

# Revenge is not blind: Testing the ability of retribution to justify dishonesty

Dar Peleg<sup>a</sup> Guy Hochman<sup>†</sup> Timothy Levine<sup>‡</sup> Yechiel Klar<sup>§</sup>

Shahar Ayal<sup>¶</sup>

## Abstract

In two studies, we tested the power of revenge as a justification mechanism that enables people to cheat with a clear conscience. Specifically, we explored the effects of prior dishonesty and unfairness towards participants on their subsequent moral behavior, as well as the physiological arousal associated with it. To this end, we employed a two-phase procedure. In the first phase, participants played one round of a bargaining game (the Ultimatum game in Study 1 and the Dictator game in Study 2) in which we manipulated whether the players had been treated (un)fairly and (dis)honestly by their opponent. In the second phase, they did a perceptual task that allowed them to cheat for monetary gain at the expense of their opponent from the first phase. In Study 1, participants also took a lie detector test to assess whether their dishonesty in the second phase could be detected. The behavioral results in both studies indicated that the opponent's dishonesty was a stronger driver than the opponent's unfairness for cheating as a form of retaliation. However, the physiological arousal results suggest that feeling mistreated in general (and not just cheated) allowed the participants to get revenge by cheating the offender while dismissing their associated guilt feelings.

Keywords: retaliation, ultimatum game, dictator game, cheating behavior, justifications, lie detector

---

\*Tel Aviv University. <https://orcid.org/0000-0002-5379-6378>

†Reichman University (IDC, Herzliya). <https://orcid.org/0000-0001-7486-8265>.

‡University of Alabama at Birmingham. <https://orcid.org/0000-0002-0293-4734>.

§Tel Aviv University.

¶Corresponding author. Baruch Ivcher School of Psychology, Reichman University (IDC, Herzliya), P.O.Box 167, Herzliya, 4610101, Israel, Email: [s.ayal@idc.ac.il](mailto:s.ayal@idc.ac.il). <https://orcid.org/0000-0002-7357-1969>.

This research was supported by the United States - Israel Binational Science Foundation (BSF) Grant # 2018365 to GH, TL, and SA.

During the revision process at the beginning of August 2021, the fourth author, Professor Yechiel (Hilik) Klar, passed away. Hilik was a great Social Psychologist and advisor. His insightful thinking and passion for research will continue to empower us.

Copyright: © 2021. The authors license this article under the terms of the Creative Commons Attribution 3.0 License.

# 1 Introduction

*Revenge is a kind of wild justice, which the more a man's nature runs to, the more ought law to weed it out.* (Francis Bacon).

Imagine two employees. One stole money from her boss because he cheated her out of her hard-earned money; e.g., by not covering her business expenses as promised. The other stole from her boss because he treated her unfairly; e.g., by forbidding her, unlike others, from working overtime. Here we examined whether people who are mistreated in different ways allow themselves to cheat their perpetrators to get revenge, regardless of the type of mistreatment. According to the Oxford English Dictionary, in the act of revenge, individuals respond to a wrong done to them or someone they care about by harming the transgressor. Here, we empirically tested whether the behavioral reaction to mistreatment (revenge) and its associated physiological arousal depends on the type of mistreatment.

According to a recent psychological account of dishonesty, when the possibility to cheat presents itself, people find themselves in a dilemma (Ayal et al., 2015). On the one hand, people strive to maintain a positive self-image (Jones, 1973; Rosenberg, 1979), and they value morality and honesty very highly (Chaiken et al., 1996; Greenwald, 1980; Sanitoso et al., 1990). On the other hand, most people are tempted to increase their personal gain through cheating even though the extent of cheating is usually rather small (Gino, Norton & Ariely, 2009; Hochman et al., 2016; Mazar, Amir & Ariely, 2008; Ordóñez et al., 2009; Shalvi et al., 2011). This conflict creates a type of psychological distress termed ethical dissonance (Ayal & Gino, 2011; Barkan et al., 2015), which is accompanied by physiological arousal (Hochman et al., 2016). This unpleasant arousal may serve as a moral gatekeeper that impedes dishonest behavior. Thus, the ability to benefit from cheating requires tension-reduction mechanisms (e.g., Barkan, Ayal & Ariely, 2012; Jordan et al., 2011; Peer et al., 2014; Sachdeva et al., 2009).

Behavioral ethics research shows that justifications can serve as a psychological tension-reduction mechanism for ethical dissonance (Shalvi et al., 2015). Justifications have been found to both facilitate dishonesty and increase the magnitude of cheating (e.g., Ariely, 2012; Ayal & Gino, 2011; Gino & Ariely, 2012; Shalvi et al., 2011; Mazar et al., 2008), while reducing guilt feelings (Gino et al., 2013; Gneezy et al., 2014). Justifications may also reduce the physiological arousal elicited by cheating behavior (Hochman et al., 2016; Hochman et al., 2021).

According to Shalvi et al. (2015), these self-serving justifications can emerge before or after ethical violations. Pre-violation justifications such as moral licensing (Merritt et al., 2010; Monin & Miller, 2001; Sachdeva et al., 2009) lessen the anticipated threat to the moral self by redefining questionable behaviors as excusable. Post-violation justifications, such as physical or symbolic cleansing (Gneezy et al., 2012; Gneezy et al., 2014; Mitkidis et al., 2017; Monin & Miller, 2001; Tetlock et al., 2000), ease the threat to the moral self through compensations that balance or lessen violations.

The current study was designed to examine whether and to what extent getting revenge against someone who engaged in mistreatment can serve as a pre-violation justification that increases dishonesty while decreasing its associated guilt feelings. Specifically, we examined whether different types of mistreatment would result in the same type of revenge. For instance, similar to our workplace example above, imagine two players who were mistreated, but in a different way. Player A was cheated by a classmate while Player B was treated unfairly by his colleague. As luck would have it, both players have a chance to get back at their offender. It is likely that Player A will feel like she has a justification for cheating her classmate back since she cheated her earlier. The question is whether Player B would feel as comfortable and entitled to cheating her colleague simply because he treated her unfairly. To test this question, our experimental design was constructed to differentiate between these two types of mistreatment and to ask whether they differed not only in the extent of cheating they elicited but also in the psychological distress they invoked. To the best of our knowledge, this distinction between unfairness and dishonesty and their effects on subsequent cheating behavior has rarely been explored in research.

### **1.1 The (iffy) link between mistreatment and revenge**

Some empirical evidence suggests that cheating may be an acceptable form of revenge for mistreatment. Mistreating others can lead to negative feelings (Pillutla & Murnighan, 1996; Van Der Zee et al., 2016). In turn, these negative feelings may increase unethical behavior. For example, Motro et al. (2016) primed participants with guilt, anger, or a neutral feeling. Then, participants engaged in the matrices task (Mazar et al., 2008), in which they could cheat to increase their monetary gain. The results showed that participants in the anger condition cheated more than twice as much as participants in the neutral condition and more than eight times as much as those in the guilt condition (Motro et al., 2016). Similarly, Gneezy and Ariely (2010) found that even a subtle form of mistreatment (a research assistant who is rude when instructing participants how to complete a task in a coffee shop), results in stealing from the offender as an act of revenge (Ariely, 2012).

A different kind of mistreatment that might lead to cheating as a form of revenge has been studied under the guise of unfairness. Research shows that people are markedly sensitive to fairness and reciprocity (e.g., Fehr & Schmidt, 1999; Houser, Vetter & Winter, 2012), and are willing to hurt and get their revenge on those who mistreat them, even at a personal cost (Gino & Pierce, 2009). For example, Houser et al. (2012) showed that people who were treated unfairly were more likely to cheat to restore the unfair 'balance' they recently experienced. Similarly, Schweitzer and Gibson (2008) found that subjects claimed they would be more likely to engage in unethical behavior when they perceived their counterpart to violate fairness standards. In a similar vein, Greenberg (1990) showed in a field study that employees who perceived a pay cut as unfair engaged more in employee theft than employees who perceived the pay cut as fair. At first glance, these findings suggest that mistreatment in general, and not specific types of wrongdoing, may justify retaliating by

cheating behavior, since people's thirst for revenge stems from their desire to get back at the person who hurt them (Bone & Raihani, 2015).

