 of 3 Amazon gift cards). If you withdraw from the study before completing it, you will not be paid. Compensation will be awarded upon completion of the entire study.

If you have any questions about this study, you may contact us at edimant@sas.upenn.edu.

Please note: *Participating in this requires your full attention. We will have some questions checking your attention.*

I have read and understood this consent form and wish to continue the study.

I do not wish to continue the study.




Figure C.3.2: Vignette Experiment in Study 1, Screen #2

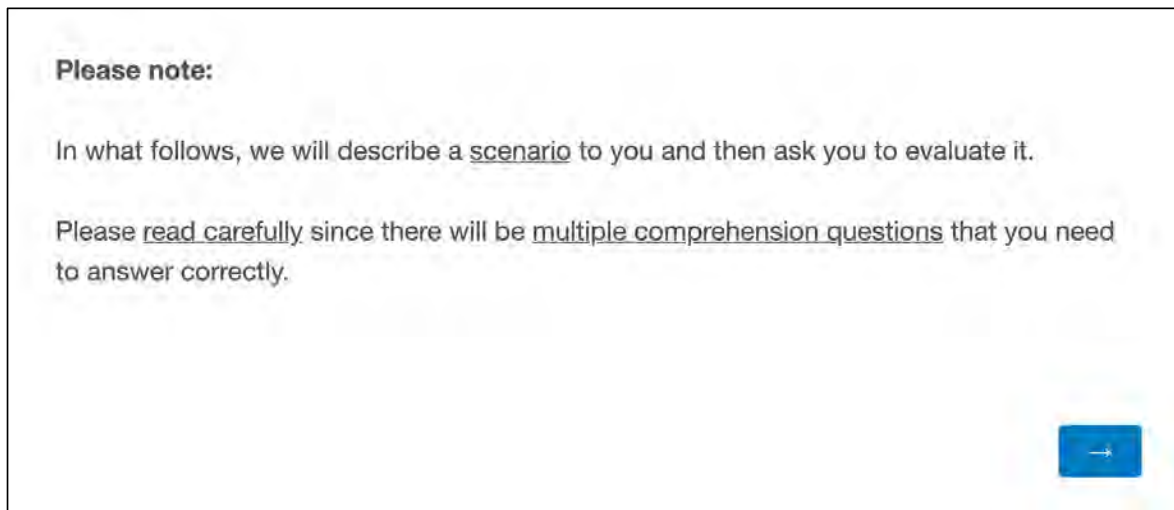


Figure C.3.3: Vignette Experiment in Study 1, Screen #3a

WELCOME!

It is important that you take the time to read all instructions and that you read questions carefully before you answer them. Previous research has found that some people do not take the time to read everything that is displayed in surveys. The questions below serves to test whether you actually take the time to do so. Therefore, if you read this, please answer "six" to the first question, divide that number by two and use the result as the answer to the second question in order to participate.

Your answer to the first question:

very little: 1	2	3	4	5	very much: 6
-------------------	---	---	---	---	-----------------

Your answer to the second question:

very little: 1	2	3	4	5	very much: 6
-------------------	---	---	---	---	-----------------

Consider the following situation:

You are an employee of a company. Alex, a colleague of yours with the same rank and level of seniority, is responsible for submitting an offer (tender) for a large project via postal mail. Well in advance of the tender's deadline, Alex sends you the tender document for review. You review and approve it. Then, you send the tender via postal mail on the same day.

However, due to unforeseen weather conditions, the post gets delayed and the tender documents do not arrive on time.

- Consequently, the **company loses out on the opportunity to win the project.**
- Since the late letter was not Alex's fault, **Alex receives an end-of-year bonus of \$10,000.**
- **You also receive a bonus of the same size (\$10,000),** which you immediately use in full to pay outstanding debt you have.

Figure C.3.4: Vignette Experiment in Study 1, Screen #3b

A few months go by and Alex is tasked with managing the remodeling of the office space in your company. Over lunch, Alex tells you that *IntDesign Ltd.* made the best offer for the remodeling.

A few weeks later, you are surprised to see another company, *CraftSpace Inc.*, doing the remodeling of your office instead. As you go to Alex's office to inquire why he did not choose the best offer by *IntDesign Ltd.*, you see it abandoned. However, you notice a document issued by *CraftSpace Inc.* It contains the following handwritten note:

"Dear Alex, in exchange for your helpful collaboration and as per our agreement, we will waive the price of \$10,000 for converting your home attic into a home cinema."

