Morality in Deceptive Environments: The puzzling effects of deceptive culture on moral behavior

Panagiotis Mitkidis (pm@mgmt.au.dk)
Aarhus University https://orcid.org/0000-0002-9495-7369

Philipp Gerlach
Hochschule Fresenius

Aaron Nichols
Questrom School of Business

Christian Elbaek
Aarhus University https://orcid.org/0000-0002-7039-4565

Sonja Perkovic
Aarhus University

Dan Ariely
Duke University

Article

Keywords: deception, social norms, morality, dishonesty, meta-analysis, experimental deception

Posted Date: March 30th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1444049/v1

Abstract

Psychologists, economists, and philosophers as well as lay people have long argued and intuited that in environments where deception is normative, moral behavior is harmed. After we confirmed this folk intuition, we challenge it and show that individuals making decisions within deceptive environments do not behave more dishonestly than in nondeceptive environments. We demonstrate the latter utilizing an example of experimental deception within established institutions, such as laboratories and institutional review boards. Specifically, we first find correlational evidence showing that experiencing deception at the laboratory level yields lower dishonesty. Then, across three studies, we empirically demonstrate that deceptive environments do not affect downstream honest behavior but that when participants were in a deceptive environment and aware of being observed, their dishonest behavior decreased. Our results show that the relationship between deception and dishonesty is more complicated than previous interpretations have suggested and expand the understanding of deception and moral behavior.

Introduction

Dishonesty, in the forms of fraud, bribing, and cheating, has been found to devastate the economy (Cebula & Feige, 2012; Gee & Button, 2019; Warren & Schweitzer, 2019), decrease societal trust (Kirchler et al., 2008; Banerjee, 2016; Butler et al., 2016), escalate corruption (Olsen et al., 2019) with shocking results (Ambraseys & Bilham, 2011), spread further unethical behaviors (Robert & Arnab, 2013), and numb our moral compass, creating deleterious cultural, organizational, and social norms (Gächter & Schulz, 2016). People regularly have opportunities to behave dishonestly (Ariely, 2012). However, these moral decisions are not made in a vacuum. People often make moral decisions collaboratively (Weisel & Shalvi, 2015) and typically within an organizational, social, or cultural environment. In other words, moral decisions are normative, which means that they are consistent with the normative principles acknowledged or rejected by others (Habermas, 1990; Stansbury, 2009). Therefore, a critical question is: what happens to our moral fiber in environments where social norms allow deception? How does an organizational culture that permits or tolerates deception affect our moral judgments and subsequent behavior? In particular, what effect does a laboratory in which experimenters deceive participants have on the moral behavior of these participants?

Some possible answers to these questions can be drawn from the psychological theory of self-concept maintenance (Mazar, Amir, & Ariely, 2008; Hilbig & Hessler, 2013; Thielman & Hilbig, 2019). According to this model, people balance two motivations: the desire to maintain a positive self-image as an honest person and the desire to use the means at their disposal (i.e., deception) to maximize personal welfare. The more justifiable a dishonest act is, the better the maintenance of the positive self-image is and therefore the higher the dishonesty is. Hence, self-concept maintenance theory predicts that people will engage in deception when it is justifiable (Schweitzer et al., 2002; Shalvi et al., 2012). Following this logic, we could hypothesize that people would deceive within an environment in which deception is publicly acknowledged since a deceptive environment would provide them with the necessary justification to do so. In other words, environments in which deception is allowed—if social norms in an

environment condone deception—make deceptive behavior justifiable because it is normative (socially appropriate behavior – Köbis et al., 2018).

Social norms (Cialdini et al., 1990; Bicchieri, 2006) and social exchange theories (Cropanzano & Mitchell, 2005) suggest that people learn about social norms by interacting with others and that they reciprocate other people's behavior. Following these predictions, in environments that permit deception, social norms will be similar, undermining honesty. In other words, in environments that permit deception, deceptive behavior can be understood by actors as a social norm[1]. Empirical data supporting the influence of the deceptive social norms hypothesis can be found in research on collaborative dishonesty and the spread of deception. Sometimes, collaborative settings (Weisel & Shalvi, 2015), market interactions (Falk & Szech, 2013), negotiation cues (Rees et al., 2019), and contexts of commitment (van Baal et al., in review) increase dishonest tendencies. For example, interpersonal deception has been shown to increase lying (Boles et al., 2000; Croson et al., 2003) and decrease trust (Schweitzer et al., 2006) and ethical behavior and has been defined as morally unacceptable to the larger community (Schweitzer & Gibson, 2008). Similarly, cues of criminal behavior might further encourage unethical behavior (Keizer et al., 2008). The influence of deceptive social norms is further corroborated by research on the contagion of dishonesty, which has found that signals of dishonesty create more dishonesty (Robert & Arnab, 2013), that bribing negatively affects downstream moral behavior (Nichols et al., in review), and that corruption undermines cooperation (Muthukrishna et al., 2017). A meta-analysis by Bellé and Cantarelli (2017) reinforces the robustness of these findings by concluding that increases in unethical behavior are predicted by exposure to unethical others or incremental increases in dishonesty.

Beyond small interpersonal settings, individual and dyadic, there is research suggesting that larger organizational and cultural settings can influence dishonesty. However, the evidence from these settings seems more mixed. For example, Cohn and colleagues (2014) found that salient cultural cues, particularly participants' professional identity in the banking business sector, undermined honesty norms. However, Rahwan and colleagues (2019) did not find such an effect. Falk and Szech (2013) found that market interactions decay morality (also: Gerlach, 2017). Ariely and colleagues (2019) extended this evidence to other economic environments, showing that long-term exposure to a specific economic system, such as socialism, can have negative implications for moral behavior. Based on a series of cross-cultural experiments, Gächter and Schulz (2016) reported a link between rule violation in society and individual dishonesty. Additionally, Hugh-Jones (2016) found evidence of a positive relationship between honesty and economic growth. However, three other cross-cultural studies found no such link (Pascual-Ezama et al., 2015; Dieckmann et al., 2016; Mann et al., 2016). One such study, however, indicated the possibility that a person's close network can predict lying tendencies (Mann et al., 2014). Finally, research exploring the connections between social context and moral behavior has found that identity (Dreber & Johannesson, 2008) and other social and religious norms (Mazar & Aggarwal, 2011; Piff et al., 2012; Mitkidis et al., 2017) can lead to differential results in regard to dishonest behavior. Overall, there seems to be a strong link between social cues and dishonesty that is not reflected by existing cross-cultural evidence.

A meta-analytical study (Gerlach et al., 2019) found evidence that some situational factors, such as the experimental setting and normative cues (socially appropriate behavior), might influence dishonest behavior but provided intriguing indications regarding the effect of deception on subsequent morality. Specifically, the authors reported that in experimental settings where participants were themselves deceived, they lied *less* compared to participants who were not deceived. Although this correlational finding seems counterintuitive, it might be a natural consequence of the aforementioned mixed evidence of cultural influence on deception and the differential effects observed when studying deception in laboratory settings (for example, see Bonetti, 1998; Ortmann & Hertwig, 2002), with most literature suggesting that there must be a morally undesired effect of laboratory deception on downstream behavior (e.g., Stricker et al., 1967; Christensen, 1977; Jamison et al., 2008).

A different, nonmutually exclusive explanation of Gerlach and colleagues' (2019) counterintuitive finding is that participants exposed to deceptive experimental designs might be suspicious of their personal anonymity. The effects of suspicion on experimental control and on participants' behavior are not negligible (2008a). When detection seems possible, participants in deceptive environments might behave more honestly because they fear being exposed as liars (Kimmel, 1996, p 68; Gneezy et al., 2018). Indeed, research has found conflicting evidence that participants' behavior was altered when they had specific but not general knowledge about experimental deception (Ortmann & Hertwig, 2002; Hertwig & Ortmann, 2008a, 2008b).

Overview of the current research

In this paper, we provide evidence that challenges the prevailing assumption and folk intuition that deceptive environments in general cause dishonesty. We argue that this understanding of deceptive environments assumes that experimental deception affects participants' expectations and, in turn, participants' subsequent behavior because people react to dishonest environments, institutions, and organizations as they to dishonest peers (Bok, 1978; Boles et al., 2000; Croson et al., 2003; Schweitzer et al., 2006; Schweitzer & Gibson, 2008). We thus propose that prior research has conflated deception with direct interpersonal lying and self-serving intentions. We therefore reason that prior research on the harmful effects of deception can refer to and account for only interpersonal dishonest behavior and that these empirical findings do not generalize to deceptive environments, such as laboratories using deception.

Consistent with prior research, we define deception as *the transmission of information that is intentionally erroneous or that intentionally misleads others* (Gino & Shea, 2012; Levine & Schweitzer, 2015). Across a survey, a meta-analysis and three experiments, we demonstrate that—despite people's intuition that deceptive environments causally increase dishonesty—deceptive environments generally do not affect dishonesty but *only* when participants' individual behavior can be observed. If so, deceptive environments *decrease* dishonesty, making people more honest, as the fear of detection is higher. Our

findings are in line with the idea that deceptive experimental designs induce detection suspicion among participants about their anonymity in experiments.

We first establish that the prevailing assumption and folk intuition is that deceptive environments harm morality by surveying lay people's intuitions about moral behavior in deceptive environments (Intuition Study). We find that people expect deceptive environments to increase dishonest behavior.

Thereafter, we use a meta-analytical study to examine the correlational effect of experimental deception on participants' dishonesty. In particular, we explore the effect of real lab policies concerning the use of deception in behavioral research across four experimental paradigms: die-roll tasks[2] (Fischbacher & Föllmi-Heusi, 2013), sender–receiver tasks[3] (Gneezy, 2005; Capraro, 2018), coin-flip tasks[4] (Bucciol & Piovesan, 2011), and matrix tasks[5] (Mazar et al., 2008). In our meta-analysis, we tested whether participants behaved more dishonestly in situations in which the laboratory policy either prohibited or permitted deception. Our results suggest that laboratory policies that allow deception were associated with less dishonest behavior. This finding is the direct opposite of lay people's intuitions (see Intuition Study). Although our meta-analysis yielded robust results, insights from meta-analyses are generally limited, as they provide only correlational results, while the underlying mechanisms remain unclear.

We therefore collected causal evidence and tested the implications of our meta-analysis by conducting a series of experiments. We began our primary investigations with a laboratory study (Study 1) in which participants were randomly assigned to read and sign one of three versions of the consent form. Crucially, the consent forms in Study 1 included a description of the laboratory's deception policy or, in the control group, did not describe the laboratory's deception policy. In the treatment conditions, participants learned that the experiment could or could not contain misleading information about the experiment. Manipulating awareness that deception could (or could not) occur in the laboratory allowed us to make causal inferences about the role of deceptive environments in dishonest behavior. After obtaining participants' consent, participants completed a repeated and incentivized private die-roll task. The results from Study 1 indicate that being in a deceptive environment did not predict consequential cheating behavior. More precisely, the decision to cheat in a die roll task was unaffected by the presence (or absence) of laboratory policies on deception.

To test the robustness of the null finding in Study 1, we report the results from two online studies (Studies 2 and 3), which tested larger samples to replicate and expand on the null effect of experimental deception on dishonest behavior. In Studies 2 and 3, participants were informed that the policy of the Institutional Review Board (IRB) allows experimental deception (vs. prohibits vs. control). In Study 2, we used a similar design to test the robustness of the null finding of Study 1 in regard to cheating behavior in a die roll task. We also extended our investigation by including one-shot die roll tasks, allowing us to test for differential effects on moral behavior due to potential learning effects in shorter (one-shot) vs. longer-term (repeated; Study 1) tasks. We also introduced a different task, the sender-receiver game, to examine dishonesty. In the sender-receiver game, lying victimizes another participant instead of the experimenter (as is the case in our die-roll task). Importantly, the sender-receiver game allowed us to directly identify individual

dishonest behavior, whereas in the die-roll task, cheating behavior is indirectly identified by comparing expected vs. real outcome distributions. The results from Study 2 were consistent with those from Study 1; again, we observed no effect of deceptive environments on dishonest behavior.