On the other hand, revenge might be more affected by the type of initial mistreatment. The term *lex talionis*, the law of retaliation, refers to the "tit for tat" principle (Jones, 1995), which suggests that people should be penalized in a way similar to the way they hurt the other. Based on this logic, people who were cheated, for example, are entitled to cheat their offender, whereas people who were treated unfairly are entitled to treat their offender unfairly. While unfairness is typically associated with the inequitable distribution of resources (e.g., Houser et al., 2012; Koning et al., 2010), dishonesty is associated with direct lying and deception committed by one party against another (e.g., Koning et al., 2010; Thielmann & Hilbig, 2019). To the extent that justifications facilitate dishonesty (Shalvi et al., 2015), people who were treated dishonestly, but not mistreated in general, might feel entitled to cheat their counterparts. For instance, a cheated individual might more easily recast stealing or cheating the other as merely claiming back what she justly deserves or taking back what was unlawfully taken away from her. Hence, revenge on a dishonest other may pave the way for guilt-free dishonest behavior (e.g., Bandura, 1999; Detert et al., 2008).

In sum, one line of research supports the claim that both unfairness and dishonesty (as well as other types of mistreatment) could serve as a justification for cheating behavior (e.g., Gneezy & Ariely, 2010; Motro et al., 2016). Other studies suggest that the effect of mistreatment on revenge is dependent on the specific type of mistreatment (e.g., Detert et al., 2008; Jones, 1995). However, none of these studies manipulated more than one type of mistreatment at a time. In the current work, participants were treated fairly or unfairly (by notifying them that they had been given an unequal share of the profits) while being treated honestly or dishonestly (by informing them they had been cheated). This made it possible to test whether cheating creates a different justification from unfairness, which authorized increased dishonesty as retaliation.

In addition to measuring the magnitude of cheating behavior, we used a lie detector test to examine the associated physiological arousal. The lie detector test measures the Galvanic Skin Response (GSR, which responds to sweat gland activity) and was used here to detect arousal linked to dishonesty. Since skin conductance increases or decreases proportionally to the individual's stress levels (Andreassi, 2000), it is considered an indication of deception (e.g., Barland & Raskin, 1975; Patrick, 2011; Krapohl, 2013). Previous research suggests that physiological arousal is correlated with ethical dissonance (Hochman et al., 2016) but that justifications can lessen the physiological arousal associated with cheating (Hochman et al., 2016; Peleg et al., 2019; Hochman et al., 2021). If revenge serves as a justification for dishonesty, it should lead to increased cheating while decreasing arousal, due to its inhibiting effect on guilt (Gino et al., 2013). The two lines of research presented above may thus imply that this physiological effect could be general and lead to similar arousal in response to cheating to avenge any type of mistreatment. Alternatively, physiological arousal in response to cheating may depend on the kind of mistreatment the victim wishes

to avenge.

In real-life situations, unfairness, as in the form of resource distribution, may be more common and visible than dishonesty, since fairness considerations are legitimate bargaining chips. In contrast, dishonesty is flatly disapproved and is by nature concealed. Moreover, deception detection is typically poor (Bond & DePaulo, 2006) and often occurs well after the fact (Park et al., 2002). Thus, in many real-life circumstances, people can decide whether and how to react to unfairness before they can detect and react to dishonesty. Consistent with this presumed time ordering, the setting of our first study exposed participants to (un)fairness, gave them a chance to decide how to react to it, and only then exposed them to (dis)honesty. By contrast, in our second study, participants were exposed simultaneously to both types of mistreatment.

## 1.2 Hypotheses

Based on the rationales of the two theoretical approaches outlined above, we derived two competing hypotheses on the role of revenge as a justification for dishonest behavior. According to the view that revenge to any type of mistreatment can serve as a justification for cheating back, a general mistreatment effect hypothesis was formulated in terms of the magnitude of cheating and its associated psychological arousal:

**Hypothesis 1a:** Participants will cheat an individual who behaved dishonestly towards them or treated them unfairly more than an honest or fair other.

**Hypothesis 2a:** Physiological arousal on the lie detector test will increase in response to cheating an honest or fair other. However, no increase will be observed in participants who cheat back in response to any kind of mistreatment.

By contrast, if we assume that revenge can justify cheating an individual back only after s/he behaved dishonestly first, but no other type of mistreatment can justify cheating, a distinctive mistreatment effect hypothesis can be formulated as follows:

**Hypothesis 1b:** Participants will cheat an individual who behaved dishonestly towards them more than an honest one and/or someone who treated them unfairly.

**Hypothesis 2b:** Physiological arousal on the lie detector test will not increase in response to cheating back a dishonest individual but will increase in response to cheating other individuals who were honest and/or treated them unfairly.

To test these competing hypotheses, we conducted two studies using a two-phase procedure. In the first phase, participants played one round of a bargaining game (the Ultimatum game in Study 1 and the Dictator game in Study 2) in which we manipulated whether the players had been treated (un)fairly and (dis)honestly by their opponent. In the second phase,

they played a perceptual task that allowed them to cheat for monetary gain at the expense of their opponent from the first phase. In Study 1, we also had a third phase, in which the participants underwent a lie detector test to determine whether we could detect their cheating level in the second phase. This research design enabled us to compare the effect of prior dishonesty or unfairness towards participants on their subsequent moral behavior as well as physiological arousal. All data are available online at <https://osf.io/5j3y7/files/>.

## 2 Study 1

Using a modified version of the Ultimatum game (Güth & Kocher, 2014), Study 1 examined the effect of the allocator's level of fairness and honesty on the participant's subsequent moral behavioral and physiological response on the lie detector. This served to test our competing hypotheses and examine whether unfairness and dishonesty have the same general effect on revenge (H1), or whether each has a distinctive effect (H2).

### 2.1 Method

#### 2.1.1 Participants

The sample size was predetermined using G\*Power (Faul et al., 2009). We used a medium-large effect size (0.25), with an  $\alpha$  of .05 and a power of .90. This calculation yielded a minimum target sample size of 185 participants. The final sample consisted of 197 Tel Aviv University students (81 males, 116 female), who volunteered to participate in the study. Their mean age was 24.41 years ( $SD= 4.218$ ). The participants received a 20 ILS (approximately 5.50 USD) show-up fee for their participation, but actual payment was contingent on their selections in the task and could amount to 40 ILS (approximately 11.00 USD).

#### 2.1.2 Design and procedure

The study had 2 computerized phases consisting of the manipulation and the retaliation task, after which the participants took a lie detector test. All phases were administered consecutively in a between-subject design with 4 conditions.

#### 2.1.3 The manipulation

In the manipulation phase, participants played a modified version of the Ultimatum Game (Güth et al., 1982). In the classic version of this game (Güth et al., 1982), one individual (the proposer) suggests how to divide a sum of money (the pie) with a second individual. The responder then has the opportunity to either accept the proposed division, in which case both players earn the amount proposed, or reject it, in which case both earn nothing. In the current study, participants played only the role of the responder. They were told that they

would play one trial with another participant on the computer. In actuality, there were no real proposers, and all offers were controlled by the computer (programmed in Visual Basic 6.0). Participants were seated in front of a computer screen and were asked to accept or reject the offer. To facilitate their decision, the following information was provided: the total pie (as reported by the proposer), the split of the total pie between them and the proposer (in percentages), and the absolute value (in new Israeli shekels, ILS, equivalent to about \$0.30 each) of what was being offered to them. To motivate their choices, participants were told that at the end of the experiment, they would be paid based on the amount allocated from that trial (if they accepted the offer).

Importantly, the “actual” amounts divided between the proposer and participants were 10 or 25 ILS. The split of the pie was either 50:50 or 80:20 (in favor of the proposer). However, participants were informed that the proposer would report the total amount to be divided at their discretion. The report could be 10, 15, 20, or 25 ILS. Participants were randomly assigned to one of the  $2 \times 2$  between-subjects conditions. As shown in Table 1, in the Fair & Honest condition, participants were offered 50% of the actual sum (5 out of 10 ILS). In the Fair & Dishonest condition, participants initially thought they had been offered 50% of the sum, but actually got only 20% (5 out of 25 ILS). In the Unfair & Honest condition, participants were offered 20% of the actual sum (2 out of 10 ILS). Finally, in the Unfair & Dishonest condition, the participants thought they had been offered 20% of the sum, but actually they got only 8% (2 out of 25 ILS).