You quickly take a photo of the document with your phone and leave the office.

Given the evidence, it becomes clear to you that Alex took a bribe from *CraftSpace Inc.* and that the best offer by *IntDesign Ltd.* was not taken. Your company has established multiple mechanisms that deal and punish such behavior. In addition, it has also instituted whistleblower-protection guidelines. However, the company has to rely on people with relevant information - you in this case - to report this deviance and associated evidence.

You can now take different actions.

Each action influences the chance that Alex's behavior is caught and punished as follows:

Figure C.3.5: Vignette Experiment in Study 1, Screen #3c

Action	Chance of detection for Alex
Take no action and look the other way	0 %
Report to another colleague and mutual friend	25%
Send a tip and the photo evidence to another colleague and mutual friend	50%
Call the company's whistleblower team and report the behavior	75%
Send a tip and the picture evidence to the company's whistleblower team	100%

Please click the "next"-button after you have read the above *carefully*.

Figure C.3.6: Vignette Experiment in Study 1, Screen #4

Please answer the following comprehension questions carefully to avoid being excluded from payment.

According to the previous description, does your action affect the chance that Alex gets caught for his behavior (accepting the bribe)?

Yes

No

Did Alex receive the same end-of-the-year (\$ 10,000) bonus as you?

Yes

No

Figure C.3.7: Vignette Experiment in Study 1, Screen #5

We now ask you to tell us how you would behave in the situation above. Depending on your action, the likelihood of Alex's behavior being detected and punished is different.

We ask you to recall the following main aspects of the situation:

- Alex, who was not responsible for the lost tender was **not held accountable** and **receives an end-of-year bonus** of \$10,000.
- **You also receive a bonus of the same size** (\$10,000), which you immediately use in full to pay outstanding debt you have.
- A few weeks later, you found a letter on Alex's desk indicating that **Alex took a bribe (amounting to \$10,000)**. You took a picture of this letter.

Please read this text and click this box in order to proceed:

In a previous study, we asked participants what they consider to be the right thing to do for Alex in the exact same situation explained above.

The majority of participants said that taking a bribe is highly unethical and not the right thing to do.

We now ask you: What action do you choose?

(you will not face any personal, monetary, or non-monetary consequences)

Take no action and look the other way
(chances of Alex being caught and punished: **0%**)

Report the behavior to another colleague and mutual friend
(chances of Alex being caught and punished: **25%**)

Send a tip and the picture evidence to another colleague and mutual friend
(chances of Alex being caught and punished: **50%**)

Call the whistleblower hotline and report the behavior
(chances of Alex being caught and punished: **75%**)

Send a tip and the picture evidence to the whistleblower hotline
(chances of Alex being caught and punished: **100%**)

Figure C.3.8: Vignette Experiment in Study 1, Screen #6



Figure C.3.9: Vignette Experiment in Study 1, Screen #7a

Please consider the following **new situation**.

NOTE: THIS SITUATION IS DIFFERENT TO THE ONE YOU HAVE SEEN BEFORE.

Please read the description carefully and see how this situation differs:

As before, the features of the situation are as following:

- Alex sent you the letter for review, you approved it and sent it via post.
- It also remains valid that the post gets delayed and the company does not win the tender for the project.

Different in the new situation: The effect of the lost tender on Alex's end-of-year bonus changes.

- Since Alex was responsible for the tender, **Alex does not receive an end-of-year bonus** of \$10,000, even though the late mail was not Alex's fault.
- **In contrast, you receive a bonus of the same size** (\$10,000), which you immediately use in full to pay the outstanding debt you have.

Also as before,

- You figure out that **Alex accepted a bribe worth \$10,000** from *CraftSpace Inc.*
- You are able to take a photo of the evidence with your phone.

As before, you can take different actions.

Each action influences the chance that Alex is caught and punished as following:

Figure C.3.10: Vignette Experiment in Study 1, Screen #7b

Send a tip and the picture evidence to the company's whistleblower team	100%
Call the company's whistleblower team and report the behavior	75%
Send a tip and the photo evidence to another colleague and mutual friend	50%
Report to another colleague and mutual friend	25%
Take no action and look the other way	0%

Figure C.3.11: Vignette Experiment in Study 1, Screen #8a

Please answer the following comprehension questions carefully to avoid being excluded from payment.

(Note that due to the differences to the first scenario some correct answers may have changed!)

