Lying to Punish

Perceived injustice in the workplace may increase the frequency of deceptive behaviors. We experimentally investigate how the kindness of an interaction between individuals affects subsequent tendency to engage in deception. In a series of two-player, two-stage games, we study punishment through deception. The first-stage varies the kindness of an encounter between two players. The second-stage gives one player the opportunity to lie in order to punish the other based on behavior in stage one. We focus on two types of deception such that lying to others is either beneficial or costly to the person who deceives from a monetary point of view. Our results show the importance of reciprocity in lying behavior: when deception is beneficial, lying rates are high but kind encounters are rewarded by honesty. Further, when lying is costly individuals use deception as a punishment device after an unkind encounter. These results are driven by intentions rather than by inequalities in payoffs.

1. Introduction

Empirical studies suggest that workers react to perceived injustice in the workplace by engaging in what is commonly summarized as 'counterproductive workplace behavior' (Greenberg, 1990; Hollinger and Clark, 1983; Krueger and Mas, 2004; Lee & Rupp, 2007; Mas, 2006, 2008). Labor disputes, negative relations within organizations or pay cuts have been shown to decrease workers' effort provision and motivation, resulting in lower quality production. For example, Krueger and Mas (2004) found that during the 1994-1996 conflict between a labor union and the management of Bridgestone/Firestone, the company produced a large amount of defective tires. Similarly, compensation disputes in police departments in New Jersey (Mas, 2006) have been shown to coincide with a decline in the number of arrests by police officers, and with an increase in crime rates.

Although workers might sometimes react to unfair treatment with a direct decrease in effort provision (i.e., by working less), some other times they might engage in alternative forms of retaliation, with similar negative effects on productivity and production quality. In particular, in response to perceived injustice workers might engage in unethical behaviors such as misreporting the quality of their work, over-reporting their work hours, or engaging in outright sabotage when their actions are not observable. Such counterproductive behaviors often involve deception—an act or statement intended to make people believe something that is not true. Since deception is by its nature difficult to detect, empirical work offers limited guidance in answering whether employees adopt deceptive behaviors as a form of retaliation. In this paper, we experimentally study deception that result from perceptions of being treated unfairly[1].

The literature in experimental economics has mostly focused on forms of punishment that can be directly employed without deception (see for example the literature on the ultimatum game by Güth et al, 1982; and the survey of reciprocal behavior by Fehr and Gächter, 2000). In many workplace situations however, direct punishment is not always available due to the fear of retribution. In such situations unethical behavior, such as deception, may serve as an alternative punishment device. However, it is not clear whether the results observed in the punishment literature would also be observed when punishment involves deception. In contrast to standard punishing tools (i.e., rejecting an offer in the ultimatum bargaining), punishing with deception may induce psychological costs. Notably, a growing experimental literature shows that some people are averse to engaging in

[1] As frequently pointed out by the popular press, unethical behavior by both employees and executives is widespread in the corporate world (see Trevino et al. 2014, for a review of unethical behavior in organizations). While previous research has investigated some of the factors that affect unethical conduct in the workplace such as, for example, incentives schemes (Belot and Schroeder, 2013; Danilov et al. 2013), relative wages (John, Loewenstein and Rick, 2014), or the effects of monitoring on other forms of counterproductive behavior (Gino, Krupka and Weber, 2013; Belot and Schroeder, 2013), it is still not clear whether perceptions of being treated unfairly increase the frequency of deceptive behavior as a form of direct punishment.

unethical behaviors due to lying or guilt aversion (Gneezy, 2005; Erat and Gneezy, 2012; Sutter, 2009; Dreber and Johannesson, 2008; López-Pérez and Spiegelman, 2013; Lundquist et al., 2009; Battigalli, Charness and Dufwenberg, 2013; Battigalli and Dufwenberg, 2007; Charness and Dufwenberg, 2006; 2010; Greenberg et al., 2014). Nonetheless, this might not be the case when deception is used to reciprocate unkind behavior, as punishing via deception in such situations might not be considered unethical.

In this paper, we experimentally investigate whether individuals use deception as a form of punishment after a negative encounter with another person, and whether deception decreases to reward a positive interaction. We focus on two types of deception. First, we look at deception that benefits the deceiver at a cost to the receiver. Such situations could happen, for example, when a worker over-reports own performance to earn more, at a cost to a colleague or the company. Second, we study deception that involves a monetary cost both for the person who deceives and her counterpart. Sabotaging the success of a team project by misreporting crucial private information constitutes an example of this kind of deception. We also study whether both kinds of deceptions are triggered by intentions—the (un)kindness of subjects' behavior—or by distributional concerns (i.e. differences in payoffs).

While the base rates of these two types of deception might differ (with higher deception rates when deception is beneficial than when it is costly for the deceiver), we expect individuals' tendency to engage in both types of deception to be sensitive to a previous encounter. We know from the extensive experimental literature on punishment that when subjects think that they were treated unfairly, they are willing to punish, even when punishment is costly for them. In particular, when punishing unfair outcomes, intentions matter (see e.g. Blount, 1995; Charness, 2004; Charness and Haruvy, 2002; Charness and Levine, 2007; Rand et al. 2014; Sutter, 2007; see also the survey in Charness and Khun, 2011; see Offerman, 2002 for an example of the importance of intentions in reciprocal behavior and Dufwenberg and Kirchsteiger, 2000 on the effect of reciprocity in workers' motivation) and in many cases they matter more than payoff inequalities (see e.g. Charness and Kuhn, 2007; Garofalo and Rott, 2014[2]). However, it is not clear whether deception toward others is more likely to be triggered by intentions or payoff asymmetries.

We explore these research questions in a series of two-player, two-stage games. In the first study, subjects first play a dictator game. Then, the receiver of the dictator game is given the opportunity to tell a "Selfish Black Lie" (in the Erat and Gneezy, 2012 taxonomy) to her counterpart. Using different variations of this experiment we test whether deception is used to reciprocate unkind

[2] Garofalo and Rott (2014) find that when individuals receive an unfair allocation compared to two other people and have the option to punish them, they punish at a greater degree the decision-maker (the person who decides how to allocate money across all the subjects) than the spokesperson (the person who communicate the decision but does not make any decision) even if the payoffs for both the decision-maker and the spokesperson are the same.

or kind behavior. In a second set of studies, we first use the same design (Study 2a) to look at the effect of the dictator's generosity on subsequent costly deception (telling a "Spiteful Black Lie" in the Erat and Gneezy, 2012 taxonomy). Then, we further disentangle the drivers of costly deception (intentions versus payoff inequality) in a different two-stage, two-player game (Study 2b). In the first stage of this game, each player in a pair chooses between two actions, and the pair's decisions jointly lead to a payoff outcome. In a 2x2 design, we vary whether the unique Nash equilibrium outcome of this game results in equal or unequal payoffs and whether the action chosen by one player is neutral or (intentionally) unkind from the perspective of the other player. In a second stage in which individuals are given an opportunity to tell a spiteful lie to their counterpart, we study how these factors affect costly punishment.

Our results show that when deception benefits the deceiver at a cost to the receiver, overall deception rates are rather high but kind actions are rewarded by less lying; lying monotonically decreases with the kindness of the previous interaction. When deception is costly, we find no spiteful lying after a neutral or kind encounter but we do find substantial deception after an unkind interaction. Further, for both types of deception intentions rather than payoff inequality explain our results.

To our knowledge, the only studies in economics that investigate deception after an initial encounter are by Ellingsen et al. (2009) and Houser et al. (2012). Similarly to our first study, Ellingsen et al. study deception that mostly benefits the deceiver and hurts the other party ("Selfish Black Lies"). They find that the propensity to be honest in bilateral bargaining is strongly affected by the prior behavior of one's counterparts. Specifically, they find that subjects do not tell a selfish black lie after a positive encounter; rather, they find that lying for one's own benefit is usually observed after a negative encounter. In their study, different than in ours, the benefit of lying depends on the valuations in the bargaining game and is not constant[3]. Houser et al. (2012) find that subjects who believe that they were treated unfairly in an interaction with another person are more likely to cheat in a subsequent unrelated decision that does not involve the other party. Our first study expands this literature as it investigates deceptive behavior as a form of direct punishment or reward mechanism when lying benefits the deceiver at the expenses of the receiver. The second set of studies expands this literature by focusing on costly punishment via deception, which to our knowledge has not been previously investigated.

[3] In their setup it is possible to lie and also hurt oneself by reducing the trade possibility, but their design does not allow for a direct inference on how often such lies happen.

2. Selfish Black Lies As a Punishment Tool

2.1 Study 1

Procedures

The experiment was conducted at Northern American University. A total of 332 individuals participated in three different variations of the experiment (216, 56, and 71 participants in treatment 1, 2, and 3, respectively[4]). The experiment was conducted online using the recruitment system of the university. The matching protocol was as follows: In the first variation of the study, each participant was randomly assigned to one of the two roles in the experiment (Player A or Player B) and was matched with another participant playing the other role. In the second and third treatments, at most one in ten B players were matched to one player A[5]. At the beginning of the experiment, subjects were informed about their type (A or B), and were told that the experiment had two stages with the same matching of participants. Subjects learned that the final payoff was determined by the sum of the payoffs obtained in both stages of the experiment, and that one out of 10 participants would be selected for payment after the completion of the experiment. Participants selected to get paid were subsequently invited to the lab and paid in person. Average earnings for the participants[6] selected to get paid were 20.1 dollars. After the experiment, participants were asked to fill out a short questionnaire indicating their background information such as gender and age. Full instructions of the study are reported in Appendix A1. A detailed description of the experiment follows.