Finally, in Study 3, we used a similar design as in previous studies and further tested a potential mechanism. Specifically, we included two additional conditions to examine whether cues of anonymity and being observed could affect dishonest behavior. We again informed participants that the policy of the IRB allows experimental deception (vs. prohibits vs. control) in conditions where their die roll performance could be observed (vs. unobserved) and then asked them to complete a one shot die roll task. We found that informing participants about nondeceptive environments leads to less dishonesty. We also found that when deception is allowed and participants are aware of being observed, then honesty increases, echoing previous theoretical accounts (Ayal et al., 2015) and findings (Zhong et al., 2010; Ernest-Jones et al., 2011; Nettle et al., 2012; Pfattheicher & Keller, 2015; Schild et al., 2019) that cues of being observed (i.e., anonymity or visibility) decrease cheating, which suggests that participants are more skeptical about their anonymity when experiencing experimental deception.

Open Science Statement and Ethics

All studies[6] were preregistered and did not deviate from the preregistered protocol. All studies followed open science practices: all material, including the source code, collected anonymized raw and processed data, consent forms, stimuli, and surveys, were shared and made publicly available as online supplementary material (*SM*) [Anonymized for peer-review] of the project (https://osf.io/ca769/?view_only=c137cb475eb74d3f831544d54b1c97f8). The codes for data management and statistical analyses were written in the statistical environment *R* (version 4.0.3). No conditions or variables were dropped from any analyses we report (except in Study 1, see footnote #10). All protocols were approved by the IRB at Duke University (#A0642) or the Committee on Health Research Ethics for Region Midtjylland (107/2019 (sagsnr. 1-10-72-1-19)) and the Aarhus University Research Ethics Committee (Journal no.: 2020-0169943, Serial Number: 2020-76).

[1] For a review on social norms as both psychological states and collective constructs and on how social norms inform action-oriented decision-making, see Legros & Cislaghi, 2020.
[2] The die-roll task is a commonly used paradigm to measure cheating behavior. In the standard form of the task (one-shot individual decision-making), the participant is asked to roll a 6-sided die under a cup to ensure nondetection and more than once to test whether the die is fair. However, the participant is instructed to report the number rolled the first time, which will determine the participant's payment. In this task, cheating means reporting a higher number than the actual number rolled the first time.
[3] The sender-receiver game is based on cheap-talk communication and is used to measure lying to another participant (instead of the experimenter). In its most standard form, it involves two players

(Player 1 and Player 2): Player 1 receives, from the experimenter, some information about a monetary allocation that is profitable to either Player 1 or Player 2. Player 1 is then asked to report the allocations to Player 2. Player 1 can thus decide to either truthfully or deceitfully convey information about the allocations to Player 2. Finally, Player 2 can decide to believe or distrust the report of Player 1.
[4] The coin-flip task is relatively similar to the die-roll task. In its standard form, the participant tosses a coin privately and reports the result. The participant knows the incentive to report the prize-winning side of the coin (given by the experimenters). Cheating in the coin-flip task is measured by comparing the reported distribution of coin flips with the expected 50/50 distribution of coin flips.
[5] The matrix task is a search task in which participants are asked to find the two numbers that add to 10 across several matrices containing a set of several, usually 3-digit, numbers. Participants are asked to report the number of correctly solved matrices and then discard or recycle their answer sheets. Unknown to participants, the matrix sheets are not thrown out but are later verified by the experimenter so that the reported number of solved matrices can be compared with the actual number of solved matrices. Cheating in the matrix task is therefore measured by comparing the actual number of solved matrices to the reported number of solved matrices.
[6] In the experimental studies, we employed deception. We argue that our decision is justifiable because it enabled us to maintain experimental control of our manipulation to solve the debate. Furthermore, the use of deception in our studies did not result in any harm to participants, except from what is to be expected in normal everyday life, which was also highlighted in our consent forms. Participants were fully debriefed.

Intuition Study

Method

Participants and Sample Size Estimation

Three hundred ninety-eight adult participants were recruited by the online platform Prolific Academic, according to our preregistered sample size estimation (see *SM*). Following preregistered exclusion criteria, individuals who failed the attention check and/or manipulation check questions or did not complete the study (screening criteria) were excluded, resulting in 182 participants remaining for the analysis (mean age = 33.09, SD = 11.89, 69.87% female, and 33% had finished at least a BA degree). There were many exclusions, as several participants incorrectly answered the attention check (n = 44) and control questions (n = 172). All participants were compensated for completing the study (avg. £8.80/hr). The study lasted approximately 5 minutes.

Materials and Procedure

Participants first saw a typical consent form informing them about the purpose, risks and benefits, time required for the study, screening criteria, and treatment of their personal data. Participants were then

asked if they wanted to continue. Participants who signed the consent form continued by reading the instructions and a detailed description of a planned experiment with three conditions (no-deception vs. deception-allowed vs. control) and two measures of dishonest behavior (die-roll and sender-receiver tasks). Initially, participants were told that *we plan to randomly assign participants to one of three different conditions. In these conditions, some participants will be led to believe that our Lab's policy is: 1. No-Deception, some other participants will be led to believe that our Lab's policy is 2. Deception-Allowed, and finally, some other participants will be informed that there is 3. No Information about our Lab's policy*. [1]

Participants were also informed about the incentives scheme on our two different measures: 1. a one-shot die-roll task, where cheating was quantified as 1 pip = 10¢, 2 pips = 20¢, 3 pips = 30¢, 4 pips = 40¢, 5 pips = 50¢, but 6 pips = 0¢), and 2. a sender-receiver game, where cheating was quantified as truthful or deceitful information sent from Player 1 to Player 2 and where the two payment options were *Option A: 40 cents to you and 50 cents to the other participant and Option B: 50 cents to you and 40 cents to the other participant*. Afterward, participants were asked several attention and control questions to ensure that they understood the instructions and nature of the experiment.

In our main measure of intuitions, participants were asked to report their opinion on sliding bars about the degree to which participants would cheat under the different conditions in the different measures. Finally, participants answered demographic items (age, gender, and education).

Results

Die roll predictions

Across the three conditions, participants in the intuition survey predicted that die roll claims would on average be 31.73¢ (SD = 11.08), which is significantly greater than the expected outcome of a fair die roll, 25¢ (one sample *t test*; $t(181) = 14.20$, 95% CI [30.79, 32.66], $d = 1.05$, $p < .001$).

At the condition level, participants predicted that die roll claims would, on average, be highest in the Deception-Allowed condition (M = 35.45¢, SD = 12.05), followed by the Control condition (M = 31.33¢, SD = 10.48), and last, the No-Deception condition (M = 28.41¢, SD = 9.48¢). A one-way analysis of variance (ANOVA) suggested significant differences in predictions between the three conditions ($F(2, 1635) = 59.51$, 90% CI [0.05, 0.09], $\eta_p^2 = 0.07$, $p < .001$), with participants predicting significantly higher die roll reports for the Deception-Allowed condition ($\beta = 7.03$, 95% CI [5.76, 8.30], $t(1635) = 10.86$, $p < .001$; *Std.* $\beta = 0.63$, 95% CI [0.52, 0.75]), as well as the Control condition ($\beta = 2.92$, 95% CI [1.65, 4.19], $t(1635) = 4.50$, $p < .001$; *Std.* $\beta = 0.26$, 95% CI [0.15, 0.38]).

Hence, compared to experiments that prohibited deception, participants intuitively believed that experimental participants would report more lucrative die roll outcomes in 1) experiments where

deception was allowed and 2) in experiments where participants were not informed about the IRB's stance on the use of deception. The distribution of predictions is illustrated in Figure 1.

Sender-Receiver predictions

Overall, participants in the intuition survey predicted that experimental participants across the three conditions would decide to send a deceitful message to the other player 52.21% of the time. Recent meta-analytical evidence has identified that people on average send a deceitful message in the sender-receiver game 51% of the time (Gerlach et al., 2019). This evidence indicates that the predictions of intuition survey participants were surprisingly close to the empirical evidence on observed behavior in these tasks and thus not statistically significantly different from this base rate (one sample *t test*; $t(181) = 1.06$, 95% CI [49.95, 54.47], $p = .292$).

Between conditions, compared to the Control condition ($M = 53.9\%$), participants identified that deceitful behavior would be significantly higher in the Deception-Allowed condition ($M = 66.14\%$; $c^2 = 345.64$, df = 84, $p < .001$) but significantly lower in the No-Deception condition ($M = 36.59\%$; $c^2 = 381.18$, df = 88, $p < .001$). Consequently, a one-way analysis of variance (ANOVA) revealed a significant effect of condition on predictions of deceitful behavior in the task ($F(2, 1635) = 202.70$, 90% CI [0.17, 0.23], $\eta_p^2 = 0.20$ $p < .001$), with positive significant effects on both the deception-allowed condition ($\beta = 29.54$, 95% CI [26.65, 32.44], $t(1635) = 20.04$, $p < .001$; *Std.* $\beta = 1.09$, 95% CI [0.98, 1.19]) and the control condition ($\beta = 17.31$, 95% CI [14.42, 20.20], $t(1635) = 11.74$, $p < .001$; *Std.* $\beta = 0.64$, 95% CI [0.53, 0.74]). The distribution of predictions is illustrated in Figure 1.

Discussion

In this intuition study, participants predicted that deceptive environments would induce dishonest behavior among participants. This prediction aligns with social norms explanations concerning the effect of deceptive environments on individuals' downstream ethical behavior. However, do these intuitions reflect empirical findings? To answer this question, we performed a meta-analysis, using as a starting point the meta-analysis by Gerlach and colleagues (2019), which suggests that deceptive laboratory policies would *decrease* participants' dishonesty.

[1] In particular, participants were informed that: *Our IRB (institutional review board) prohibits (vs. allows) the use of deception in all experimental protocols. In other words, our IRB does (not) allow giving participants misleading or erroneous information about (elements of) the study conducted.* Or that: *Our IRB (institutional review board) has approved this experimental protocol.*

Meta-analysis

Method

To test whether deceptive laboratory policies actually affect participants' dishonesty, we reanalyzed a recently published meta-analysis (Gerlach et al., 2019). The meta-analysis encompassed 565 experiments (N = 44,050) across four experimental paradigms: sender–receiver tasks (Gneezy, 2005), coin-flip tasks (Bucciol & Piovesan, 2011), die-roll tasks (Fischbacher & Föllmi-Heusi, 2013), and matrix tasks (Mazar et al., 2008).

To compare dishonest behavior across paradigms, Gerlach and colleagues (2019) prepared a standardized report called M_r. The standardized report could range from 100%, suggesting that all participants maximally misreported (e.g., in a die roll task: all participants claimed the highest reward) over 0%, suggesting that participants reported honestly (e.g., participants claimed, on average, the expected outcome of a fair die), to –100%, suggesting that all participants (oddly) claimed the lowest possible reward (e.g., all participants claimed to have observed the least rewarding pip). The standardized report thus quantifies the level of dishonest behavior and facilitates comparison of the various experimental paradigms.

Following the methodology of Gerlach and colleagues (2019), we combined standardized reports by using random effects models. Then, we tested personal and situational moderators via meta-regression models. In their meta-analysis, Gerlach and colleagues (2019) tested the following moderators: experimental paradigm (sender–receiver tasks vs. coin-flip tasks vs. die-roll tasks vs. matrix tasks), investigative setting (laboratory experiments vs. online/telephone experiments vs. field experiments), participant characteristics (noneconomics students vs. nonstudents vs. economics students vs. MTurk workers), normative cues (e.g., whether participants were reminded about standards of moral conduct), maximal externality (e.g., the inflicted loss to a partner from following a misleading message in the sender–receiver tasks), maximal gain (i.e., the maximum incentive to misreport), and *experimental deception* as explained next.

In their moderator tests, Gerlach and colleagues (2019) measured the influence of experimental deception on dishonest behavior by creating a dummy variable that was measured at the study level and indicated whether the study's experimenter provided misleading information to the participants. For instance, some experimenters misinformed participants about the purpose of conducting a die-roll task. Other experimenters led sender-receiver game participants to believe that they would interact with a partner, even if no such partner existed.

Classifying deceptive vs. nondeceptive experiments in this way is a crude measure for experimental deception. For example, a study might be classified as nondeceptive even though the study was part of a larger series of experiments in which participants were in fact deceived. For our meta-analysis, we thus calculated an additional moderator for experimental deception. Specifically, we introduced a variable

capturing whether the study was conducted in a laboratory that had an existing deception policy, i.e., if the laboratory would (vs. would not) allow deception on its premises.