According to the previous description, does your action affect the chance that Alex gets caught for his behavior (accepting the bribe)?

Yes

No

Did Alex receive the same end-of-the-year (\$ 10,000) bonus as you?

Yes

No

We now ask you to tell us how you would behave in this new situation.

Depending on your action, the likelihood of Alex's behavior being detected and punished is different.

Recall the following main aspects of the new situation:

- Alex, who was not responsible for the lost tender, was **held accountable** and **does not receive an end-of-year bonus** of \$10,000.
- **You, on the other hand receive a bonus of the same size** (\$10,000), which you immediately use in full to pay the outstanding debt you have.
- A few weeks later, you found a letter on Alex's desk indicating that Alex took a bribe (amounting to \$10,000). You took a picture of this letter.

Figure C.3.12: Vignette Experiment in Study 1, Screen #8b

Please read this text and click this box in order to proceed:

In a previous study, we asked participants what they consider to be the right thing to do for Alex in the exact same situation explained above.

The majority of participants said that taking a bribe is highly unethical and not the right thing to do.

We ask you: What action do you choose?
(you will not face any personal, monetary, or non-monetary consequences)

Take no action and look the other way
(chances of Alex being caught and punished: **0%**)

Report the behavior to another colleague and mutual friend
(chances of Alex being caught and punished: **25%**)

Send a tip and the picture evidence to another colleague and mutual friend
(chances of Alex being caught and punished: **50%**)

Call the whistleblower hotline and report the behavior
(chances of Alex being caught and punished: **75%**)

Send a tip and the picture evidence to the whistleblower hotline
(chances of Alex being caught and punished: **100%**)

Figure C.3.13: Vignette Experiment in Study 1, Screen #9

Thank you for completing the survey!
Before we finish, we would like to ask you a few questions.

What is your year of birth?

What is the gender you identify yourself with the most?

Female

Male

Other

Are you currently affiliated (active student, staff, faculty) with the University of Pennsylvania?

Yes

No, but with another university (please specify in the field below)

No current affiliation with any other university

Please enter your email if you would like to enter a lottery with the chance to win \$100.

D.4: Instructions for Study 2

The following pages display the screens for Study 2. All subjects were in the same role (determining the social appropriateness of lying) but the provided norm information and lying scenario differed between subjects. In each treatment, they were presented with either NO INFO, NORMATIVE, or EMPIRICAL information and the equity scenario was either Equity- (p13), Inequity- (p35) or Overclaiming-based (p24) lying. Here, the shown screen presents the text for the EMPIRICAL norm information (see, e.g., Figure C.4.5); the content for the other treatments was adjusted accordingly.⁵ The same applies for the Equity-scenario. Here, we present Inequity-based lying (see Figure C.4.7; p35: "Player A rolls a 3 and reports a 5"). The scenario description was adjusted accordingly for the other conditions.

⁵For NORMATIVE, the text was: "In a previous study, we asked people what they consider to be the right thing to do for participants in the role of Player A. Those people could not play the game themselves. The majority of these people stated that the right thing to do for Player A is to report the number truthfully (i.e., report exactly what the die showed)." In NO INFO, such information was not displayed.

Figure C.4.1: Experiment in Study 2, Screen #1a



University of Pennsylvania

Department of Philosophy, Politics, and Economics

Claudia Cohen Hall, Room 311

Philadelphia, PA 19104

Phone: (215)-898-3023

Fax: (215) 573-2231

Informed Consent/ Assent Form for Non-Pool Participants Earning Money

You are invited to take part in a study named *The Acceptability Determination Task*. The purpose of this research study is to explore human decision-making. You will complete a brief survey involving a determination of how acceptable you find a given behavior. If you agree to be in this study, you will need to make decisions and answer questions regarding the study materials. We will also ask you to provide demographic information. We will not ask for your name or any information that will make you identifiable. Overall, this study will take approximately 5 minutes.

For your participation in this study, you will receive a fixed payment of \$0.20. Additionally, you may receive a monetary bonus. The exact amount depends on your results in the experiment. The risks to participating are no greater than those encountered in everyday life. Your participation in this study is completely voluntary, and you may refuse to participate or withdraw from the study without penalty or loss of benefits to which you may otherwise be entitled. Compensation will be awarded upon completion of the entire study.

Figure C.4.2: Experiment in Study 2, Screen #1b

Results may include summary data, but you will never be identified. If you have any questions about this study, you may contact the Behavioral Ethics Lab at behavioralethicslab@gmail.com.