Tasks and Hypothesis

In the first stage, Player A (dictator) is given an endowment of $10 and has to decide how much to send to Player B (receiver). Player A can choose between sending $0, $2, $4, $6, $8, or $10 to Player B. Player A keeps any amount that is not sent to Player B. The amount sent to Player B is tripled. Player B does not make any decision in the first stage. In the second stage of the experiment, both participants play the deception game of Erat and Gneezy (2012). A six-sided die is rolled before the beginning of the game and its outcome is communicated to Player B. Afterwards, Player B is asked to send one of the six possible messages to Player A. The six messages are: "The outcome of

[4] In the first treatment we used a between subjects design, matching each Player A to a Player B. To have enough variation in the amount sent in stage 1, we aimed at having at least 100 subjects per role. In the second treatment we used a within subjects design. In the second and third treatment we matched 1 in 10 Player B with a Player A.

[5] In particular, in treatment 2 and 3 we randomly selected one every ten participants and matched them with a Player B. Since the total number of Players A in the two treatments was not a multiple of ten (51 people in treatment A and 65 people in treatment B), we rounded it and randomly matched 5 and 6 Players A (T2 and T3 respectively) to a Player B.

[6] In treatment 1, one out of 10 participants (22 subjects) who completed the experiment were paid.

the roll of die was x", where x can be any number from {1,2,3,4,5,6}. After receiving Player B's message, Player A declares what she believes the outcome of the die roll to be. Player A's choice and the real die outcome determine the payoffs for both players. There are two payment options, Option 1 and Option 2. Both players are informed that if Player A's choice coincides with the real outcome, payment option 1 is implemented. Otherwise, option 2 is implemented. Player A is never informed about the possible payoffs associated with the two options; she only learns her own earnings from the task. Player B knows the payoffs associated with options 1 and 2 for both players, and also knows that Player A does not know the payoffs. The payoffs are as follows. In option 1, if Player A chooses the number that is the actual outcome of the die roll, then Player A earns $20 while Player B earns $10. In option 2, if Player A chooses a number different than the actual outcome, Player A earns $10 and Player B earns $20. These payoffs make clear that option 2 is worse for Player B. Thus, assuming that Player A has no reason not to follow the recommendation she receives, Player B has an incentive to send an untruthful message to increase her own payoff at the expenses of the other player. In this paper, we refer to all messages that contain a number different than the real outcome of the die roll as deceptive.

We conducted three variations of this design. In the first treatment, each Player B is matched with a Player A, according to a between-subjects design. Player B is first informed about how much Player A decided to send her in the first stage of the game and is then asked to make a decision in the second stage of the game. In the second treatment, we replicate the study using the strategy method (*within* subjects design). In particular, Player B is asked to choose which message to send to Player A conditional on all the possible first stage choices of Player A. Whereas the first treatment establishes the general trend, the second allows us to analyze how the individual lying decisions differ across different first-stage scenarios, increasing the power of our statistical analysis.

In the third treatment we vary who determines Player B's payment in the first stage of the game. In particular, we analyze Player B decision to lie to Player A when the amount received in the first stage is not determined by Player A, but rather comes in the form of a show up fee from the experimenter.

We hypothesize that Player B's decision to lie will vary as a function of Player A's choice in the first stage. In particular, we expect lying to decrease with the amount sent by Player A. When the money is received from the experimenter rather than from another player, we expect the amount received not to affect people's tendency to deceive their counterpart. These hypotheses are summarized below:

Hypothesis 1: In treatment 1 and 2, lying decreases with the amount received from the other player in the first stage.

Hypothesis 2: In treatment 3, lying does not change with the amount received from the experimenter in the first stage.

Results

The results of the first and second treatment are depicted in Table 1. Player A sent an average of $3.60 in the first treatment. We use the amount sent by Player A as an indicator of how (un)kind the encounter between Player A and Player B was. As shown in the table, when the experiment was conducted between subjects, a substantial fraction of participants (77%) sent 0, 2 or 4. Notice that the remaining 23.3% of Players A sent more than half of the pie; much larger than what is usually observed in dictator games (see Camerer, 2003). This is probably due to the fact that participants were informed that there were two stages in the experiment, and they wanted to avoid punishment in the second stage.

Outcome of Stage 1 and Stage 2

	Money received	**0**	**6**	**12**	**18**	**24**	**30**
Treatment 1: Between Subjects	**# Offers**	17	41	26	7	3	14
	# Lying	13	26	16	3	1	5
	Percentage of lying	76.5	63.4	61.5	42.9	33.3	35.7
Treatment 2: Strategy method	**# Lying (out of 51)**	36	38	34	27	23	18
	Percentage of lying	70.6	74.5	66.7	52.9	45.1	35.3

Table 1. Outcomes of Stage 1 and Stage 2 in Treatment 1 and 2.

The fraction of Players B who lied after receiving any of the six possible amounts in the first two treatments is depicted in Figure 1. When we look at lying behavior, we see that in both treatments there is substantial lying when participants receive low amounts[7].

[7] In the second treatment, there are three people who chose not to lie when 0 is sent and lied in all other possibilities suggesting an error in their first decision. This is why the percentage of lying increases somewhat from 0 to 2. The rest of the responders display the expected pattern.

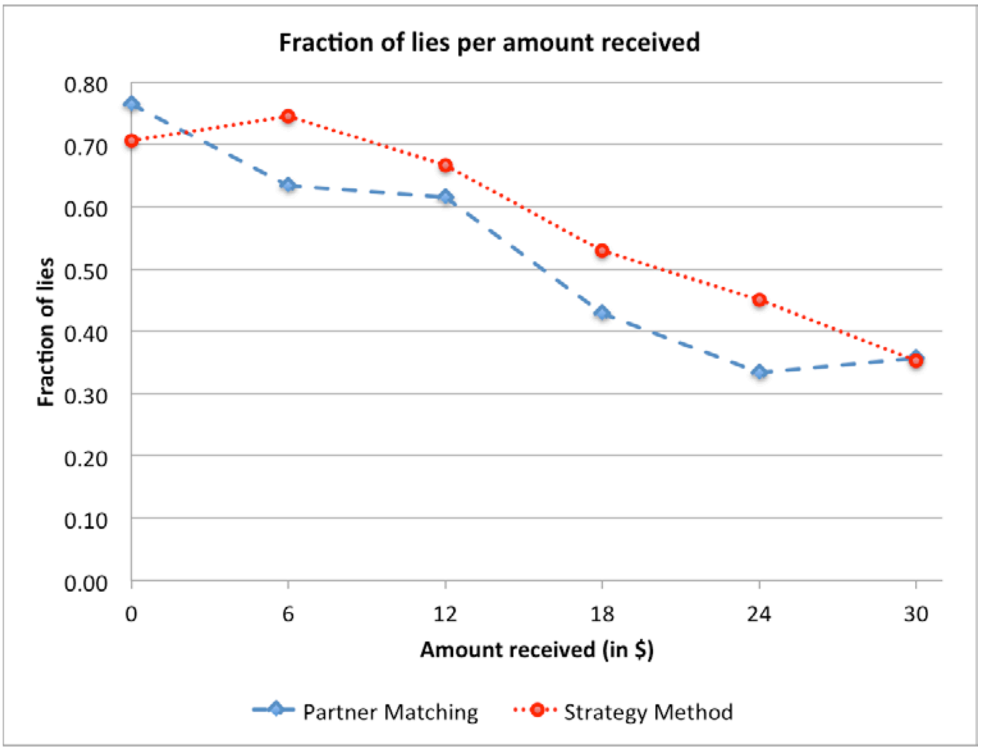

Figure 1. Fraction of lies per amount received in treatments 1 and 2

The comparison of lying rates when players receive $0 to lying rates when they receive $10 confirms that the percentage of lies is significantly lower in the latter case in both treatments (for treatment 1-between subjects, p=.026, one-sided Fisher exact test; for treatment 2-strategy method, p<.001, McNemar). Furthermore, a probit regression using lying as a dependent variable and the amount given by Player A as an independent variable confirms that in both treatments, lying rates decrease with the amount sent. The estimated marginal effects show that for each additional dollar received, the probability of lying drops by 4 percentage points. The regression tables can be found in Appendix A2.

Therefore we find support for our first hypothesis, and show that lying monotonically decreases with the amount sent in the first stage. The results obtained using the strategy method are in line with those observed using a between subjects design. When we compare lying rates per amount sent in the two treatments, none of the percentages is significantly different from each other. Considering the amount sent as a measure of the kindness of Players A' action, these results suggest that the kinder the actions, the lower the likelihood of lying in a subsequent interaction. However, these two treatments do not disentangle whether the relationship between lying rates and the amount received is due to the sender's intention or just to Player A's increase in payoffs when she receives positive amounts. To further isolate the role of intentions in punishment via deception, we study Player B's behavior when Player A does not determine the Player B's payoff in the first stage.