To classify the exact existing deception policy for all studies, we wrote to all authors of the primary studies asking them to identify their laboratory policies. Of the original 565 studies, the authors of 428 studies revealed their laboratory policies. Of these 428 studies, the laboratory policy of 294 experiments allowed deception, whereas the laboratory policy of 134 studies forbade deception. The authors of 137 studies did not reply to our emails (see *SM*). Therefore, we removed those experiments from our investigation. We then tested the moderating effect of such a *general* deception policy.

Results

To make comparisons of our and Gerlach and colleagues' (2019) findings easier, Model #1 in **Table 1** reports a direct replication of Gerlach and colleagues' moderator test (2019, table 3) using the full dataset of all 565 studies. Here, experimental deception at the *particular* study level lowered the standardized report by $b = -7.75$ percentage points (95% CI [−12.07%, −3.42%], $p < .001$).

Model #2 uses the same predictors as in Model #1 but is fitted to our subset of 428 studies for which we knew the laboratory policies. The results from Model #2 remain relatively similar to those of the original Model #1. Most importantly, from the perspective of the current research, even if fitted to the 428 studies, the results of experiencing deception in the *particular* study seemed robust: Experimental deception in the *particular* study lowered the standardized report by $b = -9.11$ percentage points (95% CI [−14.70%, −3.52%], $p = .001$). The only major difference between the results of Model #1 and Model #2 was that Model #1 suggested no difference between studies conducted in the laboratory vs. online/telephone ($b = -5.90\%$, 95% CI [−13.25%, 1.44%], $p = .115$), whereas our more limited Model #2 did ($b = -9.41\%$, 95% CI [−18.21%, −0.62%], $p = .036$). We return to the issue of laboratory setting later.

Next, in Model #3, we added our own moderator to the predictors, the *general* laboratory policy on experimental deception. The results suggest that over and above deception in the *particular* study ($b = -5.99\%$, 95% CI [−11.74%, −0.24%], $p = .041$), there was an effect of the *general* laboratory policies against deception ($b = -8.17\%$, 95% CI [−12.43%, −3.92%], $p < .001$). That is, allowing experimental deception was associated with less dishonest behavior among the participants.

Model #4 suggests that the effect of a *general* laboratory policy remained robust even when dropping the predictor of experimental deception in the *particular* study. That is, *general* laboratory policies against deception were still associated with $b = -9.43$ percentage points (95% CI [−13.54%, −5.32%], $p < .001$) less dishonest behavior among participants. Compared to all other models, Model #4 identified that studies conducted on Amazon's Mechanical Turk (MTurk) were associated with more dishonest behavior ($b = 9.69\%$, 95% CI [1.37%, 18.01%], $p = .023$). MTurk is an online platform that some researchers consider an affordable source of a fairly representative participant pool (Peer et al., 2017).

Overall, comparing Models #1 to #4, the two moderators did not seem to be robust predictors of dishonest behavior: experimental setting (laboratory vs. online/telephone) and MTurk. Both variables only partly moderated dishonest behavior. As a final step, we thus limited the dataset to studies that were conducted in physical laboratories, excluding all studies conducted via telephone, in the field or online (either on MTurk or elsewhere). As Models #5 and #6, respectively suggest, experiencing experimental deception in the *particular* study (b = −9.86%, 95% CI [−16.38%, −3.34%], p =.003) as well as in *general* (b = −17.32%, 95% CI [−22.52%, −12.11%], p <.001) still results in less dishonest behaviors in physical laboratories.

Table 1. Predictors of Different Measures of Dishonest Behavior Across the Four Paradigms

Data		*Experiments with known laboratory policies*				
					Physical laboratory only	
	Model #1	*Model #2*	*Model #3*	*Model #4*	*Model #5*	*Model #6*
Intercept	34.98%†	39.67%†	35.49% †	35.64%†	14.34%	–0.23%
	(3.78)	(4.07)	(4.77)	(4.81)	(7.82)	(7.83)
Experimental paradigm (coin-flip task)						
Sender–receiver	17.54%†	16.82%†	16.09% †	14.28%†	22.89%†	21.01%†
	(2.44)	(2.92)	(2.89)	(2.77)	(4.29)	(3.86)
Die-roll	–5.19%*	–5.37%	–7.96% **	–8.57%**	–2.12%	–9.15%*
	(2.55)	(2.99)	(3.04)	(4.46)	(4.13)	(4.01)
Matrix	–16.74% †	–20.63% †	–20.67%†	–21.46% †	–17.43% ***	–18.97% †
	(2.86)	(3.53)	(3.49)	(3.49)	(4.63)	(4.32)
Investigative setting (laboratory)						
Online/telephone	–5.90%	–9.41%*	–8.58%	–9.20%*	—	—
	(3.75)	(4.49)	(4.44)	(4.46)		
Field experiment	–12.91% †	–13.78% †	–13.45%†	–12.40% †	—	—
	(2.53)	(3.17)	(3.14)	(3.11)		
Participant sample (noneconomics students)						
Nonstudents	3.46%	–2.39%	–1.19%	–2.17%	20.87%**	29.48%†
	(3.29)	(4.21)	(4.17)	(4.17)	(7.16)	(6.95)
Economics students	0.38%	2.90%	2.08%	0.48%	–1.31%	–5.39%

	(2.62)	(3.19)	(3.16)	(3.08)	(3.91)	(3.50)
Mechanical Turk workers	5.18% (3.86)	5.49% (4.19)	8.87% (4.24)	9.69% * (4.24)	—	—
Normative cue	–15.70% (2.99) †	–14.99% (3.57) †	–13.13% (3.56) †	–11.77% (3.52) ***	–17.25% (4.14) †	–12.78% (3.91) **
Deception (no deception)						
Particular study received participants	–7.75% (2.21) ***	–9.11% (2.85) **	–5.99% (2.93) *	—	–9.86% (3.33) **	—
General laboratory policy allowed deception	—	—	–8.17% (2.17) ***	–9.43% (2.10) †	—	–17.32% (2.66) †
Maximal externality	–0.02% (0.07)	–0.02% (0.08)	0.01% (0.08)	0.02% (0.08)	–0.02% (0.10)	0.07% (0.10)
Maximal gain	–0.03% (0.06)	–0.02% (0.06)	–0.02% (0.06)	–0.01% (0.06)	–0.02% (0.06)	0.01% (0.06)
Observations	k = 558 n = 44,050	k = 428 n = 34,592	k = 428 n = 34,592	k = 428 n = 34,592	k = 243 n = 15,395	k = 243 n = 15,395
Residual heterogeneity	I^2 = 84% τ^2 = 0.02	I^2 = 84% τ^2 = 0.16	I^2 = 84% τ^2 = 0.15	I^2 = 84% τ^2 = 0.16	I^2 = 87% τ^2 = 0.16	I^2 = 85% τ^2 = 0.15
Heterogeneity accounted for	R^2 = 54%	R^2 = 48%	R^2 = 50%	R^2 = 49%	R^2 = 41%	R^2 = 49%

Note. Linear regression models with random effects at the experimental level. Values refer to beta weights with standard errors in parentheses. Except for *k, which* is the number of studies per meta-regression model with *n* being the total number of participants, I^2 is the study variance independent of the number of studies, τ^2 is the between-study variance, and R^2 is the heterogeneity accounted for by the predictors. Significant predictors are marked

as * p <.05; ** p <.01; *** p <.001; † p <.0001. Model #1 is the original model by Gerlach et al. (2019). Model #2 is the same as Model #1 but fitted to studies with known laboratory policies only. Model #3 is the same as Model #2 but with laboratory policies as an additional moderator. Model #4 is the same as Model #2 but without the original moderator of particular deception. Model #5 uses the same moderators as in Model #2, but it is limited to studies in physical laboratories only. Model #6 uses the same moderators as in Model #3, but is limited to studies in physical laboratories.

Discussion

Our meta-analysis showed an overall effect of experimental deception resulting in less dishonest behavior. At the study level, experiencing deception predicted less dishonest behavior. These results provide initial evidence against the common intuition (see Intuition study) that deceptive environments result in more dishonest behavior.

Although more precise than the original moderator of a *particular* deception, our measure of a *general* deceptive policy is also crude. Studies using experimental deception could involve naïve participants who have never encountered experimental deception. Hence, participants may not have expected deception in laboratories that did in fact allow deception. Indeed, it is likely that some participants were unaware of the laboratory's policy on deception when completing the study.

This point prompts several questions: Is the link between deceptive environments and honest behavior causal? If so, what drives the effect? An intriguing hint concerning the causal mechanism is suggested by Gerlach and colleagues' own analyses (2019, table 4). By differentiating between the four experimental paradigms, the authors showed that the effect of experimental deception in the *particular* study was driven by sender–receiver games only (see Study 2). This possibility would suggest an interaction effect of experimental deception and experimental paradigm.

Sender–receiver games differ from the other tasks in several regards. First, in sender–receiver games, participants (supposedly) interact with another participant to whom they can send a truthful vs. a misleading message. Therefore, behaving dishonestly in sender–receiver games means lying to another participant rather than the experimenter (Baumard & Sperber, 2010; Frollová et al., 2021). Second, in sender–receiver games, individual decisions can be linked to individual participants. In other tasks, dishonest behavior is measured by comparing the aggregated choices of a large group of participants to the distribution of outcomes expected from honest participants.[1] By design, die-roll tasks and coin flip tasks might therefore provide some form of anonymity to participants: individuals cannot be exposed for behaving dishonestly. For example, it is always possible that a single participant in fact observed the

highest pip claimed. The aggregated result would only seem dubious if all participants claimed the highest pip.

Hence, one reason why deceptive environments in sender–receiver games cause less dishonest behavior among participants could be that participants fear they might be exposed as liars and in response, show less dishonest behavior (Thielmann & Hilbig, 2018). Another explanation is that participants may not believe that their interaction partner truly exists. Hence, there is no reason to behave honestly and claim less for oneself. Furthermore, exposure to any form of deception may engender participant suspicion if participants become aware of it, as they may suspect that they are also being lied to about the promise of anonymity.

Certainly, it may be true that there is no such causal link in the first place. After all, the meta-analysis only provided correlational evidence. To further understand the effect of deceptive environments on ethical behavior, we thus examined the causal mechanism in a series of laboratory and online experiments, where we manipulated the information participants received about the lab policy/IRB and controlled for 1) the victim of dishonesty, 2) the duration of the behavior, and 3) whether the actual manipulation of deception was realized by participants. We began by testing experimentally whether repeated[2] die-roll tasks deliver the same results.

[1] In some form of the matrix tasks, experimenters deceive participants so that experimenters know how many matrices each participant actually solved, similar to sender–receiver games. However, the procedure is unbeknownst to participants. It allows experimenters to directly compare the number of supposedly solved matrices to the actually solved matrices (e.g., Mazar et al., 2008). This type of experimental deception is associated with less dishonest behavior among participants (Gerlach et al., 2019, table 2).
[2] In addition to being able to test the duration of the behavior, we used a *repeated* die-roll task to obtain more statistical precision to identify a small effect.

Study 1

Method

Participants and Sample Size Estimation

Two hundred and four adult participants were recruited using the Sona subject pool at Duke University, following our preregistered sample size estimation (see *SM*). Following preregistered exclusion criteria, individuals who did not complete the whole study were excluded, resulting in 202 participants remaining for the analysis (mean age = 23.51, SD = 4.41, 57.92% female). All participants were compensated for completing the study (avg. $17.98/hr).

Materials and Procedure

Initially, participants were seated in individual booths to ensure privacy. Afterward, participants were given a typical consent form informing them about the tasks, risks and benefits and time required for the study, screening criteria, and treatment of their personal data and were asked if they wanted to continue. Participants who signed this initial consent form were randomly assigned, via computer-generated randomization, to one of three conditions: No-Deception, Deception-Allowed, or Control. This manipulation was operationalized as a second form that participants were asked to sign, similar to the consent form.

Participants allocated to the no-deception condition saw a form indicating that the laboratory *prohibits the use of deception in all experimental protocols taking place in the facility*. Participants allocated to the deception-allowed condition saw a form indicating that the laboratory *allows the use of deception in all experimental protocols taking place in the facility*. Finally, participants allocated to the control condition did not receive any information regarding the laboratory's policy on deception and only saw the initial consent form.