For any questions, concerns, suggestions, or complaints that are not being addressed by the researcher, please contact the Institutional Review Board at the University of Pennsylvania at phone: (215) 898-2614.


Please feel free to print or save a copy of this consent form.

By continuing from this page you are indicating that you have read and understood this consent form and wish to continue your participation in this study.

Consent



Figure C.4.3: Experiment in Study 2, Screen #2



Thank you for choosing to participate in this study! It is important that you **read all of the instructions carefully to maximize your earnings**. This is a survey-based study where you will be asked to answer questions and complete simple tasks. At the end of the study, you will be asked to fill out a short questionnaire and demographic questions.


It is expected that this study will take less than 5 minutes. There is **no deception** in this study. Everything you see or read is true.

You begin the survey with an endowment of \$0.20 (your fixed payment) and will have the opportunity to earn more based on your answer. You will be presented with a statement and asked to determine which answer you believe that most participants would be more likely to give. Guessing correctly will earn you an additional \$0.20.

Your bonus will be determined by how well you can anticipate the answer given by the other participants of this study, who will have the same information and condition for earning a bonus as you. It is therefore important that you read the following carefully.

You may stop participating in this study at any time; however, you will not receive your payment unless you complete the entire survey through to the end. You will receive a completion at the end of the survey.

Figure C.4.4: Experiment in Study 2, Screen #3a



In what follows, we describe some scenarios to you based on a series of actual prior experiments. The scenarios below describe the original experimental setup truthfully. In particular, this experiment involved real stakes and the bonus was paid accordingly to the winning participant.

A pair of two participants in a previous experiment played a quick game for the chance to win an additional bonus. They were paid a basic wage for their participation, and the bonus for winning was worth 500% of this basic wage. However, only one of them could actually win the bonus. Who won was determined by chance.

How it worked:

- One participant within the pair, called Player A, was randomly chosen to toss a fair, five-sided, virtual die (with numbers "1", "2", "3", "4", "5"). The number reported by Player A determined how many lottery tickets Player A got.
- Player A was asked to report what the die shows and received the corresponding number of lottery tickets. Because the dice roll was blind, Player A could report a higher number than they actually rolled and earn more tickets through their deception.
- The other participant, called Player B, got a fixed amount of 3 lottery tickets.
- Player A and B's lottery tickets were put into a (virtual) raffle from which one ticket was randomly drawn. The participant to whom this ticket belonged earned the bonus.

Figure C.4.5: Experiment in Study 2, Screen #3b

Below, we provide you with a few examples. Please familiarize yourself with the following statements and check each to confirm that you have read and understand them.

In a previous study, the majority of Player As chose to report the number truthfully.

- Example: Player B already has **3** tickets. Player A rolls a **3** and reports a **1**. Player A receives **1** ticket. The chance for Player A to win the bonus is **$1/4 = 25\%$** .
- Example: Player B already has **3** tickets. Player A rolls a **3** and reports a **6**. Player A receives **6** tickets. The chance for Player A to win the bonus is **$6/9 = \text{about } 67\%$** .
- Example: Player B already has **3** tickets. Player A rolls a **3** and reports a **3**. Player A receives **3** tickets as well. The chance for Player A to win the bonus is **$3/6 = 50\%$** .


Figure C.4.6: Experiment in Study 2, Screen #4



In a previous study, people played this game for themselves. The majority of people in the Player A position chose to truthfully report the number truthfully (i.e. report exactly what the die showed).

- Confirm: The majority of people playing as Player A reported the number truthfully.

Figure C.4.7: Experiment in Study 2, Screen #5

 **Penn Law**
UNIVERSITY of PENNSYLVANIA CAREY LAW SCHOOL

In what follows, you need to make a payoff-relevant guess.

Note: this survey will also be given to a number of other participants.