The results of the third treatment are depicted in Table 2 and Figure 2. The average lying rate is 63%. Importantly, lying rates in this treatment do not seem to depend on the amount Players B earned: they are quite high regardless of the amount received in Stage 1. Among all Players B who received $0 from the experimenter, 54.5% lied, whereas among those who received $10 from the experimenter, exactly the same percentage lied in the second stage. A probit regression confirms that lying is independent of the amount received. This confirms hypothesis 2(a), showing that individuals' tendency to lie toward another subject is not affected by the initial payoffs.

Outcome of Stage 1 and Stage 2

	Money Received	**0**	**6**	**12**	**18**	**24**	**30**
Treatment 3: No intentions	**# Lying (N)**	6 (11)	10 (13)	7 (11)	10 (12)	2 (7)	6 (11)
	Percentage of lying	54.5	76.9	63.6	83.3	28.6	54.5

Table 2. Outcome of Stage 1 and Stage 2 in Treatment 3.

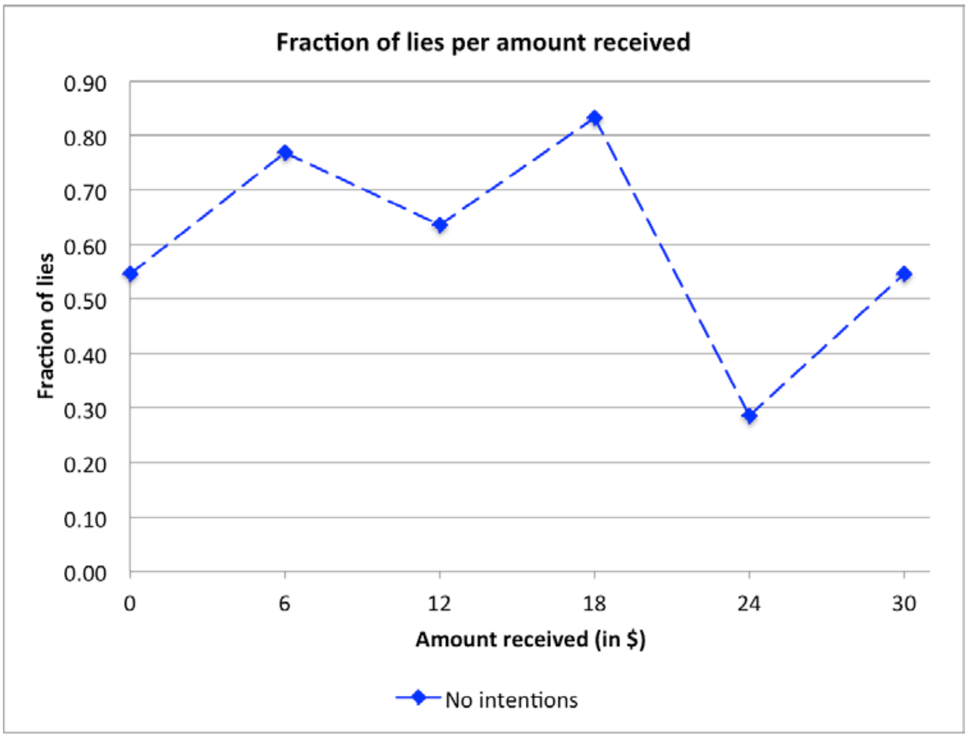

Figure 2. Fraction of lies per amount received in the third (No intentions) treatment.

Finally, we compare the lying rates in the third treatment with the ones observed in the previous treatments. If Player B engages in deception to punish unkind actions, then we should observe an increase in lying rates as a response to unkind behavior in the first stage compared to the case in which subjects receive money from the experimenter. Hence, since it is unkind to send 0, or 2 dollars in the first stage of the game[8], we expect higher lying rates when Player A sends these amounts than when 0 or 2 dollars are given by the experimenter. Since it is kind to send 6, 8, or 10 dollars, we expect lower lying rates when these amounts are received from Player A than when they are received from the experimenter.

Our data confirm the direction of the effect for kind behavior, but not for unkind behavior. If we combine observations of B players earning 18, 24, and 30, we see that 60% (18 out of 30) of them lie in treatment 3, whereas 37.5% (9 out of 24) lie in treatment 1 (p=.085, one-sided Fisher exact). If we combine observations of Players B earning 0 or 6, however, we do not see an effect of intentions: 66.7% (16 out of 24) lie in treatment 3 whereas 67.2% (39 out of 58) lie in treatment 1 (p=.577, one sided Fisher exact). Therefore, we do not find evidence for punishment of unkind actions, but we do find that kind actions are rewarded by less lying when lying benefits the deceiver and hurts the receiver. This result could be due to the fact that people are unwilling to use deception as a punishment device. However, the overall high lying rates we observe in our sample might not have allowed us to detect an increase in lying after an un(kind) encounter. If (as previous studies suggest) a minority of subjects never lies regardless of the consequences, the upper bound on the frequency of lies would be smaller than 100 percent. Thus, a 67 percent deception rate in treatment 3 makes it unlikely to detect an upward effect of unkind actions. Therefore, it is possible that a proportion of people was willing to punish unkind actions but since a large fraction of subjects already lies for personal benefit, the effect would not be observed in our data.

Finally, we explore Player's A behavior in treatment 1 after receiving the recommendation from Player B. Overall, the fraction of Players A who followed the recommendation is about 70%, suggesting that lying in this context had negative consequences for Player A. This is in line with

[8] In the second study, we will employ Rabin's (1993) categorization of kind/unkind actions, however, in the first study we will not. The reason is, Rabin's categorization is meaningful if both players have a non-empty choice set, and is therefore suited to (and developed for) 2x2 games. If we were to stick to Rabin's categorization in the first study as well, we should classify sending 0, 2, and 4 as unkind, and sending 6, 8, and 10 as kind actions. The rationale is as follows: given a set of actions, kindness of an action is measured as per the deviation it creates from the average Pareto efficient payoff of the other player. Since the average Pareto efficient payoff for Player B is 15 from the first stage, any action by Player A that gives Player B an amount above 15 is kind whereas any amount below 15 is unkind. We, however, think that categorizing sending 4 as unkind is not meaningful since sending 4 actually gives Player B more than what remains for Player A. The categorization problem arises from the fact that Player B does not have a choice to make, and by arbitrarily changing the choice set of Player A, we can manipulate what constitutes kind or unkind behavior. As an extreme example, assume that Player A has 10 units and his choice set is giving 10, 12, 14, 16, 18, and 20 to player B. Clearly, none of these are unkind, whereas we should classify 10, 12, and 14 as unkind if we were to strictly employ the rationale behind the kindness measure introduced by Rabin. More information on Rabin's theory is provided in the next section.

previous literature (Gneezy, 2005; Hurkens and Kartik, 2009). Further, we do not find Player's A propensity to follow the recommendation to depend on the amount sent to Player B on stage 1 (β=-.051, p=.27, probit regression).

To sum, the results of the first study show a relationship between amount received and lying rates that seems to be independent of the amount received in Stage1 but rather due to Player A's intentions. Further, while deception does not increase when the previous encounter is unkind, kind behavior is rewarded with significantly less deception. In the next section we investigate whether deception is used as a punishment tool in situations where lying is costly for the deceiver.

3. Spiteful Black Lies As a Punishment Tool

In two experiments, we investigate the effect of the type of an initial encounter on subsequent deception, when deception is costly for both the deceiver and the receiver. In Study 2a, we employ the same design we used in Study 1 and, using the strategy method, we investigate the effects of the dictator's generosity on costly deception. In Study 2b, we adopt a different design to look into the drivers of costly deception by manipulating the equality of payoffs as well as the type of encounter (neutral or unkind) between players.

Procedures

Both studies were conducted in two large libraries of European university. We conducted half of the sessions in each library. Subjects were randomly approached and recruited from the library areas[9]. The experiments were implemented in paper and pencil format. In each experimental session there were 12 participants, six Players A and six Players B, who were divided according to player type into two adjacent rooms. Participants could verify that there was another room with the other players. In both experiments each Player A was matched with another Player B. Subjects learned their type (A or B) and were informed that all participants in the same room had the same type. They were also informed about the payment procedure that would follow at the end of the experiment. Subjects were asked to fill out a questionnaire asking for their background information and for some game specific questions. After the end of each session one Player A and one Player B were randomly selected to be paid privately and in cash according to their earnings during the experiment. The instructions of the studies are reported in Appendices A1 and B1. A detailed description of the specific procedures follows.

[9] We approached anyone who was in the library at the times of the experiment. It is possible that people who are willing to spare their time for an experiment have different characteristics than the ones who are not, but such a difference should have no effect on the comparative statics that are informative for our research questions.

3.1. Study 2a

Study 2a implements the same design of Study 1 using costly deception. Participants were matched for two-stages and played a dictator game followed by the same sender-receiver game of Study 1. Payoffs for stage 2 were determined as follows: In option 1 both players earn 10 euros, whereas in option 2 the sender earns 8 euros and the receiver earns 4. Thus, option 2 is welfare decreasing and costly for both players. Assuming that the receiver (Player A) has no reason not to follow the sender's (Player B) recommendation, Player B should have no incentive to send an untruthful message. However, if the interaction with Player A is unkind, Player B may be willing to sacrifice some money and send an untruthful message in order to decrease Player A's payoff.

This design allows us to test whether spiteful lying decreases with the amount received in the first stage. In total 72 subjects participated in 6 different sessions. The experiment was conducted using the strategy method. That is, each Player B was asked to choose which message to send to Player A conditional on all the possible first stage choices of Player A. Average earnings for the participants that were selected to be paid were 16.7 Euros. Each session of the experiment lasted approximately 15 minutes.