The primary outcome measure was cheating quantified in a 20-shot die-roll task, where 1 pip = 10¢, 2 pips = 20¢, 3 pips = 30¢, 4 pips = 40¢, 5 pips = 50¢, and 6 pips = 60¢, in each round.[1] Immediately thereafter, participants answered a set of questionnaires assessing demographics, general honesty, and personality measures (see *SM*) before being thoroughly debriefed.

Results

Manipulation check

Participants completed the manipulation check, which asked them to show that they understood how earnings from the die rolls would be calculated. All participants correctly answered the manipulation check question and correctly estimated the hypothetical earnings of die rolls.

Die rolls

First, we analyzed dishonesty in the 20-shot die-roll task across conditions. A Kolmogorov–Smirnov test indicated that the distribution of die rolls across conditions was significantly different from a uniform distribution ($D = 0.374$, $p < .001$). The results indicated that participants on average claimed $7.68 (SD = 1.30), which is significantly higher than the expected claim of $7.00 (i.e., the sum of Ø3.5¢ claims in 20 die rolls using a fair die; one sample *t test*; $t(201) = 7.37$, 95% CI [0.37, 0.67], $d = 0.52$, $p < .001$), indicating that at least some participants across conditions used the opportunity to cheat in the game. Interestingly,

six participants reported rolling six all twenty times, maximizing the monetary outcome in the experiment, thus earning the maximum payoff of $12.

Next, we analyzed cheating between the three experimental conditions. Here, we used a one-way analysis of variance (ANOVA) to explore differences in claims between the three conditions (no-deception vs. deception-allowed vs. control). We observed no significant difference in claims between the three conditions ($F(2, 199) = 0.41$, $p = .666$; $\eta^2 = 0.00$).

To further corroborate this result, we used equivalence testing to establish whether the null hypothesis (i.e., no difference in dishonesty between conditions) was in fact more likely to be true than the hypothesis that different policies on deception would result in different levels of dishonest behavior[2]. For the two-one-sided-test (TOSTs; α =.05) between the no-deception condition and the control condition, the equivalence test was significant ($t(131.1) = 1.96$, 90% CI [-0.56 0.16], $p = .026$). Similarly, the test between the deception-allowed condition and the control condition was significant ($t(133.14) = 2.70$, 90% CI [-0.44, 0.34], $p = .004$). Finally, the test between the no-deception condition and the deception-allowed condition was significant, $t(128.64) = 2.23$, $p = .014$, 90% CI [-0.52, 0.22]. Based on the combined results of the equivalence test and the null hypothesis test, we can conclude that the observed effect is not significantly different from zero and is statistically equivalent to zero. The average claims in the die-roll task per experimental condition per round (Figure 2) suggest that dishonesty in the repeated measures die-roll task did not deviate based on the manipulation of the deception condition.

Discussion

The results from Study 1 indicate that deceptive environments, in the form of experimental deception, did not affect downstream, dishonest behavior. These results provide initial causal evidence that deceptive environments neither intensify nor inhibit dishonest behavior. These results contrast the evidence observed in the intuition study as well the associational evidence in the meta-analysis.

To replicate and further extend our findings, in the next study, we kept the die-roll task and added the sender–receiver game. This addition allowed us to test whether the effect of experimental deception on dishonest behavior is observed in specific tasks only, as suggested above (see *Discussion* under Meta-analysis). Furthermore, to eliminate potential learning effects, where participants gain experience with the task (Kroher & Wolbring, 2015), Studies 2 and 3 used one-shot tasks to measure dishonesty. This approach also allowed us to directly compare the findings by Gerlach and colleagues (2019) and our meta-analysis, which analyzed one-shot situations only.

Moreover, because we observed relatively low rates of cheating in our laboratory study, we made cheating opportunities more salient to participants by altering the payoff scheme: reporting 6 pips would result in a payment of zero, so participants who actually observed a 6 would be potentially disappointed and thus more inclined to misreport their observation (Fischbacher & Föllmi-Heusi, 2013).

Additionally, to extend the findings to samples more versatile than students, we used online samples. This choice allowed us to increase the number of participants, making it more likely to detect small effects with sufficient power. The change from laboratory to online studies meant that we also had to modify the stimuli to make them more suitable for online studies. Thus, in Studies 2 and 3, we informed participants that it is the IRB policy (rather than the laboratory's policy) that allows (vs. forbids) deception. Finally, to ensure that participants paid attention to the stimuli, we added additional attention and manipulation checks.

[1] In this study, we additionally used a Dictator Game (DG) to measure generosity. We did not use it in Studies 2 and 3, as we opted there to focus on only dishonest behavior. The analyses (also for order effects) and results of this extra variable can be found in the *SM*. Deceptive policy had no effect on giving in the DG. The die-roll task was combined with the DG in random order (block-randomized), so approximately 50% of the participants first played the die-roll task, whereas the other 50% first played the DG. Participants were also asked to test-to-play both tasks in a few trials before the actual, incentivized task and answered manipulation check questions for both tasks.
[2] See Lakens et al. 2018 for a thorough explanation of this procedure, including a description of the use of 90% confidence intervals that include zero.

Study 2

Method

Participants and Sample Size Estimation

Nine hundred and twenty-four adult participants were recruited through the online platform Prolific Academic, according to our preregistered sample size estimation (see *SM*). Following preregistered exclusion criteria, individuals who did not complete the whole study or who failed the manipulation check question were excluded[1], resulting in 640 participants remaining for the analysis (mean age = 36.52, SD = 12.41, 62.7% female, 43.6% having at least a BA degree). All participants were compensated for completing the study (avg. £22.26/hr + bonus £1). The study lasted approximately 10 minutes.

Materials and Procedure

Initially, participants were given a typical consent form informing them about the tasks, risks and benefits and time required for the study, screening criteria, and treatment of their personal data and were asked if they wanted to continue. Participants who signed this initial consent form were randomly assigned, via computer-generated randomization, to one of three conditions: no-deception, deception-allowed, or control. The conditions were operationalized as a second form, similar to the consent form, where

participants were asked to indicate that they understood the content and wanted to proceed with the experiment. Participants were made aware that they could proceed only after 20 seconds.

Participants allocated to the no-deception condition were informed that the IRB *prohibits the use of deception in all experimental protocols* and *does not allow giving participants misleading or erroneous information about (elements of) the study conducted*. Participants allocated to the deception-allowed condition were informed that the IRB *allows the use of deception in all experimental protocols* and *allows giving participants misleading or erroneous information about (elements of) the study conducted*. Finally, participants allocated to the control condition did not receive any information regarding the laboratory's policy on deception and learned only that the IRB *has approved this experimental protocol*. Afterward, participants were asked to answer the attention and manipulation check questions to ensure that they had understood and carefully read the information provided.

Participants were then asked to roll an actual physical die at their convenience, where the primary outcome measure was cheating quantified in this one-shot die-roll task as follows: 1 pip = 10¢, 2 pips = 20¢, 3 pips = 30¢, 4 pips = 40¢, 5 pips = 50¢, but 6 pips = 0¢. In addition, participants played a sender–receiver game in which cheating was quantified as sending truthful or deceitful information to another player. Unbeknownst to participants, in this version of the task, our participants were always Player 1, and there was no other participant (Player 2). Additionally, to control for potential punishment or reputation fears, we informed our participants that Player 2 would never be told about the nature of the information sent. The order in which the two tasks were completed was counterbalanced.

Subsequently, participants answered a set of questionnaires concerning demographics, honesty, and personality measures (*see SM*) before being thoroughly debriefed.

Results

Manipulation check

Based on the condition to which participants were assigned, they answered a question asking 1) whether deception was allowed in the experiment, 2) whether deception was not allowed in the experiment, or 3) whether the IRB provided any information about deception being allowed in the experiment. In accordance with our preregistration, 108 participants did not pass the manipulation check and were consequently excluded from the subsequent analysis (see *SM*).

Die rolls

Initially, we analyzed dishonesty in the individual one-shot die-roll task across conditions. A one-sample Kolmogorov–Smirnov test indicated that reported rolls across conditions were significantly different from

a uniform distribution (D = 0.181, p <.001), which provided the first indication that individuals seized the opportunity to cheat in the task. Turning to earnings in the task, we found that participants reported a mean claim of 32.61¢ (SD = 13.92¢), which is significantly higher than the expected mean claim of 25¢ in the task (one sample *t test*; $t(639) = 13.83$, $d = 0.55$, p <.001). This result indicated that at least some participants inflated their earnings in the task by acting dishonestly.

We then estimated the percentage of potentially dishonest individuals in the task, as well as the percentage of individuals who acted dishonestly to maximize their payoff. As 5.13% of participants reported a payoff of 0, we estimated the upper limit of unconditionally honest participants to be 30.78% (i.e., 6 × 5.13%)[2]. Thus, 30.78% is an upper limit for the number of honest participants. Next, we calculated the percentage of individuals acting to maximize their payoff in the task as follows: 20.16% of individuals in the sample claimed a 5. Assuming that all participants who actually rolled a 5 would also report having rolled a 5, we estimated the maximal percentage of payoff maximizers to be 4.19% (i.e., (20.16% − 1/6) × 6/5))[3], indicating that few people in the sample actually acted as such. The distribution of reported claims between conditions is illustrated in Figure 3.

Next, we analyzed cheating between conditions to identify differences in individual dishonesty patterns between the three conditions (i.e., no-deception vs. deception-allowed vs. control). Here, a one-way analysis of variance (ANOVA) suggested no difference in cheating between conditions ($F(2, 637) = 0.73$, $p = 0.935$; $\eta_p^2 = 0.00$). Thus, these results suggest that irrespective of whether participants were made aware either that 1) deception was used in the experiment or 2) no deception was used in the experiment or were 3) simply not given any information regarding the deception policy of the study, such policies (or the lack thereof) did not affect participants' inclination to behave (dis)honestly in the die roll task.

To further corroborate this result, we used equivalence testing to establish whether the null hypothesis (i.e., no difference in dishonesty between conditions) was in fact more likely to be true than the hypothesis that different policies on deception would result in different levels of dishonest behavior. The equivalence test was significant for the two-one-sided-test (TOSTs; α =.05;) between the no-deception condition and the control ($t(389.82) = -3.36$, 90% CI [-2.00, 2.60], p <.001) as well as for the test between the deception-allowed condition and the control ($t(375.69) = -1.97$, 90% CI [-0.21, 4.41], p =.025). Hence, based on the combined results of the equivalence test and the null hypothesis test, we conclude that the observed effects of condition on dishonesty were not significantly different from zero and instead were statistically equivalent to zero.

Sender-Receiver game

Across conditions, 40.8% of participants in the sender-receiver game acted dishonestly by sending a deceitful message to the other player and thus gaining a higher outcome in the task.

Between conditions, we observed this behavior in 36.1% of participants in the no-deception condition, in 42.9% of participants in the deception-allowed condition and in 44.6% of participants in the control condition. On the condition level, a X^2-test between the no-deception and control condition revealed no significant difference between the two conditions ($X^2(1) = 2.82$, $p = .093$); the same was found between the no-deception and deception-allowed condition ($X^2(1) = 1.96$, $p = .162$) and between the deception-allowed and control condition ($X^2(1) = 0.05$, $p = .821$). Furthermore, predicting sender-receiver outcomes in a logistic model with the control condition as the intercept yielded no significant predictive power of the no-deception condition ($b = -0.35$, 95% CI [-0.75, 0.04], OR = 0.70, $p = .076$) or the deception-allowed condition ($b = -0.07$, 95% CI [-0.47, 0.33], OR = 0.94, $p = .740$).

Discussion

Consistent with our laboratory study findings, individuals do not behave more dishonestly when in deceptive environments, whether cheating victimizes the experimenter (or the lab/institution) or when it affects another participant. Importantly, participants recognized the environment they were in (deceptive vs. nondeceptive), but this recognition did not affect their cheating behavior.