TABLE 1: Distribution by condition of offered and actual sums

Condition	Fair & Honest	Fair & Dishonest	Unfair & Honest	Unfair & Dishonest
Proposal (participant, proposer)	5:5 ILS	5:5 ILS	2:8 ILS	2:8 ILS
Reported total sum	10 ILS	10 ILS	10 ILS	10 ILS
Actual total sum	10 ILS	25 ILS	10 ILS	25 ILS

After deciding whether to accept or reject the offer in the Ultimatum Game, participants were informed on the screen whether they had been treated dishonestly or not by the proposer. Thus, we manipulated both fairness and honesty in the same design, and randomly assigned the participants to one of four conditions.<sup>1</sup>

#### 2.1.4 The retaliation task

In the second phase, participants could increase their monetary gain by cheating. They were told that this monetary gain would be deducted from the earnings of their opponent

<sup>1</sup>Note that the honesty/dishonesty of the proposer was revealed to participants only after they responded to the UG offer (accept/reject). In the Results and Discussion sections of Study 1, we discuss the implications of this procedure.

(the proposer in the UG). Participants played the Flexible Dot task (Hochman et al., 2016), a visual perception task. In this task, participants are presented with multiple trials of 50 non-overlapping dots appearing in different arrangements within a square. The square is divided down the middle by a vertical line. The task is to determine which side of the screen, right or left, contains more dots. Although participants are always requested to accurately judge which side of the square contains more dots, they are paid more if they say there are more dots on a specific side regardless of their accuracy. (They were told this at the outset.) Participants aiming to maximize their profits should thus indicate the highest paying side on each trial regardless of the number of actual dots appearing in the square. Variations of this task have been validated as a reliable measure of cheating behavior in previous research (e.g., Ayal & Gino, 2011; Gino et al., 2010; Hochman et al., 2016; Mazar & Zhong, 2010; Peleg et al., 2019).

To examine the extent of cheating, we calculated the difference between trials in which participants chose the high-paying side when it was also the inaccurate one (beneficial errors) and trials in which participants chose the low-paying response when it was the inaccurate one (detrimental errors; see Hochman et al., 2016). This measure represents cheating, since it captures the extent to which participants purposely violated the experimental ‘rules’ set forth by the experimenters in the instructions to increase their personal gain. Participants played 100 trials of the task for real money. They were notified in advance that their accumulated gains would be deducted from the earnings of the proposer against whom they had played in the Ultimatum Game. Thus, they were given the opportunity to retaliate if they felt they had been mistreated.

### **2.1.5 The Lie Detector test**

Finally, participants were taken to another room where they underwent a lie detector test. In the test, they were asked about their own honesty on the retaliation dots task while their physiological arousal was monitored.

To examine dishonesty in the dots task, participants’ GSR was recorded using the Limestone technologies© system (Data Pac USB Ltd.) connected to a laptop computer. The stimulus questions for the test and GSR signals were recorded using cogito software (Ltd.) by SDS © (Suspect detection systems). Participants were seated in front of the computer and connected to the GSR machine. Two 24k gold plated electrodes were attached to the palmar surface of participants’ index and ring fingertips of their right hand. The data were sampled at 60 Hz, with 16 bits per sample. The GSR data were down-sampled to 30 Hz after smoothing by 3 sample kernels. The data were then separated using wavelets to analyze the peaks of tension and temporal responses by an in-house MATLAB script. The exam itself was fully computerized. Findings have confirmed that the automation of polygraph tests increases lie detection accuracy (Honts & Amato, 2007; Kircher & Raskin, 1988) and is viewed more positively by both examiners and examinees (Novoa et al., 2017). Note



that this short-automated lie detector test is successfully used worldwide in 15 different countries as well as in international air and sea terminals (Suspect Detection Systems. Ltd).

GSR is the most frequently used measure in physiological lie detection (Vrij, 2000) and is considered the most sensitive and critical parameter of the three channels used in polygraph examinations (Kircher & Raskin, 2002). The GSR measure was the baseline-to-peak amplitude difference in the 0.5-s to 8-s time window from the response provided by the participant. The measures were composed of arousal in response to 6 changing probable lie control (PLC) questions and 3 changing irrelevant questions. The control question arousal responses were compared to the arousal in response to the two relevant test questions (see Appendix for all questions). This format is in line with the most frequently administered polygraph tests known as the Control Question Test (CQT; see Honts & Reavy, 2015; Raskin, 1986; Vrij & Fisher, 2016) and is one of the most common instruments to test people's integrity in the US, Europe, Israel, and South Africa (Honts & Reavy, 2015; Raskin, Honts & Kircher, 2014).

Questions were presented to the participants on the computer screen and aurally through a headphone set placed on their ears. Responses were recorded orally via a microphone attached to the headset. All participants were notified before the test that if they were identified as honest, they would get a 5 ILS bonus. In actuality, at the end of the study, all participants (regardless of their performance on the lie detector test) were given the bonus. Participants answered all the questions in a changing order in 3 consecutive rounds. Their GSR to the target questions (regarding cheating on the dots task) compared to the control questions were calculated to obtain the final Sympathetic Arousal Index, which fell between  $-1.2$  and  $1.8$  98% of the time. This continuum measure allowed for a precise analysis of physiological arousal levels on the target questions.

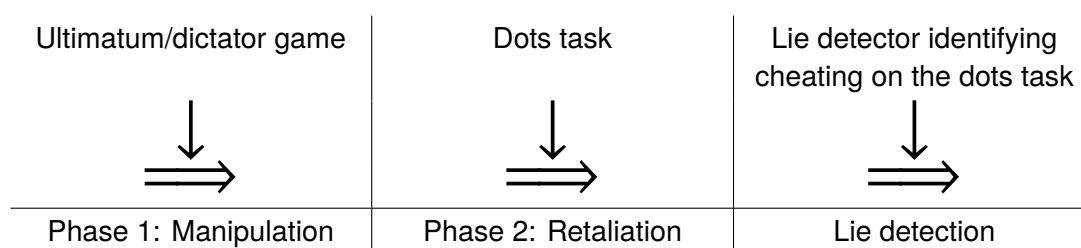


FIGURE 1: Experimental Procedure. Lie detection was implemented only in Study 1.

There is substantial body of laboratory and field research in forensic settings that supports the rationale and the validity of the PLC version of the CQT (see reviews by Honts, 2004; Raskin & Honts, 2002). Given the constraints of the laboratory setting and the administration of multiple tests, we used a semi-automated version of the standard CQT test. The procedure, which is identical to the one used in Peleg et al. (2019), was composed of a pre-test phase consisting of a brief explanation of the upcoming process followed by a short presentation of all the test questions by the interviewer (research assistants), and the participant's informed consent, which conformed closely to standard CQT procedures.

Figure 1 outlines the experimental procedure. Immediately upon completion of the study, participants were debriefed on the experimental manipulation, its purpose, the aim of the experiment and were paid.

## 2.2 Results

### 2.2.1 Behavioral results

First, we examined acceptance rate on the UG task. Since participants were informed about the actual total sum after making their decision to accept or reject the offer, we expected the acceptance rate to be affected by fairness but not by honesty, since it was unknown. Similar to previous findings (e.g., Camerer, 2003; Güth & Kocher, 2014), participants exhibited higher acceptance rates for fair than for unfair offers. The acceptance rate was 86% (74 out of 86 participants) when the offer was fair, and 32.4% (36 out of 111 participants) when the offer was not fair. A two-tailed chi-square test revealed that the difference between these two conditions was significant ( $\chi^2(1, n=197) = 56.485, p < 0.001$ ). Thus, as expected, unfair offers were rejected more often than fair offers. This tendency was not influenced by the honesty level of the proposer, which was unknown to the participants at this stage.

Next, we asked whether the cheating behavior in Phase 2 was affected by whether the proposer had been fair and honest (or not) on the UG. To examine cheating levels on the dots task, we classified the errors (selections of the side that had fewer dots) into two types: beneficial errors; that is, trials in which participants chose the high-paying side when more dots were on the low-paying side of the square, and detrimental errors; that is, errors in which participants chose the low-paying side when more dots were on the high-paying side (Hochman et al., 2016). The cheating level was calculated as the difference between the percentage of beneficial and detrimental errors, which estimates the extent to which participants purposely violated the instructions to increase their personal gain.