Results

The results are depicted in Table 3 and Figure 3. As displayed in the figure, lying rates monotonically decrease as a function of the amount received in the dictator game. Although lying rates are lower than in Study 1, the results are similar to the pattern we observed when deception benefits the deceiver. When the amount sent is 0, the lying rate is 42 percent, and it drops rapidly to 3 percent with increasingly higher transfers. A McNemar test confirms that lying is significantly higher when the offer is 0 or 2 compared to when the offer is 6, 8, or 10 (p-values for 0, 2, and 4 Euro offers are <0.001, 0.008, and 0.25, respectively). Finally, as expected, deception rates are much smaller than in the first study and statistically significant in all pairwise comparisons[10]. A probit regression confirms this result. The estimated marginal effects show that for each additional dollar received, the probability of lying drops by 3.5 percentage points (see the table in Appendix A2).

We also look at Player A's behavior after receiving the recommendation. We find that 72% of Players A choose to follow Player B's message. Thus, most of Players A rely on the advice from the better-informed party.

[10] The relevant one-sided Fisher exact test p-values are less than 0.001 for all the comparison of lying rates when 0, 2, 4, 6, 8, or 10 is sent.

	Outcome of Stage 2						
	Money received	0	6	12	18	24	30
Study 2a:	# Lying (out of 36)	15	9	4	1	1	1
Strategy method	Percentage of lying	41.7	25.0	11.1	2.8	2.8	2.8

Table 3. Outcomes of Stage 2 in Study 2a.

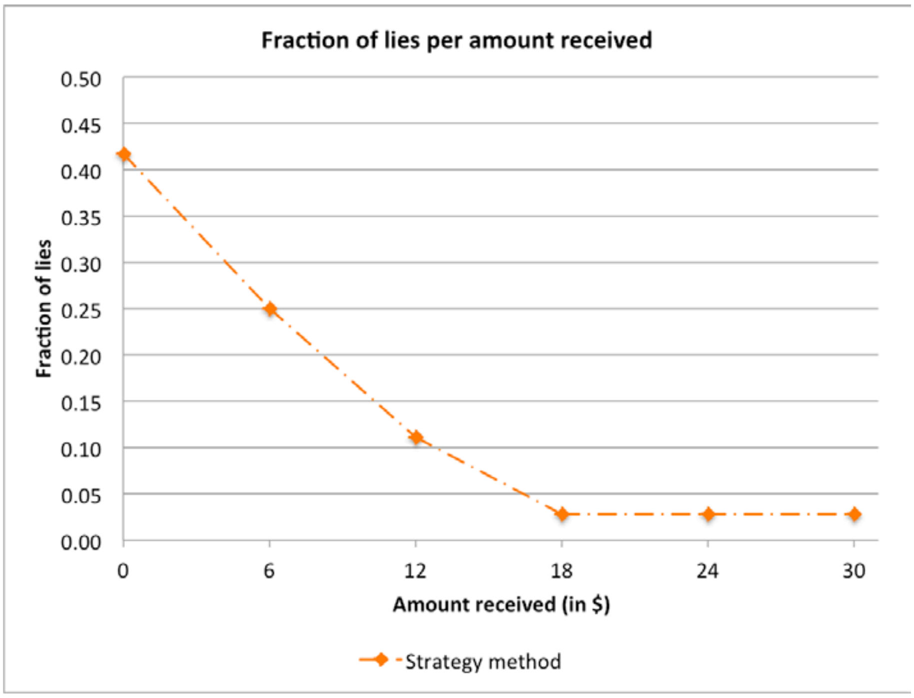

Figure 3. Fraction of lies per amount received in Study 2a

3.3. Study 2b

The results of study 2a show that people react to unfair offers by lying even if it is costly for them. To tease apart the effect of intentions from the inequality in payoffs, we run an additional study. In this study, we replace the initial dictator game with a new first stage game, while keeping the deception game and payoffs the same as in study 2a. In this game, both players choose an action and their joined decisions determine the payoffs for both of them. In a 2X2 design, we vary the Nash equilibrium outcome in a way that allows us to capture all combinations of equal/unequal payoffs and neutral/unkind actions. This enables us to determine the effect of unequal payoffs, and separate it from unkind actions. We can further test whether there is an interaction effect, i.e., whether unkind actions only matter when payoffs are unequal.

Procedures

A total of 144 subjects participated in 12 different sessions. Different than in the previous studies, subjects are not informed that there is a second stage[11,12]. At the beginning of the second part, subjects are informed about the next stage and are told that they are matched with the same partners as in the first stage. Payoffs are in tokens and each token is exchanged at a rate of 0.5 Euros for 1 token. Average earnings for the participants that were selected to get paid were 7.5 Euros. Each session of the experiment lasted approximately 30 minutes.

Tasks and Hypotheses

For all players, the first task involves choosing between two actions, R and M. After this task, players are informed about the action chosen by their matched partner. The joint decisions of the two players result determine the payoffs for both. We vary the Nash equilibrium outcome of the game across treatments in order to create equality or inequality in payoffs and neutral or unkind choices by Player A, according to a 2X2 between subjects design. We have four treatments: equal-neutral, equal-unkind, unequal-neutral and unequal-unkind. Importantly for our design, the Nash equilibrium of each game coincides with the Nash equilibrium of the inequality averse preferences modeled by Fehr and Schmidt (1999) using the commonly accepted parameter values[13] (All calculations are in Appendix B3.) In our setup, only the perception of Player B is relevant because only Player B can react to the outcome of the first task. The neutral and unkind actions are based on Rabin's (1993) intentions-based model. Rabin's kindness function measures whether the other player's action leads to a payoff outcome that is more or less than one's equitable payoff. A player's equitable payoff is defined as the average of the highest possible payoff given her action and the lowest payoff among Pareto efficient outcomes. Considering the action vector of two Players A and B, if the kindness function of Player B is positive, Player A is said to be giving to Player B more than Player B's equitable payoff and Player A is considered to be kind. Similarly, a negative kindness function for one player indicates unkind actions by the other party. Otherwise, if the kindness function is zero, Player A is giving Player B exactly her equitable payoff.

[11] In this study, the labelling of the Players was inverted. We kept the labelling consistent in the paper to make for easier reading.
[12] In the first study, subjects seem to make more kind offers due to the knowledge of further play. To make sure to have enough observations from the first stage of the game, we did not inform the subjects that they will play a second stage with the same players. It is possible, however, that subjects inferred from the length of the study that there is a second stage.
[13] See for example Yang, Onderstal, Schram (2014) for an estimation of the inequality aversion parameters using the model of Fehr and Schmidt (1999).

In our game, we follow Rabin’s categorization, and consider a neutral action as having a value of zero in the kindness function, and an unkind action as having a negative value. Finally, in each game, the pure-strategy “fairness equilibrium” proposed by Rabin coincides with the Nash equilibrium of the game. We include all fairness equilibrium calculations in Appendix B2. The treatments are summarized in Table 4. Notice that each game is a separate treatment.

		Neutral				**Unkind**		
		Player A				Player A		
Equal			**R**	**M**			**R**	**M**
	Player B	**R**	12,6	3,7	Player B	R	12,6	**7,7**
		M	1,3	**6,6**		M	1,3	6,6
Unequal			**R**	**M**			**R**	**M**
	Player B	**R**	12,6	3,7	Player B	R	12,6	**3,7**
		M	1,3	**6,9**		M	1,3	2,6

Table 4. First-stage game payoffs for the four treatments.
Note: the unique Nash equilibrium outcomes are in bold.

In all our treatments, Player A has a dominant strategy of playing M. To be able to vary both the equality of the payoffs and the kindness of Player A’s actions, in addition to changing payoffs we vary the best-response of Player B. In the neutral treatments depicted in the second column, Player B’s best-response is playing M, and Player A’s choice is considered to be neutral because a deviation by Player A leads to lower payoffs for both players. In the unkind treatments depicted in the third column, Player B’s best-response is playing R, and Player A’s choice of M is considered to be unkind by Player B. That is due to the fact that Player A could increase Player B’s payoff substantially by deviating at a very small cost to herself. .

Our baseline treatment is the equal payoff-neutral action treatment depicted in the top left part of the Table 4. The Nash equilibrium for both players is to play M, with a resulting payoff of six for each. The resulting encounter is neutral for Player B because a deviation by Player A is costly to both players.

The unequal-neutral treatment is depicted in the lower left-hand side of Table 4. The only change compared to the baseline treatment is Player A’s payoff in (M,M). As in the baseline treatment, the unique Nash equilibrium prediction is (M,M) with resulting payoffs of six for Player B and nine for Player A. In this case, as in the baseline treatment, Player A’s action is neutral from the perspective of Player B because Player A's deviation would result in a Pareto dominated payoff. However, unlike the baseline treatment, Player B earns less than Player A in equilibrium.

In the equal-unkind treatment depicted on the upper right-hand side of Table 4, the only change compared to the baseline is Player B's payoff if (R,M) is played. Subsequently, the unique Nash equilibrium is playing (R,M) with a payoff of seven for each. Contrary to the baseline treatment, the encounter is unkind from the perspective of Player B as Player A could have played R to give Player B five points extra at a cost of only one point to herself.