A crucial null finding was that deceptive environments did not cause less dishonest behavior in the sender-receiver game, as participants feared they might be exposed as liars and thus showed less dishonest behavior. Might this finding be due to the particularities of the sender-receiver game, i.e., participants not believing that their partner interaction truly exists? Additionally, as argued earlier, previous research has shown that being detected as a cheater decreases dishonest behavior (Thielmann & Hilbig, 2018); therefore, people might feel more observed in environments with deception and consequently cheat less. To directly test whether feeling observed in deceptive environments can affect behavior, in the following study, we added two conditions, observed vs. unobserved (explained below).

[1] The number of excluded participants is fairly balanced across conditions, therefore minimizing chances that our effects are due to systematic selection. Specifically, 7.58% of participants were excluded in the No-Deception condition, 12.9% in the Deception-Allowed condition and 16.1% in the control condition.
[2] Since we can calculate the base rate of unconditionally honest individuals (i.e., 5.13%) based on the participants who report a 6, we can assume that this base rate holds across every outcome (i.e., 1-6) and in this way can thus estimate the total number of unconditionally honest individuals, as suggested by Fischbacher & Föllmi-Heusi, 2013.
[3] The multiplication with 6/5 is essential to include to account for the “payoff maximizers” who actually rolled a 5.

Study 3

Method

Participants and Sample Size Estimation

One thousand and four adult participants were recruited through the online platform Prolific Academic, according to our preregistered sample size estimation (see *SM*). Following preregistered exclusion criteria, individuals who did not complete the whole study or who failed to correctly answer the manipulation check question were excluded, resulting in 832 participants remaining for the analysis (mean age = 34.49, SD = 11.65, 64.18% female, 43% having at least BA degree). All participants were compensated for completing the study (avg. £18.95/hr + bonus £0.50). The study lasted approximately 8 minutes.

Materials and Procedure

The design of the study (primary outcome measure, set of questionnaires, and debriefing) was identical to Study 2, with two alterations: 1) we did not include the sender-receiver game, and 2) we included two additional conditions (observed vs. unobserved), as described below.

Initially, participants saw a typical consent form informing them about the tasks, risks and benefits and time required for the study, screening criteria, and treatment of their personal data and were asked if they wanted to continue. Participants who signed this initial consent form were randomly assigned to one of 3 × 2 conditions: no-deception vs. deception-allowed vs. control treatment × observed vs. unobserved. Similar to Study 2, the manipulation of deception (no-deception vs. deception-allowed vs. control) was operationalized as a consent form. The manipulation of observation at the individual level (unobserved vs. observed) was operationalized as two different ways of rolling the die. Participants either rolled an actual die at their convenience (unobserved) or rolled a die that we programmed within the online survey program (observed). In the unobserved condition, we thus measured dishonesty on the aggregate level only, as in Studies 1 and 2. That is, the unobserved condition only yielded the distribution of reported pips. In contrast, in the observed condition, we measured dishonesty on the individual level. That is, individuals could be identified as cheaters by comparing the pip(s) they saw against their actual reporting.

Results

Manipulation check

Based on the condition to which participants were assigned, they answered a question asking 1) whether deception was allowed in the experiment, 2) whether deception was not allowed in the experiment or 3) whether the IRB provided any information about deception being allowed in the experiment. Passing the

manipulation check was interpreted as a successful implementation of the respective stimuli. All participants passed their respective manipulation check.

Die rolls

First, we analyzed dishonesty in the die roll task across conditions. A one-sample Kolmogorov–Smirnov test indicated that reported rolls across conditions were significantly different from a uniform distribution (D = 0.129, p <.001). This result provided an initial indication that at least some participants took the opportunity to cheat in the task.

Next, we analyzed the reported claims in the die roll task across conditions. Here, we found that participants reported a mean claim of 27.90¢ (SD = 16.51¢), which is significantly higher than the expected mean claim of 25.00¢ in the task (one sample *t test*; t(831) = 5.06, d = 0.18, 95% CI [0.11, 0.24], p <.001), thus indicating that at least some participants inflated their earnings in the task by acting dishonestly.

Using simple probability statistics, we then estimated the percentage of potentially dishonest individuals in the task, as well as the percentage of individuals who acted dishonestly to maximize their payoff. Here, we assumed that if unconditionally honest individuals roll a uniform distribution of numbers, then it is reasonable to take the number of people reporting a payoff of 0 to estimate the percentage of honest people in each number reported (Fischbacher & Föllmi-Heusi, 2013). As 12.86% reported a payoff of 0, we estimated the percentage of unconditionally honest participants to be 77.16% (i.e., 6 x 12.86%). Thus, 77.16% is an upper limit for the number of honest participants. Next, we calculated the percentage of individuals acting to maximize their payoff in the task. This was calculated as follows: 18.75% of individuals in the sample reported having rolled a 5. Assuming that all participants who actually rolled a 5 would also report having rolled a 5, we estimated the maximal percentage of payoff maximizers to be at 2.5% (i.e., (18.75% − 1/6) × 6/5)), indicating that few individuals in the sample actually acted as such. The distribution of the reported claims between conditions is illustrated in Figure 4.

Next, before analyzing differences between the six conditions, we analyzed differences in claims between the unobserved vs. observed conditions. Here, we found that participants in the unobserved condition reported a significantly higher outcome than participants in the observed condition (t(184.08) = -2.58, d = -0.25, 95% CI [-0.44, -0.06], p =.011). Importantly, our results showed that the mean claim in the observed condition was not significantly different from the expected claim (25.00¢)[1] in the task (t(134) = -0.38, d = -0.03, 95% CI [-0.20, 0.14], p =.707), whereas the mean claim in the Unobserved condition was significantly higher than the expected claim (t(696) = 5.77), d = 0.22, 95% CI [0.14, 0.29], p <.001.). Focusing on the three conditions in which we could actually observe what participants rolled, we tested whether the reported results were significantly different from the actual number that participants rolled in

their first die roll in the experiment across conditions. We found no significant difference between what participants rolled in their first die roll and what they subsequently claimed to have rolled (t(772.72) = 0.73, d = 0.04, 95% CI [-0.07, 0.16], p =.465).

Next, we analyzed cheating between the six conditions to identify differences in dishonesty patterns based on 1) whether participants were presented with a no-deception IRB policy (vs. deception-allowed vs. control) and 2) whether participants rolled an observed (vs. unobserved) die.

Here, a linear model (OLS) predicting claim by condition type and moderated by whether the condition was observed vs. unobserved suggested small significant differences in cheating between conditions (F(5, 826) = 9.17, p <.001; η_p^2 = 0.05). Setting the control conditions as our intercept in this model (β = 32.87, 95% CI [30.08, 35.65], t(826) = 23.16, p <.001), our results showed that die-roll claims were significantly lower in the no-deception conditions (β = -7.27, 95% CI [-11.07, -3.47], t(826) = -3.76, p <.001; *Std.* β = -0.39, 95% CI [-0.55, -0.22]), whereas we found no such effect for the deception-allowed conditions (β = 0.81, 95% CI [-4.67, 3.04], t(826) = -0.41, p =.679; *Std.* β = -0.23, 95% CI [-0.39, -0.07]) or for the main effect of whether participants were observed or unobserved (β = -2.94, 95% CI [-6.84, 0.96], t(826) = -1.48, p =.139; *Std.* β = -0.09, 95% CI [-0.21, 0.03]). Importantly, however, we found a statistically significant and negative interaction effect between the deception-allowed conditions and whether participants were observed (vs. unobserved) (β = -6.04, 95% CI [-11.45, -0.62], t(826) = -2.19, p =.029; *Std.* β = -0.18, 95% CI [-0.35, -0.02]). No such interaction effect was found between the no-deception conditions and whether individuals were observed or unobserved (β = 1.79, 95% CI [-3.63, 7.20], t(826) = 0.65, p =.517; *Std.* β = 0.05, 95% CI [-0.11, 0.22]). These results are shown in Table 2.

Table 2. | Predicting claims per condition moderated by observed/unobserved

	Die-roll Claims		
Predictors	*Estimates*	*95% CI*	*p*
Intercept (Control)	32.87	[30.08, 35.65]	**<.001**
Deception-Allowed	-0.81	[-4.67, 3.04]	.679
No-Deception	-7.27	[-11.07, -3.47]	**<.001**
Observed	-2.94	[-6.84, 0.96]	.139
Deception-Allowed x Observed	-6.04	[-11.45, -0.62]	**.029**
No-Deception x Observed	1.79	[-3.63, 7.20]	.517
Observations	832		
R^2 / R^2 adjusted	0.053 / 0.047		

Discussion

Study 3 reveals an intriguing result: being informed about the lack of deception in the experiment seems to impact participants' behavior, as it leads to lower levels of dishonest behavior than in the control. This finding is to some degree in line with prior research conducted by Ortmann and Hertwig (Ortmann & Hertwig, 2002; Hertwig & Ortmann, 2008a, 2008b).

Additionally, while dishonest behavior occurred across all conditions, our results show that when participants were in a deceptive environment *and* were made aware that their performance in the die-roll task was being observed, such participants cheated significantly less—even comparable to the conditions where no deception was allowed, which to some degree aligns with previous findings on the effects of cues of anonymity or being observed (i.e., visibility) on cheating behavior (Zhong et al., 2010; Ernest-Jones et al., 2011; Nettle et al., 2012; Pfattheicher & Keller, 2015; Schild et al., 2019).

[1] The expected claim in a single die-roll was calculated as follows: (0¢+10¢+20¢+30¢+40¢+50¢)/6 = 25¢.

General Discussion

Human societies are not immune to deception. However, to what degree does individual moral behavior depend on the institutional setting of an organization or a culture allowing or prohibiting deception? Contrary to the widespread intuition that deceptive environments necessarily breed dishonest behavior, the present studies provide novel evidence that individuals making moral decisions within deceptive environments do not *always* behave more dishonestly than they would in nondeceptive environments.

More specifically, in a prediction survey, we confirmed the widespread intuition that deceptive environments should increase individual dishonest behavior, both toward the experimenter and another participant. Subsequently, in a meta-analytical study, where we test whether deceptive laboratory policies affect participants' honesty, we found that experiencing experimental deception at the laboratory policy level actually results in less dishonest participants. Surprised by these contradicting and counterintuitive findings, in Studies 1-3, we examined how different types of environmental deception affect the dishonest behavior of participants.

In Study 1, we conducted in-person laboratory research and found that deception operationalized at the level of the laboratory policy does not affect repeated, downstream honest behavior. In Study 2, we investigated in online settings how deception operationalized at the level of the IRB influences morality. We introduced a one-shot die-roll task to avoid learning effects and one additional measure for cheating (sender-receiver game), where the victim of dishonesty was not the experimenter (or the lab/institution) but another participant, to gauge the effects of directed dishonesty when social preferences are present. We replicated the main pattern of findings from the previous study. In Study 3, we introduced an additional observed (vs. unobserved) treatment to test whether anonymity and cues of being observed in deceptive environments affect honest behavior. Here, we found evidence that informing participants about the nondeceptive policy of an environment reduces subsequent dishonesty. Moreover, we also found some evidence that when participants were in a deceptive environment and were aware of being observed (vs. not), their dishonest behavior decreased.

Contributions and implications

Prior research on moral psychology has singled out deception as particularly harmful for moral behavior (Bok, 1978; Boles et al., 2000; Croson et al., 2003; Schweitzer et al., 2006; Schweitzer & Gibson, 2008). This work, however, has conflated deception with direct lying and self-serving intentions and has only looked at deception at the interpersonal level. We find that deceptive environments do not always directly negatively affect downstream moral behavior. Our results show that the relationship between deception and dishonesty is much more complicated than previous interpretations have suggested. Being and acting in a deceptive environment does not necessarily breed dishonesty.

Our research contributes to the deception and moral behavior literature in numerous ways. First, we demonstrate and echo the “the importance of studying a broader range of deceptive behaviors” (Levine & Schweitzer, 2015). Deception is pervasive, yet we know surprisingly little about its consequences on behavior in general and even less about the effect of potentially deceptive environments on downstream, individual moral behavior. While research assumes that deception is harmful (Boles et al., 2000; Croson et al., 2003; Schweitzer et al., 2006; Schweitzer & Gibson, 2008), here we show that this assumption might not always hold.

Second, we find some evidence that simply being informed about the lack of deception within an environment can have positive implications for downstream honest behavior. This might possibility leads to a significant methodological implication, i.e., researchers need to be cautious regarding the type of information they share with participants. This finding should be interpreted with caution since it is based on an interaction effect.