Analyzing the cheating results in the second phase, we realized that our original design had shortcomings. As previously mentioned, the (dis)honesty of the proposer was revealed to participants only in the second phase after they had already decided to accept or reject the offer. Thus, this methodological issue actually eliminated one of the experimental conditions: in the Fair & Dishonest condition, participants received a fair offer (5:10). Still, they then discovered after making their choice that the proposer was not honest. That is, they were informed that instead of getting 50% of the total pie (5 out of 10), they had actually been offered 20% of the total pie (5 out of 25). Thus, from the second phase onward, our paradigm did not include a Fair & Dishonest condition, but rather a High-Unfair & Dishonest (2 out of 25), and Low-Unfair & Dishonest (5 out of 25) condition. Since there were no differences in the subsequent cheating levels on the dots task between the High- and Low-Unfair & Dishonest conditions ( $M = 44.44, SD = 49.3$  for the High Unfair & Dishonest and  $M = 45.79, SD = 44.85$  for the Low Unfair & Dishonest,  $t = 0.142, p = 0.888$ , Cohen's  $d = 0.028$ ), these two treatments were collapsed into one condition. The behavioral

results on the dots task were thus analyzed for three experimental conditions: Fair & Honest (5:10), Unfair & Honest (2:10), and Unfair & Dishonest (High, 2:25; and Low, 5:25).

A 3 X 2 between subject ANOVA analysis was conducted to examine the main effects of the experimental condition (fair & honest, unfair & honest, and unfair & dishonest) and the UG response (accept/reject) on cheating behavior. This analysis revealed a significant main effect for condition ( $F(2,194) = 3.06, p = 0.049, \eta_p^2 = 0.031$ ) but not for the UG response ( $F(1,195) = 0.001, p = 0.994, \eta_p^2 = 0.001$ ). More importantly, adding the interaction between the two factors to the model revealed a significant interaction effect ( $F(2,191) = 3.012, p = 0.052, \eta_p^2 = 0.031$ ).

Post-hoc multiple comparisons tests (LSD) were conducted to further examine how unfairness and dishonesty of the proposer affected the cheating level of the participants who accepted or rejected the offer on the UG. When participants rejected the offer ( $n = 87$ ), there was no difference in cheating level between the Unfair & Honest ( $M = 43.6, SD = 42.1$ ) and the Unfair & Dishonest ( $M = 35.3, SD = 42.3; p = 0.374$ ) conditions. Thus, if the participants had already retaliated by rejecting the offer, they tended to cheat to a similar extent, regardless of whether they were mistreated unfairly or dishonestly (since only 4 participants rejected the offer in the Fair & Honest condition, we could not examine the difference between this condition and the other two). However, participants who accepted the offer ( $n = 110$ ) tended to cheat an Unfair & Dishonest proposer more ( $M = 50.7, SD = 48.4$ ) than a Fair & Honest proposer ( $M = 24.0, SD = 49.8; p = 0.016, 95\% C.I. = [5.007, 48.294]$ ) or a Unfair & Honest proposer ( $M = 20.6, SD = 47.6; p = 0.023, 95\% C.I. = [4.262, 55.817]$ ). No difference was found between the Fair & Honest and Unfair & Honest conditions ( $p = 0.817$ ). These results are presented in Figure 2.

Thus overall, participants who retaliated by rejecting an unfair (or unfair and dishonest) offer appeared to be less likely to also cheat their offender. This was true even if they were treated both unfairly and dishonestly. However, participants who had not already retaliated tended to cheat those who had previously cheated them more, not those who had only treated them unfairly. These latter results support the distinctive mistreatment effect hypothesis (H1b) and suggest that revenge is sensitive to the type of mistreatment.

### 2.2.2 Physiological results

To test the possible justification effect of cheating others to get revenge for their wrongdoing, we examined physiological arousal on a lie detector test. Specifically, after the dots task, participants underwent a lie detector test in which their moral conduct on the task was evaluated. This was done to assess whether the arousal associated with cheating behavior (Hochman et al., 2016; Wang, Spezio & Camerer, 2010) was affected by whether the participants were treated unfairly, dishonestly (or both), and by their actual cheating level. Importantly, we examined whether cheating in general, which is assumed to be reflected in physiological arousal on the lie detector test (Raskin et al., 2014), could be justified by a different kind of mistreatment. Thus, we divided participants into high versus low

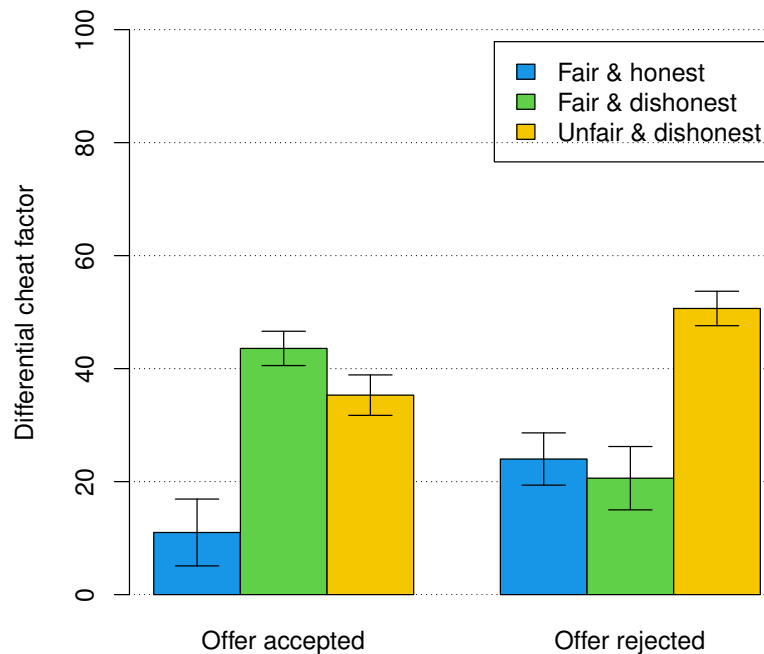


FIGURE 2: Participants' differential cheat Factor as a function of the initial proposer's offer and the UG response (accept/reject): Offer rejected: Fair & Honest ( $n = 4$ ), Unfair & Honest ( $n = 48$ ), Unfair & Dishonest ( $n = 35$ ); Offer accepted: Fair & Honest ( $n = 29$ ), Unfair & Honest ( $n = 18$ ), Unfair & Dishonest ( $n = 63$ ). Vertical lines indicate the standard errors.

cheaters (based on the median split, which was the 36.0% difference between beneficial and detrimental errors; Peleg et al., 2019; Hochman et al., 2021).

Since we aimed to examine the relationship between physiological arousal and actual cheating behavior, we compared the arousal of participants who cheated to a large extent on the dots task to the arousal of participants who cheated to a small extent. A 3 X 2 (Condition X Cheating level) between-subjects ANOVA was conducted to examine the main effects of the experimental conditions (Fair & Honest, Unfair & Honest, Unfair & Dishonest) and the cheating level (high vs. low) on the physiological arousal index on the lie detector test. This analysis revealed a significant main effect for experimental condition ( $F(2, 189) = 3.31, p = 0.039, \eta_p^2 = 0.034$ ), but not for cheating level ( $F(1, 190) = 0.002, p = 0.966, \eta_p^2 = 0.001$ ). Adding the interaction between the two factors to the model revealed a non-significant interaction effect ( $F(1, 186) = 1.351, p = 0.262, \eta_p^2 = 0.014$ ). Post-hoc multiple comparisons tests (LSD) were conducted to delve into the significant main effect of the experimental condition. This analysis indicated that the physiological arousal index in the Fair & Honest condition ( $M=0.698, SD = 0.77$ ) was significantly higher than the Unfair & Honest ( $M=0.312, SD = 0.60, p = 0.013, 95\% C.I. = [0.083, 0.689]$ ) and the Unfair & Dishonest conditions [ $M=0.390, SD = 0.76, p = 0.034, 95\% C.I. = [0.023, 0.592]$ ). However, the difference between the two latter conditions (Unfair & Honest vs. Unfair & Dishonest) was not significant ( $p = 0.499$ ). These results are presented in Figure 3.