The unequal-unkind treatment is depicted in the bottom right-hand side of Figure 4. In this game the only change compared to the baseline treatment is for Player B's payoff in (M,M). In the unique Nash equilibrium (R,M) payoffs are unequal: Player B earns three points and Player A earns seven. Moreover, by playing M, Player A is unkind to Player B, because if Player A would deviate to R, Player B would gain nine at a cost of one to Player A.

In the second stage, subjects play the same sender-receiver game they played in Study 2a. That is, if option 1 is implemented, i.e., if Player A chooses the number that is the actual outcome of the die roll, then both players earn 10 tokens. In option 2, i.e., if Player A chooses a number different than the actual outcome, Player A earns 4 tokens and Player B earns 8 tokens[14].

In this setting, Players B has the option to use deception as a punishment device to punish both unequal payoff outcomes and unkind actions. In the absence of unequal payoffs or unkind actions (equal-neutral treatment), we expect no lying. Following the findings of the literature on punishment, we conjecture that both introducing unequal payoffs and introducing unkind actions will increase deception of Player B to punish Player A in comparison to our baseline treatment. Furthermore, we expect unfair actions to lead to more punishment than unequal outcomes. We derive the following hypotheses:

Hypothesis 1: There is no lying in the equal-neutral treatment.

Hypothesis 2: There is lying in (a) equal-unkind, (b) unequal-neutral, (c) unequal-unkind treatments.

Hypothesis 3: Compared to equal-unkind treatment, there is more lying in the unequal-unkind treatment.

Hypothesis 4: Compared to unequal-neutral treatment, there is more lying in the unequal-unkind treatment.

[14] Note that there could be higher order concerns on the part of player B. If player B anticipates that player A will not trust player B's message, then player B might want to send a message different than the true one in the hope of increasing the chances of having option 1 implemented. Although we cannot rule out the possibility of non-truthful behavior occurring for other reasons than to punish the other player, such behavior should be observed in all our treatments. Therefore, our hypotheses regarding the effect of unequal payoffs and unkind action will remain unaffected by such behavior.

Results

Table 5 presents the results of the two tasks. The number of Players B who lied in the second task is depicted in parentheses. Bold numbers represent the outcome that is both the unique Nash equilibrium prediction as well as the unique Rabin fairness equilibrium. These numbers also constitutes the Nash equilibrium assuming Fehr and Schmidt inequity aversion preferences with the standard parameter values.

	Outcomes of the First and Second Task			
	M, M	**R, M**	**M, R**	**R, R**
Equal-Neutral	**10** **(0)**	4 (0)	4 (1)	0 (0)
Equal-Unkind	2 (0)	**12** **(5)**	0 (0)	4 (0)
Unequal-Neutral	**11** **(0)**	6 (0)	1 (0)	0 (0)
Unequal-Unkind	0 (0)	**12** **(4)**	2 (0)	4 (0)

Table 5. Outcomes observed in Stage 1 and Stage 2 in Study 2b.
Note: The number of Players B who lied in the second part is given in parentheses.

We have 18 observations in each treatment. For testing our hypotheses, we focus on deception that follows Nash equilibrium play in the first task. Hence, we leave out non-equilibrium results and focus only on the second stage game of the 10 observations in the equal-neutral treatment, 12 observations in the equal-unkind treatment, 11 observations in the unequal-neutral treatment, and 12 observations in the unequal-unkind treatment. Table 6 summarizes the results on deception. The second column shows that there was no deception in the neutral treatments. Thus, we confirm hypothesis 1 and we reject hypothesis 2b: when the actions are neutral, we find no punishment both when payoffs are equal and when they are unequal. In contrast, the unkind treatments show substantial deception. In the equal-unkind treatment, 42% of the players lied in their message. In the unequal-unkind treatment, 33% of the players did so. These results reject hypothesis 3, which states that unequal payoffs lead to punishment via deception (one-sided Fisher exact test, p-value 0.500). The alternative hypothesis that lying rates are the same across equal and unequal payoff treatments when controlling for the type of encounter cannot be rejected. Hence, subjects do not lie to reduce disadvantageous inequality if the intentions of the other player are perceived to be neutral.

By contrast, we find strong evidence that unkind actions lead to punishment via deception. The difference in proportions between the equal-neutral and equal-unkind treatments is significant as is the difference between the unequal-neutral and unequal-unkind treatments (both one-sided Fisher exact test, p-values 0.030 and 0.056, respectively)[15]. Thus, we confirm hypotheses 2a, 2c, and 4. Our result allows us to conclude that subjects punish unkind actions by using deception and do so regardless of whether the resulting payoffs were equal or unequal[16].

	Neutral	**Unkind**	p-values[a]
Equal	0 %	42 %	0.030
Unequal	0 %	33 %	0.056
***p*-values[a]**	-	0.500	

Table 6. Percentage of Player B's who lied in the deception task
[a]One-sided Fisher exact test

When we look at the percentage of Players A who followed Player B's message, we see that out of 45 observations across the four treatments, 33 Players A (73%) choose to do so. The percentage of Players A who followed B's recommendation is slightly lower in the unequal-unkind treatment than the others, but this difference is not statistically significant in any pairwise comparison[17].

4. Conclusion

Perceived injustice in the workplace has been shown to have detrimental effects on the quality of employees' performance and on their productivity. In this paper, we investigate whether unfair treatment affects another kind of counterproductive behavior: people's tendency to engage in deception toward others. In particular, we study whether individuals use deception as a punishment device both in situations in which lying creates a monetary advantage at the receivers' expenses, and when lying is costly. We also explore whether deception in both circumstances is triggered by inequalities in payoffs between individuals or by others' unkind behavior.

[15] The difference between the unequal-unkind and equal-kind treatment is also significant at a 10% level (one-sided Fisher exact test p-value=0.068). Furthermore, if we include all the observations in our experiment in which Player B was unkind or neutral and compare the frequency of lying when Player B was considered to be neutral with the frequency of lying when Player B was considered to be unkind, the difference is significant (one-sided Fisher exact test p-value=0.007).

[16] Contrary to what is reported in the literature (Dreber and Johannesson, 2008; Erat and Gneezy, 2012), except from the third treatment of Study 1, we find no gender differences in the propensity to lie.

[17] The percentages in each treatment are as follows: Equal-Neutral: 80%, N=10; Equal-Unkind: 75%, N=12; Unequal-Neutral: 82%, N=11; Unequal-Unkind: 58%, N=12.

Our results show that in situations in which engaging in deception is beneficial, individuals reward kind actions by being honest to the receiver. In this setting we find lying to be substantial after a negative encounter and to monotonically decrease with the kindness of the previous encounter with the possible target of deception. When engaging in deception is costly, we find that individuals use deception as a punishment tool after an unkind encounter, regardless of whether the encounter results in equal or unequal payoffs.

This study contributes to our understanding on the prevalence and causes of deception. Our results suggest that in an organizational context, workers' perception of being treated fairly is important to prevent them from using deceptive behaviors as a tool of revenge in the workplace. Future research can further look into whether the desire to punish overrides deception aversion, that is, whether people who do not lie for own benefit would lie to punish. Furthermore, the role of emotions as a driving force of deceptive punishment behavior is worth further exploration.

References

Battigalli, P., Charness G., Dufwenberg, M., (2013). Deception: the role of guilt. Journal of Economic Behavior & Organization, 93: 227-232

Battigalli, P., Dufwenberg, M., (2007). Guilt in games. American Economic Review, Papers and Proceedings 97, 170–176

Belot, M., and Schroeder M., **(**2013). "Sloppy work, lies and theft: A novel experimental design to study counterproductive behaviour." Journal of Economic Behavior and Organization, 99: 233-238.

Blount, S., (1995). "When social outcomes aren't fair: The effect of causal attributions on preferences." Organizational Behavior and Human Decision Processes 63, 131-144

Camerer, C. (2003). Behavioral Game Theory: Experiments in Strategic Interaction. Princeton University Press

Charness, G., (2004). "Attribution and Reciprocity in an Experimental Labor Market", Journal of Labor Economics 22, 665-688

Charness, G., and Dufwenberg M., (2006), "Promises and Partnership". Econometrica, 74:1579–1601.

Charness, G., and Dufwenberg M., (2010), "Bare promises: An experiment". Economics Letters, 107 (2), 281-283

Charness, G. and Levine D. (2007), "Intention and Stochastic Outcomes: An Experimental Study",Economic Journal 117(522): 1051-1072

Charness, G. and Haruvy, E. (2002) Altruism, fairness and reciprocity in a gift-exchange experiment: an encompassing approach. Games and Economic Behavior. 40, 203–231

Charness, G and Kuhn P.(2007), "Does Pay Inequality Affect Worker Effort? Experimental Evidence", Journal of Labor Economics 25(4): 693-724

Charness, G., and Kuhn P., **(**2011), "Lab Labor: What can Labor Economists Learn from the Lab?" In Handbook of Labor Economics Vol. 4 Part A. ed. Orley Ashenfelter and David Card, 229–330. San Diego and Amsterdam: North-Holland

Danilov, A., Biemann T., Kring T., Sliwka D., (2013), The dark side of team incentives: experimental evidence on advice quality from financial service professionals. Journal of Economic Behavior and Organization, 93, 266-272

Dreber, A. and M, Johannesson (2008): "Gender differences in Deception." Economics Letters, 99, 197-199

Dufwenberg, M., Kirchsteiger, G., 2000. Reciprocity and wage undercutting. European Economic Review, 44, 1069–1078.