Third, our research has another important methodological implication. Academics have debated for decades about the use of deception in experimental settings (Mills, 1976; Davis & Holt, 1993; Friedman et al., 1994; Bonetti, 1998; Hey, 1998; Jamison et al., 2008; Hertwig & Ortmann, 2008b; Krawczyk, 2019). The use of experimental deception is controversial and associated with different schools of thought. On the one hand, some researchers—mainly psychologists—suggest that the use of deception, confederates, and cover stories is defendable as a last resort to not bias participants’ responses and that there is a potential overall benefit in data validity and experimental control (Bonetti, 1998; Hilbig et al., 2021). Deception is justified “by the study’s significant prospective scientific, educational, or applied value and that effective nondeceptive alternative procedures are not feasible” (American Psychological Association, 2002). On the other hand, other researchers—mainly from the field of experimental economics—argue that the use of deception should be avoided, as it violates moral academic conduct (Hertwig & Ortmann, 2008b), affects “participants’ expectations, suspicions, and future behavior” (Jamison et al., 2008) and pollutes the participant pool (Friedman et al., 1994). Our results show that deception might harm internal validity if participants are suspicious of being monitored. We thus urge researchers to take this possibility into account when designing studies. Although, in principle, we believe that the best policy is no deception, we could value the insights that a different approach might offer, realize that deception may be unavoidable (Hertwig & Ortmann, 2008b), and call for unambiguous and clear guidelines (Hersch, 2015; Krawczyk, 2019) that will make the trade-off of using deception easier to determine.

Fourth, we found that lay people’s intuitions about moral behavior in deceptive environments are in direct contrast to empirical findings. This mismatch between predictions and outcomes emphasizes the importance of experiments to understand human behavior. Here, we want to echo Hertwig & Ortmann (2008b) that “the evaluation of our methodological standards and policy should be *evidence-based.*”

Finally, and crucially, we find that in deceptive environments, it is beneficial to introduce cues of visibility, as doing so beneficially impacts moral behavior, which does not seem to be necessary in nondeceptive environments. These findings have central policy implications and should lead to fruitful directions

for policy-related research. A cost (visibility cues)–benefit (decrease dishonesty) analysis might be beneficial and worth exploring in future research.

Limitations and future directions

We encourage future research to further examine the underlying mechanisms and boundary conditions of the effects of deceptive environments on moral behavior. For example, researchers should directly test whether there is a difference between deceptive individuals vs. deceptive environments on downstream moral behavior. In that case, it might be that interpersonal deception negatively affects moral behavior. Studies targeting this particular distinction would be helpful.

Relatedly, different types of deception may result in different types of dishonest behavior. In this article, we focus on one specific type of deception (i.e., at the laboratory/IRB level) and dishonest behavior—i.e., cheating and lying. However, our results might not be generalizable to all types of deception or dishonest behavior. For example, future work could study alternating levels of deception within environments and its effects on moral behavior to explore the dynamic and relative nature of deception and its relation to morality.

In terms of moral behavior in general, anonymity and visibility cues may influence the decision to be honest or not. As these factors are crucial for organizations, we call for further research and replication studies, following open science principles. Although our meta-analytical and experimental results offer initial insights into the causal effect of deception on moral behavior, one should be cautious in extending these findings to the real world. Thus, we call for real-world investigations on deception and morality that involve cross-cultural and diverse sampling methods combined with experiments within organizations. The latter entail creative ways of measuring (un)ethical behavior and resolving potential selection effects.

Conclusion

Transparency International (2021) considers honesty essential to interpersonal relationships and organizations. (Dis)honesty is a social phenomenon and should be studied within social contexts. To our knowledge, this is the first article to study the effects of organizational deception, i.e., if a whole environment is deceptive, on dishonesty. Through a survey, a meta-analysis, and three experiments, we show that there is a common assumption that deception trumps honesty, and we challenge this assumption. We provide evidence that the relationship between deceptive environments and morality is much more complicated than assumed and call for further debate and research on the topic.

References

Ambraseys, N. & Bilham, R. (2011). Corruption kills. *Nature*, *469*(7329), 153-155.

American Psychological Association. 2002. Ethical Principles of Psychologists and Code of Conduct. Retrieved from http://www.apa.org/ethics/.

Ariely, D. (2012). *The (honest) truth about dishonesty: How we lie to everyone - especially ourselves*. Harper Collins Publisher.

Ariely, D., Garcia-Rada, X., Gödker, K., Hornuf, L., & Mann, H. (2019). The impact of two different economic systems on dishonesty. *European Journal of Political Economy*, *59*, 179-195.

Ashton, M. C., Lee, K., & De Vries, R. E. (2014). The HEXACO Honesty-Humility, Agreeableness, and Emotionality factors: A review of research and theory. *Personality and Social Psychology Review*, *18*(2), 139-152.

Ayal, S., Gino, F., Barkan, R., & Ariely, D. (2015). Three principles to REVISE people's unethical behavior. *Perspectives on Psychological Science*, *10*(6), 738-741.

Banerjee, R. (2016). Corruption, norm violation and decay in social capital. *Journal of Public Economics*, *137*, 14-27.

Baumard, N., & Sperber, D. (2010, Jun). Weird people, yes, but also weird experiments. *Behavioral and Brain Sciences, 33*(2-3), 84-85.

Bellé, N., & Cantarelli, P. (2017). What causes unethical behavior? A meta-analysis to set an agenda for public administration research. *Public Administration Review*, *77*(3), 327-339.

Bicchieri, C. (2006). *The grammar of society: The nature and dynamics of social norms*. United Kingdom: Cambridge University Press.

Bok, S. (1978). Lying: Moral choices in public and private life. New York, NY: Pantheon.

Boles, T. L., Croson, R. T., & Murnighan, J. K. (2000). Deception and retribution in repeated ultimatum bargaining. *Organizational Behavior and Human Decision Processes, 83*(2), 235–259.

Bonetti, S. (1998). Experimental economics and deception. *Journal of Economic Psychology*, *19*(3), 377-395.

Bucciol, A. & Piovesan, M. (2011). Luck or cheating? A field experiment on honesty with children. *Journal of Economic Psychology*, *32*(1), 73–78.

Butler, J., Giuliano, P., & Guiso, L. (2016). Trust and cheating. *The Economic Journal*, *126*(595), 1703-1738.

Cebula, R. J. & Feige, E. L. (2012). America's unreported economy: Measuring the size, growth and determinants of income tax evasion in the U.S. *Crime, Law and Social Change*, *57*(3), 265–285.

Christensen, L. (1977). The negative subject: Myth, reality, or a prior experimental experience effect?. *Journal of Personality and Social Psychology*, *35*(6), 392.

Cialdini, R. B., Reno, R. R., & 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,* 1015– 1026.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*, 2nd ed. Hillsdale, NJ: Erlbaum.

Cohn, A., Fehr, E., & Maréchal, M. A. (2014). Business culture and dishonesty in the banking industry. *Nature*, *516*(7529), 86-89.

Cropanzano, R. & Mitchell, M. S. (2005). Social exchange theory: An Interdisciplinary review. *Journal of Management, 31,* 874–900. http:// dx.doi.org/10.1177/0149206305279602.

Croson, R. & Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic Literature, 47,* 448–474.

Croson, R., Boles, T., & Murnighan, J. K. (2003). Cheap talk in bargaining experiments: Lying and threats in ultimatum games. *Journal of Economic Behavior & Organization*, *51*(2), 143–159.

Davis, D. D. & Holt, C. A. (1993). Experimental Economics Princeton University Press. *Princeton, New Jersey*.

Dieckmann, A., Grimm, V., Unfried, M., Utikal, V., & Valmasoni, L. (2016). On trust in honesty and volunteering among Europeans: Cross-country evidence on perceptions and behavior. *European Economic Review*, *90*, 225-253.

Ernest-Jones, M., Nettle, D., & Bateson, M. (2011). Effects of eye images on everyday cooperative behavior: a field experiment. *Evolution and Human Behavior*, *32*(3), 172-178.

Falk, A. & Szech, N. (2013). Morals and markets. *Science*, *340*(6133), 707-711.

Fischbacher, U. & Föllmi-Heusi, F. (2013). Lies in disguise—an experimental study on cheating. *Journal of the European Economic Association*, *11*(3), 525-547.

Friedman, S., Friedman, D., & Sunder, S. (1994). *Experimental methods: A primer for economists*. Cambridge university press.

Frollová, N., Vranka, M., & Houdek, P. (2021). A qualitative study of perception of a dishonesty experiment. *Journal of Economic Methodology*, 1-17.

Gächter, S. & Schulz, J. F. (2016). Intrinsic honesty and the prevalence of rule violations across societies. *Nature*, *531*(7595), 496-499.

Gee, J. & Button, M. (2019). *The Financial Cost of Fraud 2019: The Latest Data from Around the World*. The Financial Cost of Fraud, Issue. http://www.crowe.ie/wp-content/uploads/2019/08/The-Financial-Cost-of-Fraud-2019.pdf.

Gerlach, P (2017): The Games Economists Play: Why Economics Students Behave More Selfishly than Other Students. *PLOS ONE*, 12, e0183814.

Gerlach, P., Teodorescu, K., & Hertwig, R. (2019). The truth about lies: A meta-analysis on dishonest behavior. *Psychological Bulletin*, *145*(1), 1.

Gino, F. & Shea, C. (2012). Deception in negotiations. *The Oxford handbook of economic conflict resolution*, 47-60. New York: Oxford University Press.

Gneezy, U. (2005). Deception: The role of consequences. *American Economic Review*, *95*(1), 384-394.

Gneezy, U., Kajackaite, A., & Sobel, J. (2018). Lying aversion and the size of the lie. *The American Economic Review, 108,* 419–453. http://dx.doi.org/10.1257/aer.20161553.

Hersch, G. (2015). Experimental economics' inconsistent ban on deception. *Studies in History and Philosophy of Science Part A*, *52*, 13-19.

Hertwig, R. & Ortmann, A. (2008a). Deception in experiments: Revisiting the arguments in its defense. *Ethics & Behavior*, *18*(1), 59-92.

Hertwig, R. & Ortmann, A. (2008b). Deception in social psychological experiments: Two misconceptions and a research agenda. *Social Psychology Quarterly*, *71*(3), 222-227.

Hey, J. D. (1998). Experimental economics and deception: A comment. *Journal of Economic Psychology*, *19*(3), 397-401.

Hilbig, B. E., & Hessler, C. M. (2013). What lies beneath: How the distance between truth and lie drives dishonesty. *Journal of Experimental Social Psychology*, *49*(2), 263-266.

Hilbig, B. E., Thielmann, I., & Böhm, R. (2021). Bending our ethics code: Avoidable deception and its justification in psychological research. *European Psychologist.* Advance online publication. http://dx.doi.org/10.1027/1016-9040/a000431

Hoogeveen, S., Sarafoglou, A., & Wagenmakers, E. J. (2020). Laypeople Can Predict Which Social-Science Studies Will Be Replicated Successfully. *Advances in Methods and Practices in Psychological Science*, *3*(3), 267-285.

Hugh-Jones, D. (2016). Honesty, beliefs about honesty, and economic growth in 15 countries. *Journal of Economic Behavior & Organization*, *127*, 99-114.

Jamison, J., Karlan, D., & Schechter, L. (2008). To deceive or not to deceive: The effect of deception on behavior in future laboratory experiments. *Journal of Economic Behavior & Organization*, *68*(3-4), 477-488.

Jamison, J., Karlan, D., & Schechter, L. (2008). To deceive or not to deceive: The effect of deception on behavior in future laboratory experiments. *Journal of Economic Behavior & Organization*, *68*(3-4), 477-488.

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1986). Fairness and the assumptions of economics. *Journal of Business*, S285-S300.

Keizer, K., Lindenberg, S., & Steg, L. (2008). The spreading of disorder. *Science, 322,* 1681–1685.

Kimmel, A. J. (1996). *Ethical issues in behavioral research: A survey*. Cambridge, MA: Blackwell.

Kirchler, E., Hoelzl, E., & Wahl, I. (2008). Enforced versus voluntary tax compliance: The “slippery slope” framework. *Journal of Economic Psychology*, *29*(2), 210-225.