This pattern of results runs counter the behavioral results. Although participants in the

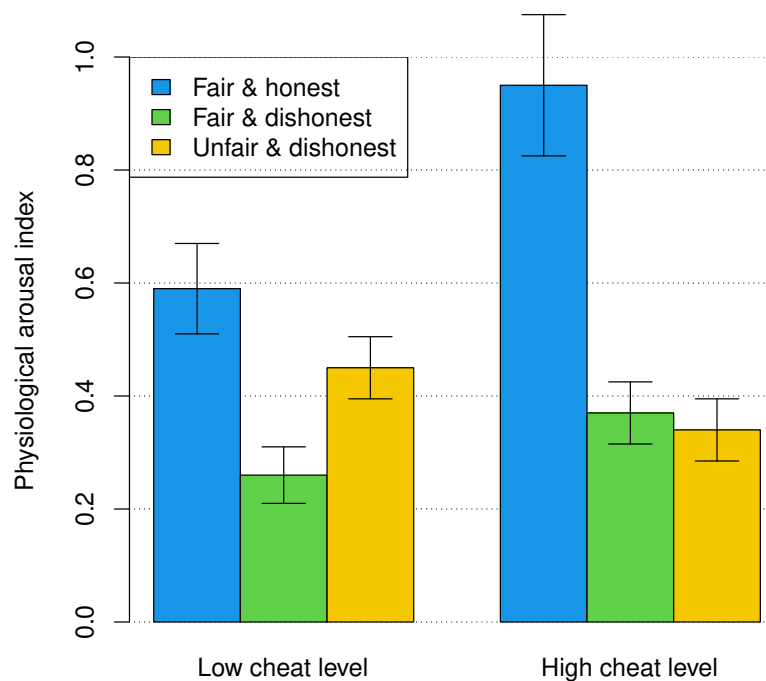


FIGURE 3: Participants' physiological arousal index as a function of cheat level and initial proposer's offer. Low cheat level: Fair & Honest ( $n = 23$ ), Unfair & Honest ( $n = 32$ ), Unfair & Dishonest ( $n = 42$ ); High cheat level: Fair & Honest ( $n = 10$ ), Unfair & Honest ( $n = 31$ ), Unfair & Dishonest ( $n = 54$ ). Vertical lines represent the standard errors.

Fair & Honest condition exhibited the lowest level of cheating, they also showed higher levels of physiological arousal. By contrast, participants who experienced any type of mistreatment cheated to a greater extent but also experienced lower levels of physiological arousal. As indicated in the literature, cheating behavior is associated with increased physiological arousal (e.g., Hochman et al., 2016, Wang et al., 2010). Thus, this pattern of results supports the general mistreatment hypothesis (H2a) and suggests that both the unfairness and dishonesty of the proposer could have served as a justification that diffused the ethical dissonance associated with cheating behavior. Presumably, when the participants were mistreated in general (and not just cheated), they may have felt entitled to get revenge by cheating the proposer, while maintaining a clear conscience.

### 2.3 Discussion

The results provide initial support for the claim that behaviorally, revenge is not blind. Similar to previous findings (Brethel-Haurwitz et al., 2016), our participants were prone to rejecting unfair offers and punishing the proposer even at a personal cost. Second, participants who had not already punished the proposer tended to cheat the dishonest proposer more than the unfair one. These results support the distinctive mistreatment effect (H2a) since they indicate a differential retribution response to unfairness vs. dishonesty. This pattern of results suggests that the participants may have adapted their retaliation to fit

the way they were mistreated. Thus, on the behavioral level, revenge might be considered a sensitive form of an 'eye for an eye' between the type of mistreatment and the nature of the retaliation. On the other hand, the physiological results suggest that both types of mistreatment served as a justification to increase the level of cheating while attenuating the tension associated with it.

One main characteristic of our experimental setting is that participants were first exposed to unfairness and only then to dishonesty. This sequential order resembles many realistic social interactions in which fairness is more visible than honesty. Individuals often know immediately if someone has treated them fairly or not, but dishonesty is not usually detectable until later. However, this sequential nature also makes the distinction between unfairness and dishonesty more difficult to determine. This stems from the fact that participants could retaliate by cheating only after they already had a chance to avenge the unfair offer by rejecting it. Presumably, the revenge that exacted by participants on unfair proposers by rejecting their offer affected the follow-up decision concerning the magnitude of cheating. Alternatively, after rejecting the offer, participants perhaps no longer cared about whether the proposer was honest or not toward them, because it did not affect them in the end.

The second issue related to this sequential design is a consequence of this information gap between the two phases. The first phase of the design precluded examining a fair and dishonest proposer. Once the participants had been informed of the proposer's dishonesty in the second phase, they also found out that the seemingly fair offer (5:10) was also unfair (5:25).

To address these two limitations, we first isolated specific comparisons and focused solely on those participants who had accepted the offer and thus were not influenced by earlier revenge. The results showed that participants who accepted the offer cheated those who were Low Unfair & Dishonest (5:25) more than those who were Unfair & Honest (2:10) ( $t(61) = 2.147, p = 0.036, \text{Cohen's } d = 0.559$ ). That is, participants in the Low Unfair & Dishonest condition cheated more even though they received a greater absolute monetary sum, and when their relative part of the total pie (20%) was held constant. This lends credence to the claim that they preferred to take revenge in the same way they had been mistreated. To further test this interpretation, we ran Study 2 with a revised procedure that exposed unfairness and dishonesty at the same time and allowed the participants only one opportunity to get revenge.

### 3 Study 2

Study 2 was designed to address the limitations in Study 1 by using the Dictator Game (Engel, 2011). This allowed us to simultaneously manipulate the opponent's levels of unfairness and dishonesty and test which effect would lead to stronger revenge. Specifically, after participants were informed about how much money they received, they were given a

chance to retaliate by cheating back on a subsequent task. Unfortunately, due to COVID-19 restrictions, this study was run online and did not include a lie detector test.

### **3.1 Method**

#### **3.1.1 Participants**

The sample size was predetermined using G\*Power (Faul et al., 2009). We used a medium-large effect size (0.30), with an  $\alpha$  of .05 and a power of .90. This calculation yielded a minimum target sample size of 195 participants. The final sample consisted of 231 Prolific academic participants (84 males, 145 females, 2 other) who volunteered to participate in the study. The mean age was 36.4 years ( $SD = 13.98$ ). Participants received €0.2 for their participation, and up to an additional €4.8, contingent upon their selections in the task.

#### **3.1.2 Design and procedure**

Similar to the first two phases of Study 1, the design implemented 2 computerized phases: the manipulation and the retaliation task. The phases were administered consecutively in a 2 (Fair vs. Unfair)  $\times$  2 (Honest vs. Dishonest) between-subject design.

#### **3.1.3 The manipulation**

In the first phase, participants played a variation of the Dictator Game (DG; Engel, 2011). In the classic version of this game, one individual (the allocator) suggests how to divide a sum of money (the pie) with a second individual (the receiver). The receiver does not affect the allocated division. In the current study, participants played only the role of the receiver. They were told that they would play one trial with another participant via the computer. In actuality, there were no real allocators, and all offers were controlled by the computer. Participants were told that the other player received an initial amount of chips (5000, 10000, 15000, or 25000, with equal probability) from the experimenter. This initial amount is only known to the other player, not to the participant. Then, the other player would decide how much of this amount to keep and how much to give to the participant. The participants were further informed that the other player would notify them about the initial amount (the total pie) and their proposed allocation. To make this clear, the participants were also presented with the split of the pie in percentages. Finally, the participants were informed that each chip was worth €0.25, and that they would be paid at the end of the experiment based on their allotted sum.

In all conditions, the participants were offered 1000 chips (i.e., €2.5), and they differed solely in the information provided to them about the total pie and the split, as follows. In the Fair & Honest condition, participants were told that the other participant got a total of 2000 chips, and decided to give them 1/2 of the overall amount. In the Fair & Dishonest condition, participants were informed that the other participant got a total of 1000 chips,

and decided to give them all that amount. In the Unfair & Honest condition, participants were told that the other participant got a total of 3000 chips, and decided to give them 1/3 of the overall amount. Finally, in the Unfair & Dishonest condition, the participants were told that the other participant got a total of 3000 chips, and decided to give them 1/3 of the overall amount. The participants were randomly assigned to one of these 4 conditions.