Ellingsen, T., M. Johannesson, J. Lilja, H. Zetterqvist, (2009), “Trust and truth”. Economic Journal, 119(534) 252–276

Erat, S. and U. Gneezy (2012). “White lies”. Management Science, 58(4): 723–733

Fehr, E., and S. Gächter, (2000). “Cooperation and punishment in public goods experiments”. American Economic Review, 90, 980–994.

Fehr, E. and K.M. Schmidt, (1999). “A theory of fairness, competition, and cooperation”. Quarterly Journal of Economics, 114, 817–868

Garofalo, O. and Rott, C. (2014). “Spokesperson—To be or not to be?”. Working paper

Gino, F., Krupka E., and Weber R.A. (2013). “License to Cheat: Voluntary Regulation and Ethical Behavior”, Management Science, 59, 2187-2203

Gneezy, U., (2005). “Deception: the role of consequences.” American Economic Review, 95: 384-394

Greenberg, A. E., Smeets, P., and Zhurakhovska, L. (2014). Lying, Guilt, and Shame. Working paper.

Greenberg, J., (1990). “Employee Theft as a Reaction to Underpayment Inequity: The Hidden Cost of Pay Cuts.” Journal of Applied Psychology, 75(5), 561-568

Güth, Werner, Rolf Schmittberger, and Bernd Schwarze, (1982) "An Experimental Analysis of Ultimatum Bargaining," Journal of Economic Behavior and Organization, III (1982), 367-388

Hollinger, R. D., and J. P. Clark, (1983). Theft by employees. Lexington, MA: Lexington Books.

Houser, D., S. Vetter and J. K. Winter (2012). "Fairness and Cheating". European Economic Review, 56, 1645-1655.

Hurkens, S., and N. Kartik, (2009). “Would I lie to you? On social preferences and lying aversion.” Experimental Economics,12, 180-192.

John L., Loewenstein G. and Rick S., (2014). Cheating more for less: Upward social comparisons motivate the poorly compensated to cheat, Organizational Behavior and Human Decision Processes 123, 101–109

Krueger AB, Mas A., (2004). Strikes, Scabs, and Tread Separations: Labor Strife and the Production of Defective Bridgestone/Firestone Tires. Journal of Political Economy 112:253-89.

Lee D, Rupp NG., (2007). Retracting a Gift: How Does Employee Effort Respond to Wage Reductions. Journal of Labor Economics 25:725-61

López-Pérez, R., and Spiegelman, E. (2013). "Why do people tell the truth? Experimental evidence for pure lie aversion." Experimental Economics, 16(3):233-247.

Lundquist, T., Ellingsen, T., Gribbe, E., & Johannesson, M. (2009). "The aversion to lying." Journal of Economics Behavior and Orgaization, 70(1), 81-92.

Mas A., (2006). Pay, Reference Points, and Police Performance. The Quarterly Journal of Economics 121:783-821

Mas A., (2008). Labor Unrest and the Quality of Production: Evidence from the Construction Equipment Resale Market. Review of Economic Studies, 75:229-58

Offerman, T., (2002). "Hurting Hurts More Than Helping Helps." European Economic Review 46, 1423-37

Rabin, M., (1993). "Incorporating Fairness into Game Theory and Economics", American Economic Review, 83(5), 1281-1302

Rand, D. G., Fudenberg D., and Dreber A. (2014). It's the Thought That Counts: The Role of Intentions in Noisy Repeated Games (June 23, 2014). Working paper. Available at SSRN: http://ssrn.com/abstract=2259407 or http://dx.doi.org/10.2139/ssrn.2259407

Sutter, M., (2007). Outcomes versus intentions: On the nature of fair behavior and its development with age. Journal of Economic Psychology 28, 69-78

Sutter, M., (2009). "Deception through telling the truth?! experimental evidence from individuals and teams." The Economic Journal 119: 47-60.

Treviño, L. K., den Nieuwenboer, N. A., & Kish-Gephart, J. J. (2014). (Un)Ethical Behavior in Organizations. Annual Review of Psychology, 65(1), 635-660

Y. Yang, S. Onderstal and A. Schram, (2014). "Inequity aversion revisted." CREED Working Papers at the University of Amsterdam

Appendix A.

A1. Instructions for Study 1[18]

Welcome to this short study in decision-making. Please read the instructions carefully. If you pay attention you may earn a considerable amount of money.

The study is composed of two stages. For the entire duration of the study you will be paired with another randomly selected participant. You will be paired with the same participant during both stages. Neither of you will ever know the identity of the other. Your earnings will depend on the decisions that you and the other participant make during both stages. At the end of the study we will randomly select 1 participant out of 10 and pay them.
You will engage in two different tasks during this study. The second task will be completely different from the first one. At the end of Stage 1, you will receive the link for Stage 2.

Please enter the unique ID you received when you enrolled to this study. This ID will be used to track your participation in both stages and to calculate your compensation.

Click the arrow below to read the instructions for Stage 1

Stage 1 (Player A)

You have been assigned the role of Player A. The participant you are matched with has been assigned the role of Player B.

You start this stage with a total of $10. Player B starts with no money. You now have the opportunity to send an amount of money from $0 to $10 to the other participant. The other participant will receive the amount you decided to send him/her, multiplied by a factor of 3. Thus, for every $1 that you send, he/she will earn $3. Any amount not sent to the other participant will be yours to keep.

Player B will be informed of the choice you face and your decision. Then, both of you will receive the instructions for Stage 2, in which the two of you will engage in a different task. Remember that you will be matched with the same participant during both stages.

How much do you wish to transfer to the other participant?

- o I wish to transfer $0 to Player B and keep $10 for myself
- o I wish to transfer $2 to Player B and keep $8 for myself
- o I wish to transfer $4 to Player B and keep $6 for myself
- o I wish to transfer $6 to Player B and keep $4 for myself
- o I wish to transfer $8 to Player B and keep $2 for myself
- o I wish to transfer $10 to Player B and keep $0 for myself

----

Stage 1 (Player B)

[18] We report the instructions from the first treatment in Study 1. The instructions for the second and third treatment as well as for the Study 2b, are the same with the required modifications where needed. The full set of instructions are available from the authors.

You have been assigned the role of Player B. The participant you are matched with was assigned the role of Player A.

You start with no money. At the beginning of the study, Player A was given $10 and had the opportunity to send you an amount of money from $0 to $10. At this stage, you will receive the amount that the other participant decided to send you, multiplied by a factor of 3.
Thus, for every $1 that the other participant decided to send, you will receive $3. The other participant will keep any amount that he or she does not send to you. That amount will not be multiplied by a factor of 3.

You do not have a decision to make during Stage 1. Click the arrow below to find out how much money Player A decided to send you.

Player A chose to send you $___ and keep $____ for himself/herself. Since the amount is multiplied by a factor of 3, you receive $___.

You have reached the end of Stage 1. Click the arrow below to read the instructions for Stage 2.

Stage 2 (Player B)

Please remember that you are still paired with the same participant (Player A) during this stage.

At the beginning of the study, we have rolled a 6-sided die, and obtained the outcome 4. Player A does not know that the outcome was 4.

We now ask you to send a message to Player A. The message will correspond to a number from 1 to 6. The six possible messages are:

Message 1: "The outcome from the roll of the die is 1"
Message 2: "The outcome from the roll of the die is 2"
Message 3: "The outcome from the roll of the die is 3"
Message 4: "The outcome from the roll of the die is 4"
Message 5: "The outcome from the roll of the die is 5"
Message 6: "The outcome from the roll of the die is 6"

Your message will be shown to Player A, and then he or she will be asked to choose a number between 1 and 6. The message you send is the only information he or she will have regarding the roll of the die.

Importantly, Player A does not know the outcome of the roll but will see your message before choosing. The number chosen by Player A will determine the payments for this stage of the study. The two possible monetary payments are:

If Player A chooses the actual outcome of the roll of the die, 4, then **you will be paid $10** and **he/she will be paid $20**.
If Player A chooses a number different than the actual outcome, **you will be paid $20** and **he/she will be paid $10**.

Player A will never know the actual payments associated with each choice or whether your message was true or not. Moreover, he or she will never know the sums to be paid to you according to the different choices. All he or she will know is the amount of money they will be paid.

The message I wish to send to Player A is:

- o "The outcome from the roll of the 6-sided die is 1"
- o "The outcome from the roll of the 6-sided die is 2"
- o "The outcome from the roll of the 6-sided die is 3"
- o "The outcome from the roll of the 6-sided die is 4"
- o "The outcome from the roll of the 6-sided die is 5"
- o "The outcome from the roll of the 6-sided die is 6"

Stage 2 (Player A)

Please remember that you are still paired with the same participant (Player B) during this stage.

Before starting this study, we have rolled a 6-sided die, and revealed the outcome to the other participant (Player B), but we are not going to tell it to you.

After being informed of the roll of the die, Player B has sent a message to you. The message corresponds to a number from 1 to 6.
The 6 possible messages are:

- Message 1: "The outcome from the roll of the die is 1"
- Message 2: "The outcome from the roll of the die is 2"
- Message 3: "The outcome from the roll of the die is 3"
- Message 4: "The outcome from the roll of the die is 4"
- Message 5: "The outcome from the roll of the die is 5"
- Message 6: "The outcome from the roll of the die is 6"

The message Player B sent is:
Message: "The outcome from the roll of the 6-sided die is __"

Now we ask you to choose a number between 1 and 6. The message you received is the only information you will have regarding the roll of the die. Your choice of a number will determine the payments in the experiment according to two different options (option A and option B), known only to Player B.