Köbis, N. C., Iragorri-Carter, D., & Starke, C. (2018). A social psychological view on the social norms of corruption. In *Corruption and norms* (pp. 31-52). Palgrave Macmillan, Cham.

Krátký, J., McGraw, J. J., Xygalatas, D., Mitkidis, P., & Reddish, P. (2016). It depends who is watching you: 3-D agent cues increase fairness. *PLoS One*, *11*(2), e0148845.

Krawczyk, M. (2019). What should be regarded as deception in experimental economics? Evidence from a survey of researchers and subjects. *Journal of Behavioral and Experimental Economics*, *79*, 110-118.

Kroher, M. & Wolbring, T. (2015). Social control, social learning, and cheating: Evidence from lab and online experiments on dishonesty. *Social Science Research*, *53*, 311-324.

Lakens, D., Scheel, A. M., & Isager, P. M. (2018). Equivalence Testing for Psychological Research: A Tutorial. *Advances in Methods and Practices in Psychological Science*, *1*(2), 259-269.

Larney, A., Rotella, A., & Barclay, P. (2019). Stake size effects in ultimatum game and dictator game offers: A meta-analysis. *Organizational Behavior and Human Decision Processes*, *151*, 61-72.

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

Levine, E. E. & Schweitzer, M. E. (2015). Prosocial lies: When deception breeds trust. *Organizational Behavior and Human Decision Processes*, *126*, 88-106.

Lewicki, R. J. (1983). Lying and deception: A behavioral model. In M.H. Bazerman & R.J. Lewicki (Eds.), *Negotiating in organizations* (pp. 68–90). Beverly Hills, CA: Sage.

Mann, H., Garcia-Rada, X., Hornuf, L., Tafurt, J., & Ariely, D. (2016). Cut from the same cloth: Similarly dishonest individuals across countries. *Journal of Cross-Cultural Psychology*, *47*(6), 858-874.

Mann, H., Garcia-Rada, X., Houser, D., & Ariely, D. (2014). Everybody else is doing it: Exploring social transmission of lying behavior. *PloS One*, *9*(10), e109591.

Mazar, N. & Aggarwal, P. (2011). Greasing the palm: Can collectivism promote bribery?. *Psychological Science*, *22*(7), 843-848.

Mazar, N., Amir, O., & Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research*, *45*(6), 633-644.

Mills, J. (1976). A procedure for explaining experiments involving deception. *Personality and Social Psychology Bulletin*, *2*(1), 3-13.

Mitkidis, P., Ayal, S., Shalvi, S., Heimann, K., Levy, G., Kyselo, M., ... & Roepstorff, A. (2017). The effects of extreme rituals on moral behavior: The performers-observers gap hypothesis. *Journal of Economic Psychology*, *59*, 1-7.

Nichols, A., Garcia-Rada, X., Chituc, V., Mann, H., Campbell, T., Mitkidis, P., & Ariely D. (in review). The Moral Degradation of Bribes.

Muthukrishna, M., Francois, P., Pourahmadi, S., & Henrich, J. (2017). Corrupting cooperation and how anti-corruption strategies may backfire. *Nature Human Behavior*, *1*(7), 1-5.

Nettle, D., Nott, K., & Bateson, M. (2012). 'Cycle thieves, we are watching you': Impact of a simple signage intervention against bicycle theft. *PloS One*, *7*(12), e51738.

Olsen, A. L., Hjorth, F., Harmon, N., & Barfort, S. (2019). Behavioral dishonesty in the public sector. *Journal of Public Administration Research and Theory*, *29*(4), 572-590.

Ortmann, A. & Hertwig, R. (2002). The costs of deception: Evidence from psychology. *Experimental Economics*, *5*(2), 111-131.

Pascual-Ezama, D., Fosgaard, T. R., Cardenas, J. C., Kujal, P., Veszteg, R., de Liaño, B. G. G., ... & Branas-Garza, P. (2015). Context-dependent cheating: Experimental evidence from 16 countries. *Journal of Economic Behavior & Organization*, *116*, 379-386.

Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology*, *70*, 153-163.

Piff, P. K., Stancato, D. M., Côté, S., Mendoza-Denton, R., & Keltner, D. (2012). Higher social class predicts increased unethical behavior. *Proceedings of the National Academy of Sciences*, *109*(11), 4086-4091.

Rahwan, Zoe, Erez Yoeli, & Barbara Fasolo. "Heterogeneity in banker culture and its influence on dishonesty." *Nature, 575*.7782 (2019): 345-349.

Rees, M. R., Tenbrunsel, A. E., & Bazerman, M. H. (2019). Bounded ethicality and ethical fading in negotiations: Understanding unintended unethical behavior. *Academy of Management Perspectives*, *33*(1), 26-42.

Rigby, R. A., & Stasinopoulos, D.M. (2005). Generalized additive models for location, scale and shape (with discussion), *Applied Statistics*, *54*, part 3, 507-554.

Robert, I. & Arnab, M. (2013). Is dishonesty contagious?. *Economic Inquiry*, *51*(1), 722-734.

Schild, C., Heck, D. W., Ścigała, K. A., & Zettler, I. (2019). Revisiting REVISE:(Re) Testing unique and combined effects of REminding, VIsibility, and SElf-engagement manipulations on cheating behavior. *Journal of Economic Psychology*, *75*, 102161.

Schweitzer, M. E. & Gibson, D. E. (2008). Fairness, feelings, and ethical decision-making: Consequences of violating community standards of fairness. *Journal of Business Ethics*, *77*(3), 287-301.

Schweitzer, M. E. & Hsee, C. K. (2002). Stretching the truth: Elastic justification and motivated communication of uncertain information. *Journal of Risk and Uncertainty*, *25*, 185–201.

Schweitzer, M. E., Hershey, J. C., & Bradlow, E. T. (2006). Promises and lies: Restoring violated trust. *Organizational Behavior and Human Decision Processes, 101*(1), 1–19.

Shalvi, S., Eldar, O., & Bereby-Meyer, Y. (2012). Honesty requires time (and lack of justifications). *Psychological science*, *23*(10), 1264-1270.

Stansbury, J. (2009). Reasoned moral agreement: Applying discourse ethics within organizations. *Business Ethics Quarterly*, 33-56.

Stricker, L. J., Messick, S., & Jackson, D. N. (1967). Suspicion of deception: Implications for conformity research. *Journal of Personality and Social Psychology*, *5*(4), 379.

Thielman, I & Hilbig, BE (2019). No gain without pain: The psychological costs of dishonesty. *Journal of Economic Psychology. 71*, 126-137.

Thielmann, I & Hilbig, BE (2018). Daring dishonesty: On the role of sanctions for (un)ethical behavior. *Journal of Experimental Social Psychology*, *79*, 71-77

Transparency International. (2019). *Ethics and Integrity.* Retrieved 25th of February from: https://www.transparency.org/en/the-organisation/ethics-integrity.

van Baal S., Chennells M., Walasek L., Mitkidis P., & Michael J. (in review). The Social Scaffolding of Corruption.

Warren, D. E., & Schweitzer, M. E. (2021). When weak sanctioning systems work: Evidence from auto insurance industry fraud investigations. *Organizational Behavior and Human Decision Processes*, *166*, 68-83.

Weisel, O., & Shalvi, S. (2015). The collaborative roots of corruption. *Proceedings of the National Academy of Sciences*, *112*(34), 10651-10656.

Zhong, C.-B., Bohns, V. K., & Gino, F. (2010). Good lamps are the best police: Darkness increases dishonesty and self-interested behavior. *Psychological Science, 21*(3), 311–314.

Declarations

Acknowledgments

We thank Arnault-Quentin Vermillet for assistance in the meta-analysis and Elif Tunaboylu

and Daisy Fang for assisting with data collection for the laboratory study.

Author Contributions

P. Mitkidis developed the studies' concept, proposed the initial designs, interpreted the data, wrote the initial draft, offered suggestions on the data analysis plan and meta-analysis, and managed the OSF repository of the project. P. Gerlach made critical revisions to the design of Studies 2 and 3, wrote the R code, conducted the meta-analysis, and made critical revisions to the manuscript. A. Nichols made critical revisions to the first study design and managed the first study. C.T. Elbæk wrote the analysis code, conducted all studies' analyses, and managed the OSF repository of the project. S. Perkovic made critical revisions to all study designs, offered suggestions on the data analysis plan, and made critical revisions to the manuscript. D. Ariely made critical revisions to the first study design and to the manuscript. All authors have approved the final version of the manuscript for submission. The authors have no conflict of interest.

Figures

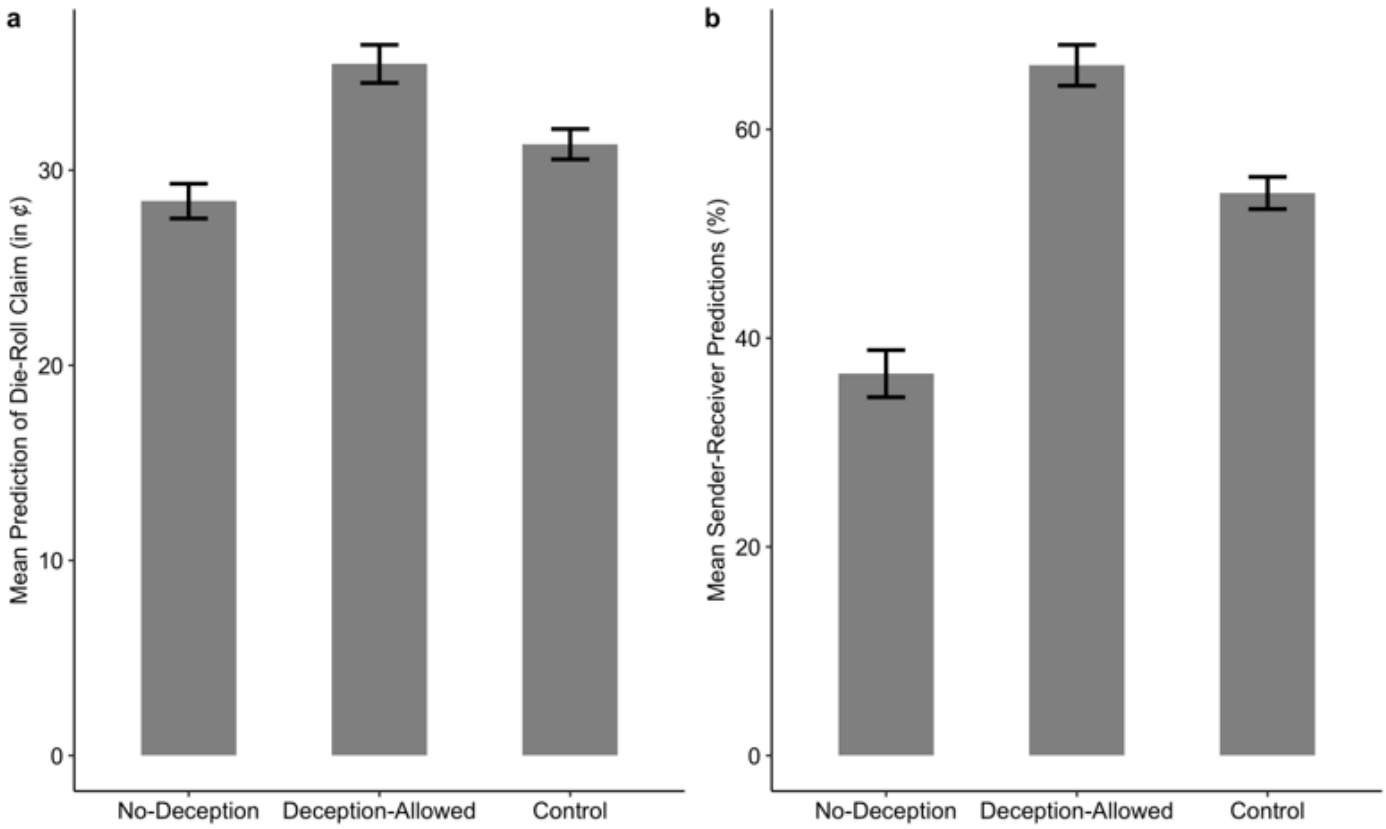

Figure 1

The graphs above contain the intuition survey results. On the left (a), mean distributions of intuitive predictions of die roll reports between the three experimental conditions. On the right (b), the distribution of average intuitive predictions (percentages) of deceptive messages in the sender-receiver game between conditions. Error bars indicate 95% CIs.