After the offer in the DG was presented to them, the participants were informed whether they had been treated dishonestly and unfairly (or not) by the allocator. Specifically, in the the Fair & Honest and Unfair & Honest conditions, they were told that indeed the other player got the reported amount of chips. By contrast, in the Fair & Dishonest condition, the participants were informed that the actual amount was 2000 chips (rather than the reported 1000), and in the Unfair & Dishonest condition, 5000 chips (rather than the reported 3000). Thus, we were able to manipulate both fairness and honesty in the same design. These experimental conditions are summarized in Table 2.

TABLE 2: Distribution by condition of offered and actual sums.

Condition	Fair & Honest	Fair & Dishonest	Unfair & Honest	Unfair & Dishonest
Allocation (participant, allocator)	1000:1000 (€2.5: €2.5)	1000:0 (€2.5; €0)	1000:2000 (€2.5: €5.0)	1000:2000 (€2.5: €5.0)
Reported total sum (in chips)	2000	1000	3000	3000
Actual total sum	2000	2000	3000	5000

To validate our manipulation and examine whether participants perceive an unfair offer as unfair and dishonest offer as dishonest, we ran a pilot on 203 Prolific Academic participants (69 males, 134 females;  $M = 32.72$  years,  $SD = 12.42$ ). Participants received €0.88 for their participation. The experiment was identical to the  $2 \times 2$  between participants design of the DG described above, except that the amounts were presented in EUROS and not in chips. After the participants saw what the allocator reported getting and offered to them, and then learned the actual total sum, we asked them to rate the extent to which they perceived the allocator to be a fair and honest person (on a 1–10 scale).

The results of this pilot are presented in Table 3.  $2 \times 2$  between subjects ANOVA analyses were conducted to compare the effects of fairness (fair vs. unfair) and honesty (honest vs. dishonest) in the DG on the two dependent measures (perceived fairness and perceived honesty of the allocator). For perceived fairness, there was a significant main effect for fairness ( $F(1,198) = 254.799$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.563$ ) and a weaker but significant main effect for honesty ( $F(1,198) = 98.853$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.333$ ). In addition, there was no significant interaction between the two factors ( $F(1,198) = 0.587$ ,  $p = 0.445$ ,  $\eta_p^2 = 0.003$ ). Similarly, the analysis of perceived honesty yielded a significant main effect for honesty  $F(1,198) = 333.894$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.628$ , and a weaker but significant main effect for



fairness ( $F(1,198) = 67.484, p < 0.001, \eta_p^2 = 0.254$ ). Here again, there was no significant interaction between the two factors ( $F(1,198) = 0.00, p = 0.996, \eta_p^2 = 0.00$ ).

TABLE 3: Mean perceived fairness (upper panel) and honesty (lower panel) of the allocator in the pilot, as a function of the experimental design. SDs are in parentheses.

		Fair	Unfair
Perceived fairness	Honest	9.41 (1.7)	5.45 (2.1)
	Dishonest	7.02 (2.2)	2.66 (1.4)
Perceived honesty	Honest	9.22 (1.6)	7.08 (2.0)
	Dishonest	4.46 (2.2)	2.32 (1.4)

These results lend credence to the validity of our manipulation since the level of fairness and honesty were accurately reflected in the way participants perceived the allocator. That is, participants perceived allocators who made fair offers as fair, and honest offers as honest. At the same time, however, the results also highlight the interplay between these two factors, since unfair allocators were also perceived as dishonest (albeit to a lesser extent) and vice-versa.

### 3.1.4 The retaliation task

In the second phase, participants played the same Flexible Dot task (Hochman et al., 2016) as in Study 1, for actual money that was deducted from the earnings of their opponent (the allocator in the DG). The participants got 1 chip (equal to 0.025 penny) for indicating that there were more dots on the left side and 10 chips (equal to 0.25 penny) for indicating that there were more dots on the right-hand side.

## 3.2 Results

A  $2 \times 2$  between subject ANOVA analysis was conducted to examine if cheating behavior in Phase 2 was affected by whether the allocator was fair and honest (or not) on the DG. This analysis revealed a nearly significant main effect for fairness ( $F(1,229) = 3.397, p = 0.067, \eta_p^2 = 0.015$ ) and a significant main effect for honesty ( $F(1,229) = 7.521, p = 0.007, \eta_p^2 = 0.032$ ). Adding the interaction between the two factors to the model, revealed a non-significant interaction effect ( $F(1,227) = 0.073, p = 0.787, \eta_p^2 < 0.001$ ). Planned contrasts were conducted to examine how unfairness and dishonesty on the part of the allocator affected the cheating level. While participants tended to cheat a Fair & Honest allocator ( $M = 15.5, SD = 36.5$ ) less than an Unfair & Honest one ( $M = 25.5, SD = 35.3$ ), this difference was not significant ( $t(227) = 1.418, p = 0.157, \text{Cohen's } d = 0.277$ ). However, in the Fair & Honest condition, cheating was significantly lower than in the Fair & Dishonest condition ( $M = 29.9, SD = 39.9; t(227) = 2.084, p = 0.038, \text{Cohen's } d = 0.376$ ), or the Unfair &

Dishonest one ( $M = 37.3$ ,  $SD = 36.4$ ;  $t(227) = 3.206$ ,  $p = 0.002$ ,  $Cohen's d = 0.596$ ). Finally, a nearly significant difference was found between participants who cheated an Unfair and Dishonest proposer and those who cheated an Unfair and Honest one ( $t(227) = 1.705$ ,  $p = 0.089$ ,  $Cohen's d = 0.328$ ). These results are presented in Figure 4.

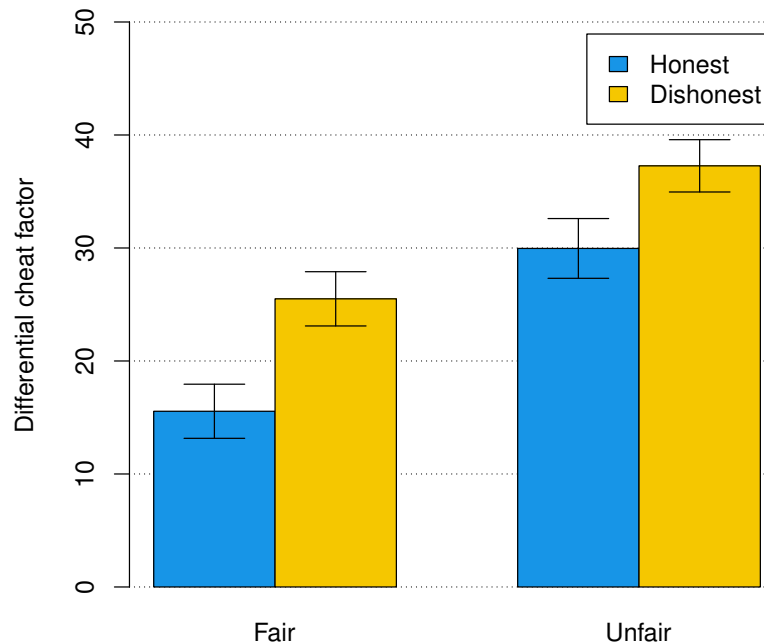


FIGURE 4: Participants' differential cheating factor as a function of the proposer's offer in the DG: Fair & Honest ( $n = 58$ ), Unfair & Honest ( $n = 54$ ), Fair & Dishonest ( $n = 57$ ), and Unfair & Dishonest ( $n = 62$ ). Vertical lines indicate the standard errors.

Thus overall, this pattern of results suggests that participants' cheating level was more affected by the honesty than the fairness of the proposer. In fact, the most cheating was observed against dishonest proposers (regardless of whether they were fair or not), and the main effect of fairness was only marginally significant. As in Study 1, these behavioral results support the distinctive mistreatment effect hypothesis (H1b) and suggest that revenge is not blind. Moreover, the magnitude of mistreatment was directly associated with the magnitude of retaliatory cheating.

## 4 General discussion

The current studies were designed to assess whether retaliation can serve as a pre-violation justification that increases cheating while at the same time decreasing its associated guilt feelings. Although some studies have concluded that all types of mistreatment can justify retaliating in the form of cheating (Motro et al., 2016), other findings indicate that different kinds of mistreatment may result in different patterns of revenge (Schurr & Ritov, 2016). To disentangle these hypotheses, we designed a distinctive setting in which we manipulated whether participants were treated (un)fairly and (dis)honestly. This made it possible to

examine how they chose to retaliate to different types of mistreatment, and how this affected their associated physiological arousal.