If you will choose the same number as the number that came up in the roll of the die, both of you will be paid according to option A. If you will choose a number different than the actual number, you will both be paid according to option B.

The number I choose is:

- o 1
- o 2
- o 3
- o 4
- o 5
- o 6

A2. Regression Results

TABLE 1A—PROBIT REGRESSIONS OF LYING (STUDY 1A)

Lying	(1)	(2)
Amount sent	-.041**	-.041**
	(.16)	(.016)
Female		-.025
		(.096)
Observations	108	108

***p<.01, ** p<.05, *p<.10.

Note: The table presents marginal effects estimated from probit regression. Dependent variable: Lying (1 false message and 0 truthful message). Robust standard errors are in parenthesis.

TABLE 1B—PROBIT REGRESSIONS OF LYING (STUDY 1B)

Lying	(1)	(2)
Amount sent	-.041***	-.041***
	(.10)	(.010)
Female		.097
		(.11)
Observations	306	306
Clusters	51	51

***p<.01, ** p<.05, *p<.10.

Note: The table presents marginal effects estimated from probit regression. Dependent variable: Lying (1 false message and 0 truthful message). Clustered standard errors are in parenthesis.

TABLE 1C—PROBIT REGRESSIONS OF LYING (STUDY 1C)

Lying	(1)	(2)
Amount sent	-.004	-.004
	(.006)	(.006)
Female		-.29**
		(.11)
Observations	108	108

***p<.01, ** p<.05, *p<.10.

Note: The table presents marginal effects estimated from probit regression. Dependent variable: Lying (1 false message and 0 truthful message). Robust standard errors are in parenthesis.

TABLE 1D—PROBIT REGRESSIONS OF LYING (STUDY 2A)

Lying	(1)	(2)
Amount sent	-.035***	-.035***
	(.008)	(.009)
Female		-.008
		(.064)
Observations	216	216
Clusters	36	36

***p<.01, ** p<.05, *p<.10.

Note: The table presents marginal effects estimated from probit regression. Dependent variable: Lying (1 false message and 0 truthful message). Clustered standard errors are in parenthesis.

Appendix B

B1. Study 2b. Instructions for the Baseline, Player A[19]:

Welcome to this experiment on decision-making. We will first go through the instructions together. Talking is strictly forbidden during the experiment. If you have a question, please raise your hand and an experimenter will come to you to answer your question.

The experiment will last about 20 minutes. If you follow the instructions carefully, and depending on the decisions of some other player, you may earn some money. During the experiment your earnings will be denoted in tokens. After the experiment your earnings will be converted into money at a rate of 1 token is 0.5 Euros. For instance, if at the end of the experiment you have 12 tokens and <u>if you are selected to receive payment</u>, you will be paid 0.5*12=6 Euros.

In this experiment you will be given the label of player A. At the beginning of the experiment, you will randomly be matched with another player B. **All the participants in this room are player A.** All the players B are located in another part of the library. Each participant A is paired with a participant B. You will not learn with whom you are matched with during or after the experiment. Your player labels will stay the same throughout the whole experiment. Additionally, the person you are matched with will stay the same.

Your earnings in this experiment will depend upon your individual decisions and the decisions of the Player B you are matched with. We will randomly choose one Player A and one Player B to be paid privately and in cash at the end of the experiment.

<u>Procedure</u>: In this experiment, you and Player A, will both make a choice. You can choose between two actions R and M. When you make your decision, you will not know what Player B will choose. In the table below you can see the number of tokens that are associated with each decision you and Player B make. The **bold** number in the bottom left corner of every field represents your payoff, and the number in the upper right corner represents Player's B payoff.

<table>
<tr><td rowspan="4">Player A
(You)</td><td></td><td colspan="2">Player B</td></tr>
<tr><td></td><td>R</td><td>M</td></tr>
<tr><td>R</td><td>12
6</td><td>1
3</td></tr>
<tr><td>M</td><td>3
7</td><td>6
6</td></tr>
</table>

So, there are four possibilities:

-if you choose R and Player A chooses R, you will earn 6 tokens and Player B will earn 12 tokens.
-If you choose R and Player B chooses M, you will earn 3 tokens and Player B will earn 1 token.
-If you choose M and Player A chooses R, you will earn 7 tokens and Player B will earn 3 tokens.
-if you choose M and Player A chooses M, you will earn 6 tokens each.

After you have made a decision, you will write your decision in the decision sheet that is given to you. When all player As have made their decision, we will collect the decision sheets. We will then

[19] We only report the instructions for the baseline. The instructions for the rest of the treatments and for Study 2a are the same with some modifications where required.

determine your payoff depending on your and Player's B decision. You will then be informed of Player's B choice, and your payoff via the decision sheet that will be handed back to you.

If you have any questions please raise your hand and one of the experimenters will come to help you.

Stage 2. Player A

In the second and final part of this experiment, you are paired again with the **same** participant as in the previous part. You are again Player A and have to make a decision. Your earnings in this part will depend upon your individual decision and the decision of Player B you are matched with.

You will face one task in this part of the experiment. Your earnings from this part will be in addition to your earnings from the previous part. As mentioned at the beginning of the experiment, we will randomly choose 1 pair and therefore pay one Player A and one Player B at the end of the experiment, privately and in cash.

In this part you will face the following task.

Before starting this part of the experiment, Player B has rolled a 6-sided die. You will not be informed about the actual outcome of the die roll. Player B has made a decision to send a message to you. The message corresponds to a number from 1 to 6. There are six possible messages:

Message 1: "The outcome from the roll of the die is 1."

Message 2: "The outcome from the roll of the die is 2."

Message 3: "The outcome from the roll of the die is 3."

Message 4: "The outcome from the roll of the die is 4."

Message 5: "The outcome from the roll of the die is 5."

Message 6: "The outcome from the roll of the die is 6."

After delivering this message to every player A we will ask you to choose a number between 1 and 6 and write it down on the decision sheet that will be given to you. Your decision will involve filling out the empty space in the following sentence: "The outcome from the roll of the die is ______."

The number you choose will determine the payoffs in the experiment according to two different options (option 1 and option 2). If the number you choose corresponds to the actual outcome of die roll, both of you will be paid according to option 1. If you will choose a number different than the actual number, you will both be paid according to option 2. Player's B and your payoff associated with these two options are made known to Player B. At the end of the experiment you will only be informed only about your **own** final payoff.

To sum up: Player B rolls a die. He decides to send a message to you about the outcome. You have to choose a number between 1 and 6. Your choice will determine your and Player's B earnings in this part of the experiment. If the number you choose corresponds to the actual outcome of the die roll, both of you will be paid according to option 1. If you will choose a number different than the actual number, you will both be paid according to option 2. However, **only** Player B will know the actual outcome of the die roll and the monetary outcomes attached to the choices 1 and 2.

If you have any questions please raise your hand and one of the experimenters will come to help you.

Instructions for the Baseline, Player B

INSTRUCTIONS:

Welcome to this experiment on decision making. We will first go through the instructions together. Talking is strictly forbidden during the experiment. If you have a question, please raise your hand and an experimenter will come to you to answer your question.

The experiment will last about 20 minutes. If you follow the instructions carefully, and depending on the decisions of some other player, you may earn some money. During the experiment your earnings will be denoted in tokens. After the experiment your earnings will be converted into money at a rate of 1 token is 0.5 Euros. For instance, if at the end of the experiment you have 12 tokens and <u>if you are selected to receive payment</u>, you will be paid 0.5*12=6 Euros.

In this experiment you will be given the label of player B. At the beginning of the experiment, you will randomly be matched with another player A. **All the participants in this room are player B.** All the player As are located in another part of the library. Each participant A is paired with a participant B. You will not learn with whom you are matched with during or after the experiment. Your player labels will stay the same throughout the whole experiment. Additionally, the person you are matched with will stay the same.

Your earnings in this experiment will depend upon your individual decisions and the decisions of the Player A you are matched with. We will randomly choose one Player A and one Player B to be paid privately and in cash at the end of the experiment.

<u>Procedure</u>:

In this experiment, you and Player A will both make a choice. You can choose between two actions R and M. When you make your decision, you will not know what Player A will choose. In the table below you can see the number of tokens that are associated with each decision you and Player A make. The **bold** number in the bottom left corner of every field represents your payoff, and the number in the upper right corner represents Player A's payoff.

		Player A	
		R	M
Player B (You)	R	6 **12**	7 **3**
	M	3 **1**	6 **6**

So, there are four possibilities:

- if you choose R and Player A chooses R, you will earn 12 tokens and Player A will earn 6 tokens.
- If you choose R and Player A chooses M, you will earn 3 tokens and Player A will earn 7 tokens.
- If you choose M and Player A chooses R, you will earn 1 tokens and Player A will earn 3 tokens.
- if you choose M and Player A chooses M, you will earn 6 tokens each.

After you have made a decision, you will write your decision in the decision sheet that is given to you. When all player Bs have made their decision, we will collect the decision sheets. We will then determine your payoff depending on your and Player's A decision. You will then be informed of Player's A choice, and your payoff via the decision sheet that will be handed back to you.