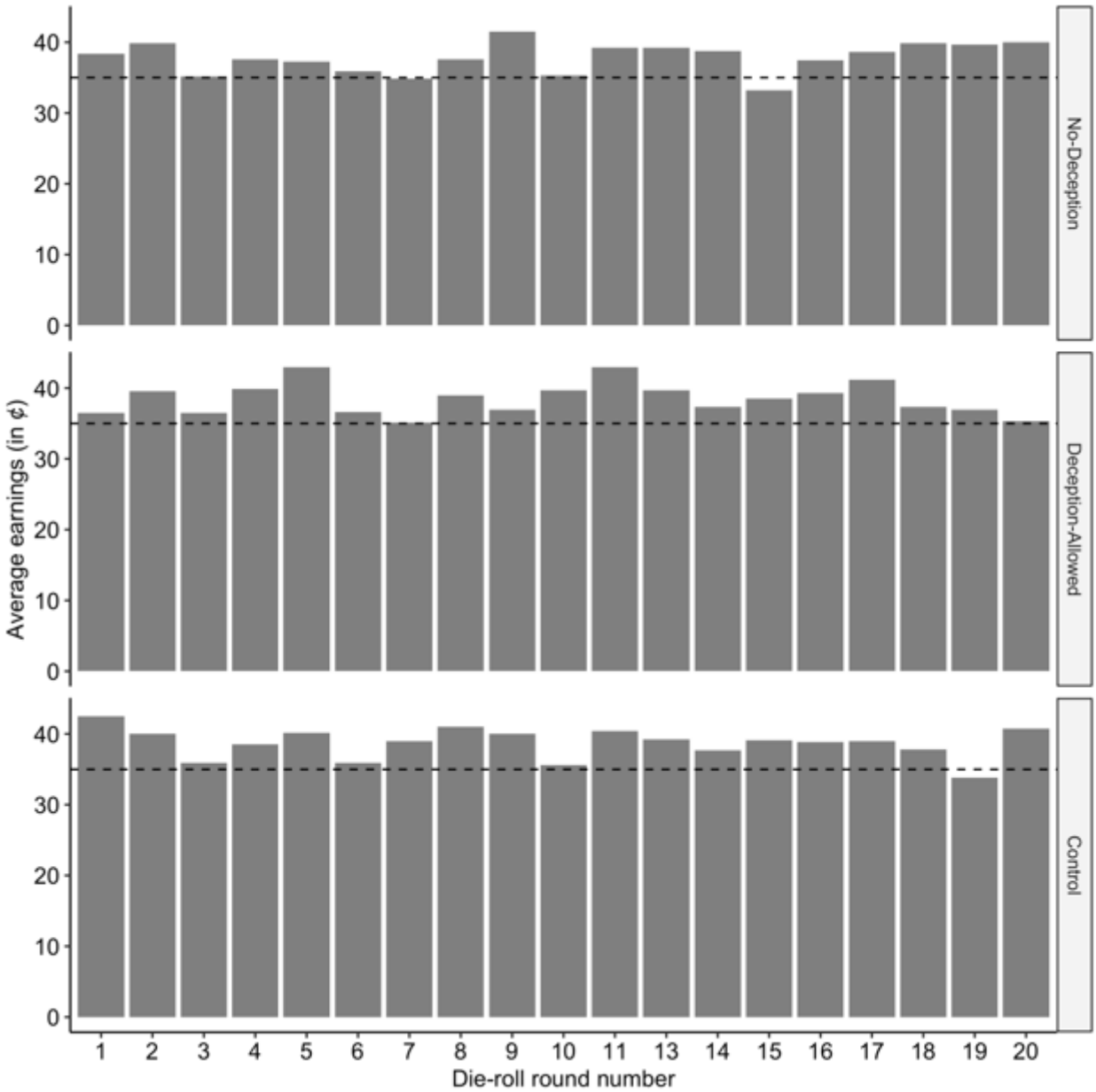

Figure 2

Average claims in the die roll task per experimental condition across the 20 rounds.

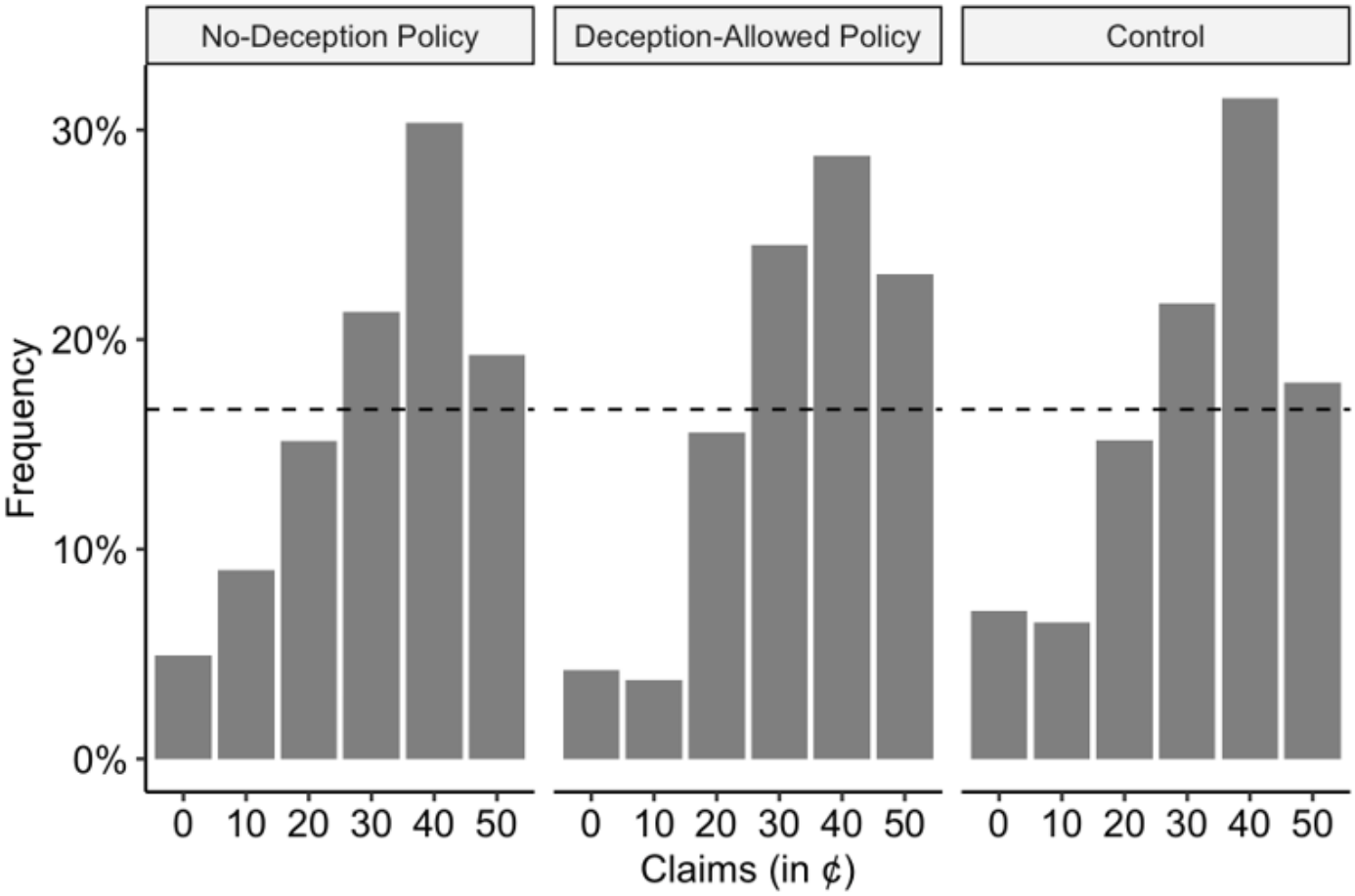

Figure 3

Distribution of claims in the die roll task in the three conditions. Dotted lines represent the expected uniform distribution of die-rolls, while bars represent the actual observed distribution.

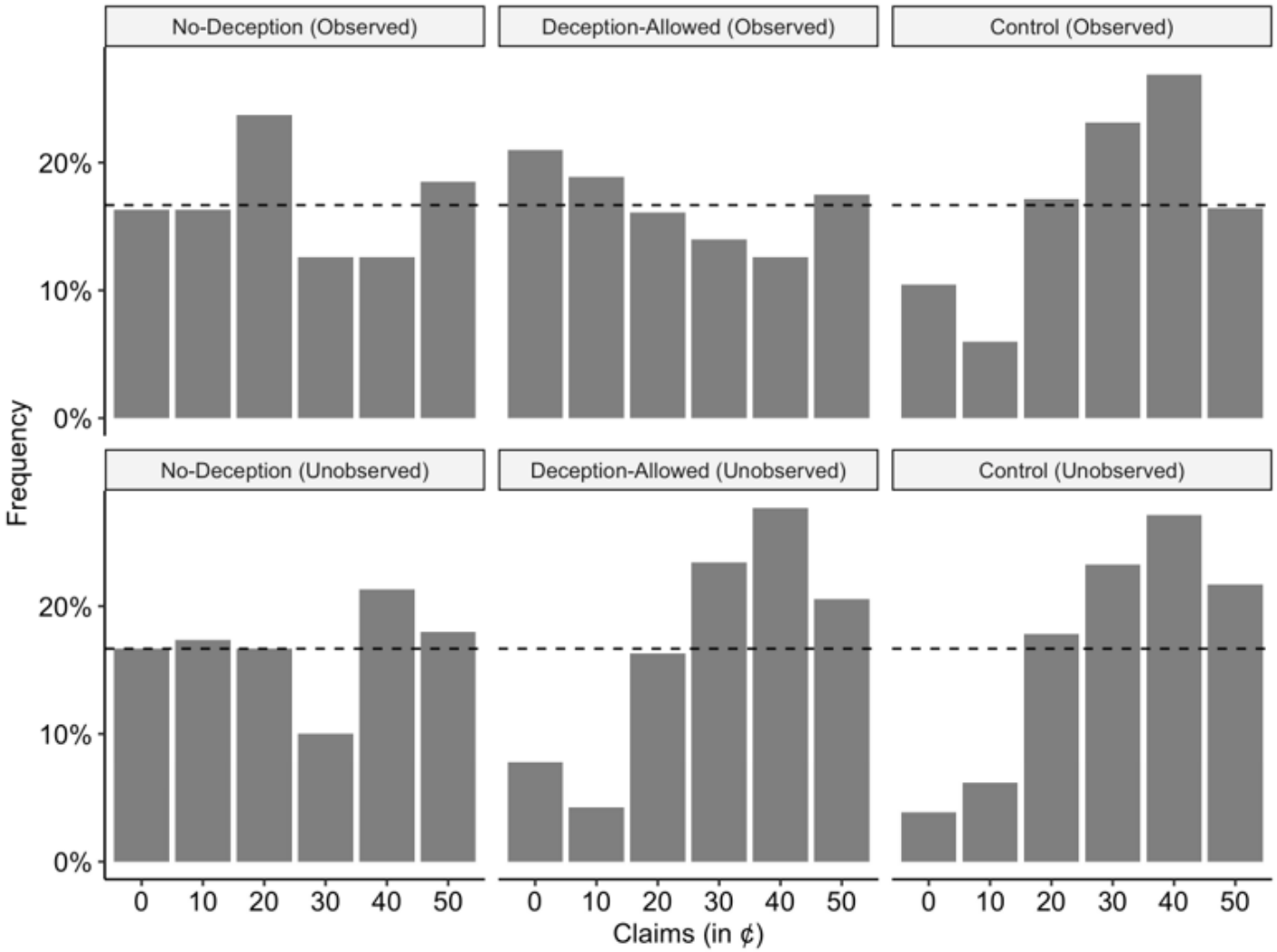

Figure 4

Distributions of reported claims between the six conditions. Dotted lines represent the expected distribution of die roll claims (i.e., 1/6), while bars represent the actual observed distribution.