Behaviorally, both studies supported the distinctive mistreatment effect (H2a) since they indicated that the opponent's dishonesty was a stronger driver for cheating as a form of retaliation than the opponent's unfairness. This was true in the sequential design of Study 1, where participants who had not already punished the proposer by rejecting the offer in the UG tended to cheat dishonest proposers more than the unfair ones. Similarly, in Study 2 participants tended to cheat the dishonest allocators more, whereas the effect of unfairness was not quite significant. However, the results of the pilot study and the finding that the participants cheated to some extent those who cheated them or treated them unfairly suggest that it may be hard in general for people to tease apart these two different types of mistreatment.

It is worth noting that the current results cannot be explained by the absolute monetary sum of the different offers or by the relative amount of money earned by the opponent. This is because the two designs manipulated honesty and fairness while controlling for the absolute payment of the participants as well as the percentage out of the total pie. Thus, comparing the effects of a payment of 5 offered out of 10 vs. 5 out of 25 in the UG in Study 1 kept the participants' payoff constant while manipulating fairness. Similarly, a comparison of the offers of 2 out of 10 vs. 5 out of 25 preserved a constant ratio while manipulating honesty. The results showed that participants chose to cheat more when the offer was 2 or 5 out of 25 (dishonest) than when the offer was 2 or 5 out of 10 (honest), even if they were offered the same or a lower amount of money. Moreover, the fact that in Study 2, participants in all conditions received the same absolute amount (€2.5) and still cheated the dishonest allocators more further undermines this alternative explanation.

The designs in these two studies differed in some critical ways. In Study 1, participants were first exposed only to the (un)fairness of the proposal. They were then required to react to the proposer's offer by rejecting or accepting it, and only then were informed of the proposer's (dis)honesty and had an opportunity to retaliate by cheating. In Study 2, however, participants were exposed to both types of mistreatment (unfairness and dishonesty) simultaneously and could only react to this mistreatment by cheating on the retaliation task. While the variety of tasks certainly contributed to the internal validity of our findings, they may also contribute to its external validity, as they represent different real-life situations. The sequential design in Study 1 reflects many realistic social interactions in which fairness is more visible than honesty. Individuals may know immediately if someone has treated them fairly or not, by contrast to dishonesty which may not be immediately detectable. If indeed unfairness can be immediately exposed and responded to, it may no longer sustain negative feelings. Conversely, if dishonesty is frequently made known in retrospect, it could generate, sustain, and perhaps even inflate negative feelings. These feelings could pave the way for avenging and cheating back as a possible mechanism for reducing victimhood without additional guilt. Moreover, Study 2 showed that even when we simultaneously

manipulated (un)fairness and (dis)honesty, dishonesty appeared as a much stronger driver for vindictive cheating. Thus, at least when revenge can take the form of cheating, prior dishonesty was perceived as a more severe violation than unfairness. Future research should test whether the unique severity of dishonesty also holds when participants consider revenge in terms of unfairness or other types of mistreatment.

Taken together, the current results suggest that revenge is not blind when individuals choose how to retaliate by cheating. However, it is blind when they examine their conscience since, at least in terms of physiological arousal, the participants seemed to feel less accountable when cheating someone who mistreated them. These results contribute to the field of lie detection by providing another example where cheating is less likely to be detected when the unethical behavior can be justified (Hochman et al., 2021). Alongside the fact that people value morality and honesty very highly (Chaiken et al., 1996; Greenwald, 1980; Sanitoso et al., 1990) and strive to maintain a positive moral image (Jones, 1973; Rosenberg, 1979), the results nevertheless suggest that preferring sincerity over revenge in these situations might also take a psychological toll. Thus, previous mistreatments may allow individuals to balance the scale by choosing to cheat to a certain extent without feeling its associated guilt.

## References

- Andreassi, J. L. (2000). *Psychophysiology: Human behavior & physiological response* (4th ed.). New Jersey: Lawrence Erlbaum Associate.
- Ariely, D. (2012). *The (honest) truth about dishonesty: How we lie to everyone—Especially ourselves*. New York: Harper.
- Ayal, S., & Gino, F. (2011). Honest rationales for dishonest behavior. *The social psychology of morality: Exploring the causes of good and evil*, 149–166.
- 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.
- Bandura, A. (1999). Moral disengagement in the preparation of inhumanities. *Personal and Social Psychology Review*, 3, 193–209.
- Barkan, R., Ayal, S., & Ariely, D. (2015). Ethical dissonance, justifications, and moral behavior. *Current Opinion in Psychology*, 6, 157–161.
- Barkan, R., Ayal, S., Gino, F., & Ariely, D. (2012). The pot calling the kettle black: Distancing response to ethical dissonance. *Journal of Experimental Psychology: General*, 141(4), 757–773.
- Barland, G. H., & Raskin, D. C. (1975). An evaluation of field techniques in detection of deception. *Psychophysiology*, 12, 321–330.
- Bond, C. F., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality & Social Psychology Review*, 10, 214–234.

- Bone J. E. & Raihani N. J. (2015) Human punishment is motivated by both a desire for revenge and a desire for equality. *Evolution and Human Behavior*, 36, 323–330.
- Brethel-Haurwitz, K. M., Stoycos, S. A., Cardinale, E. M., Huebner B., & Marsh, A. A. (2016). Is costly punishment altruistic? Exploring rejection of unfair offers in the Ultimatum Game in real-world altruists. *Scientific Reports*, 6, 18974.
- Camerer C. F. (2003). *Behavioral game theory: Experiments in strategic interaction*. Princeton, NJ: Princeton University Press.
- Chaiken, S., Giner-Sorolla, R., & Chen, S. (1996). Beyond accuracy: Defense and impression motives in heuristic and systematic information processing. In P. M. Gollwitzer & J. A. Bargh (Eds.). *The psychology of action: Linking cognition and motivation to behavior* (pp. 553–578). New York, NY: Guilford Press.
- Detert, J. R., Treviño, L. K., & Sweitzer, V. L. (2008). Moral disengagement in ethical decision making: A study of antecedents and outcomes. *Journal of Applied Psychology*, 92, 374–391.
- Engel, C. (2011). Dictator games: A meta study. *Experimental Economics*, 14, 583–610.
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G\*Power 3.1: tests for correlation and regression analyses. *Behavioral Research Methods*, 41, 1149–1160.
- Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition and cooperation. *Quarterly Journal of Economics*, 114, 817–868.
- Gino, F., Ayal, S., & Ariely, D. (2009). Contagion and differentiation in unethical behavior: The effect of one bad apple on the barrel. *Psychological Science*, 20, 393–398.
- Gino, F., Ayal, S., & Ariely, D. (2013). Self-serving altruism? The lure of unethical actions that benefit others. *Journal of Economic Behavior & Organization*, 93, 285–292.
- Gino, F., & Ariely, D. (2012). The dark side of creativity: Original thinkers can be more dishonest. *Journal of Personality and Social Psychology*, 102, 445–459.
- Gino, F., Norton, M., & Ariely, D. (2010). The counterfeit self: The deceptive costs of faking it. *Psychological Science*, 21, 712–720.
- Gino, F., & Pierce, L. (2009). Dishonesty in the name of equity. *Psychological Science*, 20, 1153–1160.
- Gneezy, A., & Ariely, D. (2010). Don't get mad get even: On consumers' revenge. *Manuscript, Duke University*.
- Gneezy, A., Imas, A., Brown, A., Nelson, L. D., & Norton, M. I. (2012). Paying to be nice: Consistency and costly prosocial behavior. *Management Science*, 58(1), 179–187.
- Gneezy, U., Imas, A., & Madarász, K. (2014). Conscience accounting: Emotion dynamics and social behavior. *Management Science*, 60(11), 2645–2658.
- Greenberg, J. (1990). Employee theft as a reaction to underpayment inequity: The hidden cost of pay cuts. *Journal of Applied Psychology*, 75, 561–568.
- Greenwald, A. G. (1980). The totalitarian ego: Fabrication and revision of personal history. *American Psychologist*, 35, 603–618.