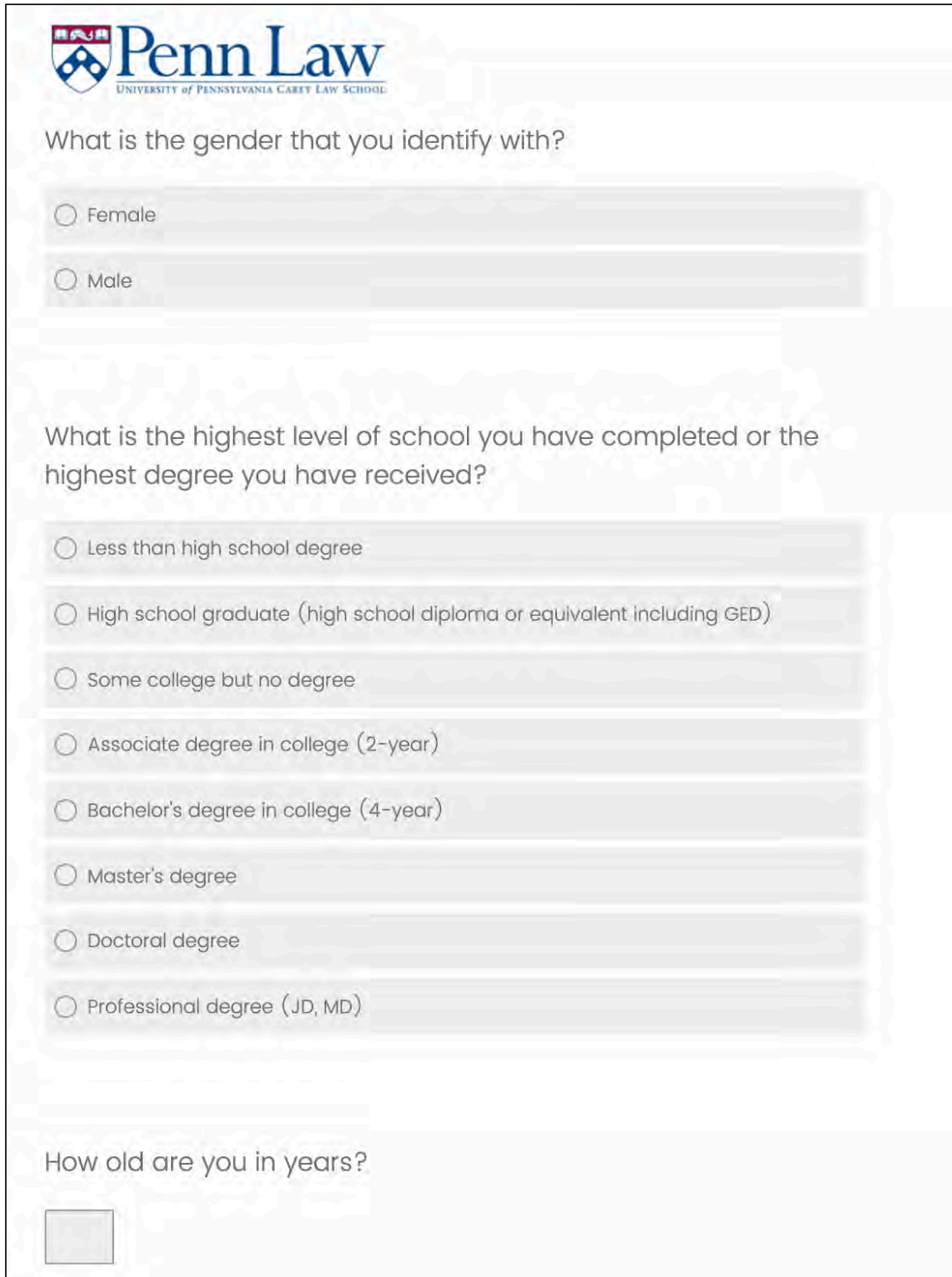
In order to earn additional money, you need to identify the answer that will be given the most often by the other participants. If you identify the most frequently given answer to the following statement, you will earn an additional \$0.20.


Remember: In a previous study, people played this game for themselves. The majority of people in the Player A position chose to truthfully report the number truthfully (i.e. report exactly what the die showed).

Please rate the appropriateness of the following behavior based on what you believe the most frequent answer will be in this survey:

	Very Socially Inappropriate	Somewhat Socially Inappropriate	Somewhat Socially Appropriate	Very Socially Appropriate
Player A rolls a 3 and reports a 5	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Figure C.4.8: Experiment in Study 2, Screen #6



 **Penn Law**
UNIVERSITY of PENNSYLVANIA CAREY LAW SCHOOL

What is the gender that you identify with?

Female

Male

What is the highest level of school you have completed or the highest degree you have received?

Less than high school degree

High school graduate (high school diploma or equivalent including GED)

Some college but no degree

Associate degree in college (2-year)

Bachelor's degree in college (4-year)

Master's degree

Doctoral degree

Professional degree (JD, MD)

How old are you in years?