- Güth, W., & Kocher, M. G. (2014). More than thirty years of ultimatum bargaining experiments: motives, variations, and a survey of the recent literature. *Journal of Economic Behavior & Organization*, 108, 396–409.
- Güth, W., Schmittberger, R., & Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior & Organization*, 3, 367–388.
- Hochman, G., Glöckner, A., Fiedler, S., & Ayal, S. (2016). “I can see it in your eyes”: Biased processing and increased arousal in dishonest responses. *Journal of Behavioral Decision Making*, 29(2–3), 322–335.
- Hochman, G., Peleg, D., Ariely, D., Ayal, S., (2021). Robin Hood meets Pinocchio: Justifications increase cheating behavior but decrease physiological tension. *Journal of Behavioral and Experimental Economics*, 92, 101699.
- Honts, C. R. (2004). The physiological detection of deception. In P. A. Granhag & L. A. Stromwall (Eds.). *Deception detection in forensic contexts* (pp. 103–123). Cambridge: Cambridge Press.
- Honts, C., & Amato, S. (2007). Automation of a screening polygraph test increases accuracy. *Psychology, Crime, & Law*, 13, 187–199.
- Honts, C. R., & Reavy, R. (2015). The comparison question polygraph test: A contrast of methods and scoring. *Physiology and Behavior*, 143, 15–26.
- Houser, D., Vetter, S., & Winter, J. (2012). Fairness and cheating. *European Economic Review*, 56, 1645–1655.
- Jones, S. C. (1973). Self and interpersonal evaluations: Esteem theories versus consistency theories. *Psychological Bulletin*, 47(4), 242–245.
- Jones, T. M. (1995). Instrumental stakeholder theory: A synthesis of ethics and economics. *Academy of Management Review*, 20, 404–437.
- Jordan, J., Mullen, E., & Murnighan, J. K. (2011). Striving for the moral self: The effects of recalling past moral actions on future moral behavior. *Personality and Social Psychology Bulletin*, 37, 701–713.
- Kircher, J. C., & Raskin, D. C. (1988). Human versus computerized evaluation of polygraph data in a laboratory setting. *Journal of Applied Psychology*, 73, 291–302.
- Kircher, J.C., & Raskin, D.C. (2002). Computer methods for the psychophysiological detection of deception. Chapter 11, pp. 287–326. In *Handbook of Polygraph Testing*. M. Kleiner, ed. London: Academic Press.
- Koning, L., E. Van Dijk, I. Van Beest and W. Steinel (2010). An instrumental account of deception and reactions to deceit in bargaining. *Business Ethics Quarterly*, 20, 57–73.
- Krapohl, D. (2013) Polygraph principals: A literature review. *Polygraph*, 42(1), 35–60.
- 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.
- Mazar, N., & Zhong, C. B. (2010). Do green products make us better people?. *Psychological Science*, 21, 494–498.
- Merritt, A. C., Effron, D. A., & Monin, B. (2010). Moral self-licensing: When being good

- frees us to be bad. *Social and personality psychology compass*, 4(5), 344–357.
- Mitkidis, P., Ayal, S., Shalvi, S., Heimann, K., Levy, G., Kyselo, M., Wallot, S., Ariely, D., & Roepstorff, A. (2017). The effect of extreme rituals on moral behavior: The performers-observers gap hypotheses. *Journal of Economic Psychology*, 59, 1–7.
- Monin, B., & Miller, D. T. (2001). Moral credentials and the expression of prejudice. *Journal of Personality and Social Psychology*, 81(1), 33–43.
- Motro, D., Ordonez, L., Pittarello, A., & Welsh, D. (2016). Investigating the Effects of Anger and Guilt on Unethical Behavior: A Dual-Process Approach. *Journal of Business Ethics*, 152, 133–148.
- Novoa, M., Malagon, M., & Kraphol, D. (2017). Attitudes of Polygraph Examiners and Examinees. *Polygraph*, 46 (2), 172–186.
- Ordóñez, L. D., Schweitzer, M. E., Galinsky, A. D., & Bazerman, M. H. (2009). Goals gone wild: The systematic side effects of overprescribing goal setting. *Academy of Management Perspectives*, 23, 6–16.
- Park, E. S., Levine, T. R., Harms, C. M., & Ferrara, M. H. (2002). Group and individual accuracy in deception detection. *Communication Research Reports*, 19, 99–106.
- Patrick, C. J. (2011). Science on the rise: Birth and development of the Concealed Information Test. *Memory Detection: Theory and Application of the Concealed Information Test*. Cambridge University Press, New York, New York, 3–11.
- Peer, E., Acquisti, A., & Shalvi, S. (2014). “I cheated, but only a little”: Partial confessions to unethical behavior. *Journal of personality and social psychology*, 106, 202–217.
- Peleg, D., Ayal, S., Ariely, D., & Hochman, G., (2019). The Lie Deflator – The effect of polygraph test feedback on subsequent (dis)honesty. *Judgment and Decision Making*, 14(5), 728–738.
- Pillutla MM, & Murnighan JK. (1996). Unfairness, anger, and spite: emotional rejections of ultimatum offers. *Organizational Behavior & Human Decision Processes*, 68, 208–224.
- Raskin, D. C. (1986). The polygraph in 1986: Scientific, professional and legal issues surrounding application and acceptance of polygraph evidence. *Utah Law Review*, 29, 29–74.
- Raskin, D. C., & Honts, C. R. (2002). The comparison question test. In M. Kleiner (Ed.). *Handbook of polygraph testing* (pp. 1–47). San Diego, CA: Academic Press.
- Raskin, D. C., Honts, C. R., & Kircher, J. C. (2014). *Credibility assessment: scientific research and applications*. Amsterdam: Academic Press.
- Rosenberg, M. (1979). *Conceiving the self*. New York. Basic Books.
- Sachdeva, S., Iliev, R., & Medin, D. L. (2009). Sinning saints and saintly sinners the paradox of moral self-regulation. *Psychological Science*, 20, 523–528.
- Sanitosa, R., Kunda, Z., & Fong, J. T. (1990). Motivated recruitment of autobiographical memories. *Journal of Personality and Social Psychology*, 59, 229–241.
- Schurr, A. and Ritov, I. (2016) Winning a competition predicts dishonest behavior. *Proceedings of the National Academy of Sciences*, 113(7), 1754–1759.

- 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, 287–301.
- Shalvi, S., Dana, J., Handgraaf, M. J., & Dreu, C. K. (2011). Justified ethicality: Observing desired counterfactuals modifies ethical perceptions and behavior. *Organizational Behavior and Human Decision Processes*, 115(2), 181–190.
- Shalvi, S., Gino, F., Barkan, R., & Ayal, S. (2015). Self-Serving justifications: Doing wrong and feeling moral. *Current Directions in Psychological Science*, 24(2), 125–130.
- Tetlock, P. E., Kristel, O., Elson, B., Green, M., & Lerner, J. (2000). The psychology of the unthinkable: Taboo trade-offs, forbidden base rates, and heretical counterfactuals. *Journal of Personality and Social Psychology*, 78, 853–870.
- Thielmann, I., & Hilbig, B. E. (2019). No gain without pain: The psychological costs of dishonesty. *Journal of Economic Psychology*, 71, 126–137.
- Van der Zee, S., Anderson, R., & Poppe, R. (2016). When lying feels the right thing to do. *Frontiers in Psychology*, 7, 734.
- Vrij, A. (2000). *Detecting lies and deceit: The psychology of lying and the implications for professional practice*. Chichester, UK: Wiley.
- Vrij, A., & Fisher, R. (2016). Which lie detection tools are ready for use in the criminal justice system? *Journal of Research in Memory and Cognition*, 5, 302–307.
- Wang, J. T-Y., Spezio, M., & Camerer, C. F. (2010). Pinocchio's pupil: Using eye tracking and pupil dilation to understand truth telling and deception in sender-receiver games. *American Economic Review*, 100(3), 984–1007.

## Appendix

Questions on the CQT (Raskin, 1986) in Studies 1 and 2.

### Test questions

- 1 - Did you cheat on the task you just completed in this study?
- 2 - In the previous task, did you select the side that would give higher payoffs regardless of whether it was the side with more dots?

### Control questions

- 1 - Is there a door in this room?
- 2 - Have you ever lied to avoid taking responsibility for your actions?
- 3 - Have you ever acquired something in a dishonest way?
- 4 - Is there a chair in this room?
- 5 - Have you ever cheated out of self-interest?
- 6 - Are we currently in Tel Aviv, Israel?