If you have any questions please raise your hand and one of the experimenters will come to help you.

Stage 2. Player B

In the second and final part of this experiment, you are paired again with the **same** participant as in the previous part. You are again Player B and have to make a decision. Your earnings in this part will depend upon your individual decision and the decision of Player A you are matched with.

You will face one task in this part of the experiment. Your earnings from this part will be in addition to your earnings from the previous part. As mentioned at the beginning of the experiment, we will randomly choose 1 pair and therefore pay one Player A and one Player B at the end of the experiment, privately and in cash.

In this part you will face the following task.

We will first roll a die with every player B privately. Only the experimenter and you will see the outcome of the die roll. Player A will not be informed about the outcome of the die roll. However, he will be told that you have rolled a die. After everyone in the room is finished with rolling a die, we will ask you to send a message to Player A and write it down to the decision sheet you are provided. The message will correspond to a number from 1 to 6.

There are six possible messages:

Message 1: “The outcome from the roll of the 6-sided die is 1.”

Message 2: “The outcome from the roll of the 6-sided die is 2.”

Message 3: “The outcome from the roll of the 6-sided die is 3.”

Message 4: “The outcome from the roll of the 6-sided die is 4.”

Message 5: “The outcome from the roll of the 6-sided die is 5.”

Message 6: “The outcome from the roll of the 6-sided die is 6.”

Your message will be shown to Player A and then he will be asked to choose a number between 1 and 6. The choice of the number by Player A will determine the payoffs for this part of the experiment. There are two possible outcomes:

Alternative 1: If Player A chooses the actual outcome of the roll of the die, then you will both earn 10 tokens.

Alternative 2: If Player A chooses a number different than the actual outcome, you will earn 8 tokens and he will earn 4 tokens

Finally, only you will know of the above mentioned monetary values connected to each message. Player A will not be informed of these monetary values. However, he will be told that there are two possible alternatives (1 and 2) and if he chooses the actual outcome of the die alternative 1 will be implemented, otherwise Alternative 2 will be implemented, without knowing the monetary values that correspond to these alternatives. He will also be informed that you are aware of the monetary value that corresponds to each one of them. At the end of the experiment, he will be informed **only** about his **own final** payoff and not for the alternatives.

To sum up, you will roll a die, you will send a message to Player A, Player A will make a choice. His choice will determine the earnings in this part of the experiment. Player A will **not** know the actual outcome of the die roll nor the monetary outcomes attached to the choices.

If you have any questions please raise your hand and one of the experimenters will come to help you.

B2. Study 2b. Calculation of the fairness equilibrium

In Rabin's fairness-based model, players' perceptions of their opponents' intentions play a central role via their effect on the utility function. Whether an action is preferred over an alternative action depends upon the direct material payoff, the belief about whether rival players are being helpful or harmful and whether this action helps or hurts rival players. Perceived kindness does not depend on inequity between payoffs. According to his model the fairness equilibrium is a situation in which no player can increase his utility by changing his strategy, given everyone else's strategy, and in which beliefs are correct (including beliefs about others' strategy, and beliefs about others' beliefs about one's own strategy). Rabin formalized fairness using a two person-two action game in which *i* is the person whose utility is being measured. The utility of person *i* is as follows:

$U_i(a_i,b_j,c_i)=\pi_i(a_i,b_j)+\widetilde{f}_{\mathrm{j}}(b_j,c_i)\left[1+f_i(a_i,b_j)\right]$, where:

a_i represents player *i*'s strategy, b_j is player *i*'s belief about what player *j*'s strategy is, and c_i is player *i*'s belief about player *j*'s belief about player *i*'s strategy. $\pi_i(a_i,b_j)$ represents the payoffs player *i* receives. Player *i*'s belief about how kind player *j* is being to her is given by $\widetilde{f}_{\mathrm{j}}(b_j,c_i)=\dfrac{\pi_i(c_i,b_j)-\pi_i^e(c_i)}{\pi_i^h(c_i)-\pi_i^{\min}(c_i)}$. In this function, $\pi_i^{\min}(c_i)$ denotes player *i*'s worst possible payoff when c_i is played, and $\pi_i^e(c_i)$ depicts the average Pareto efficient payoff of player *i* when c_i is played. That is, $\pi_i^e(c_i)=\dfrac{\pi_i^h(c_i)-\pi_i^l(c_i)}{2}$,where $\pi_i^h(c_i)$ is player *i*'s highest payoff and $\pi_i^l(c_i)$ is player *i*'s lowest payoff among points that are Pareto efficient when playing c_i. Likewise, player *i*'s kindness to player *j* is given by $f_i(a_i,b_j)=\dfrac{\pi_j(b_j,a_i)-\pi_j^e(b_j)}{\pi_j^h(b_j)-\pi_j^{\min}(b_j)}$. $\pi_j^e(b_j)=\dfrac{\pi_j^h(b_j)-\pi_j^l(b_j)}{2}$,where $\pi_j^h(b_j)$ is player *j*'s highest payoff and $\pi_j^l(b_j)$ is player *j*'s lowest payoff among points that are Pareto efficient. $\pi_j^{\min}(b_j)$ is the worst possible payoff for player *j* when playing b_j.

Player *i* chooses a_i to maximize the utility $U_i(a_i,b_j,c_i)$. The function $\widetilde{f}_{\mathrm{j}}(b_j,c_i)$ represents player *i*'s beliefs about how kindly player *j* is treating him. This function is equivalent to the function $f_j(\alpha_j,b_i)$. Therefore, they are only computed once for the calculations that follow.

1. *Predictions for the Equal-Neutral Treatment:*

The payoff matrix (on the left) and the utility matrix (on the right) of the equal-neutral treatment under the condition that beliefs are correct are as follows:

		Player A	
		R	M
Player B	R	6 12	7 3
	M	3 1	6 6

		Player A	
		R	M
Player B	R	6 12.5	6.75 2.25
	M	3 1	5.5 6

<u>Outcome R,R</u>

$$f_A(R,\ R) = f_A(b_2,\ c_B) = \frac{\pi_B(c_B,\ b_A)\ -\pi_B{}^{e}(c_B)}{\pi_B{}^{h}(c_B)\ -\pi_B{}^{min}(c_B)} = \frac{\pi_B(R,\ R)\ -\pi_B{}^{e}(R)}{\pi_B{}^{h}(R)\ -\pi_B{}^{min}(R)} = \frac{12-\left(\frac{12+3}{2}\right)}{12-3} = 0.5$$

$$f_A(M,R) = f_A(b_A,\ c_B) = \frac{\pi_B(c_B,\ b_A)\ -\pi_B{}^{e}(c_1)}{\pi_B{}^{h}(c_1)\ -\pi_B{}^{min}(c_1)} = \frac{\pi_B(\mathrm{R,M})\ -\pi_B{}^{\mathrm{e}}(\mathrm{R})}{\pi_B{}^{\mathrm{h}}(\mathrm{R})\ -\pi_B{}^{\mathrm{min}}(\mathrm{R})} = \frac{3-\left(\frac{12+3}{2}\right)}{12-3} = -0.5$$

$$\tilde{f}_B(R,\ R) = \tilde{f}_B(a_B,\ b_A) = \frac{\pi_A(b_A,\ a_B)\ -\pi_A{}^{e}(b_A)}{\pi_A{}^{h}(b_A)\ -\pi_A{}^{min}(b_A)} = \frac{\pi_A(R,\ R)\ -\pi_A{}^{e}(R)}{\pi_A{}^{h}(R)\ -\pi_A{}^{min}(R)} = \frac{6-\left(\frac{6+6}{2}\right)}{6-3} = 0$$

$$\tilde{f}_B(M,R) = \tilde{f}_B(a_B,\ b_A) = \frac{\pi_A(b_A,\ a_B)\ -\pi_A{}^{e}(b_A)}{\pi_A{}^{h}(b_A)\ -\pi_A{}^{min}(b_A)} = \frac{\pi_A(R,\ M)\ -\pi_A{}^{e}(R)}{\pi_A{}^{h}(R)\ -\pi_A{}^{min}(R)} = \frac{3-\left(\frac{6+6}{2}\right)}{6-3} = -1$$

The utilities are then:

$$\begin{aligned}
U_A(R,R,R) &= U_A(a_A,\ b_B,\ c_A)\ = \pi_A(a_A,\ b_B)\ +\tilde{f}_B(b_B,\ c_A)\left[1+f_A(a_A,\ b_B)\right] \\
&= \pi_A(R,R)\ +\tilde{f}_B(R,R)\left[1+f_A(R,R)\right] =\ 6+0(1+0.5)=6 \\
U_A(M,R,R) &= \pi_A(M,R)\ +\tilde{f}_B(R,R)\left[1+f_A(M,R)\right] = 7+0(1-0.5)=7 \\
U_B(R,R,R) &= \pi_B(R,R)\ +\tilde{f}_A(R,R)\left[1+f_B(R,R)\right] = 12+0.5(1+0)=12.5 \\
U_B(M,R,R) &= \pi_B(M,R)\ +\tilde{f}_A(R,R)\left[1+f_B(M,R)\right] = 1+0.5(1-1)=1
\end{aligned}$$

Outcome R,M

$f_A(M,R)=-0.5,\ f_A(R,R)=0.5,\ \tilde{f}_B(R,M)=0.5,\ \tilde{f}_B(M,M)=-0.5,$
$U_A(M,R,M)=6.75,\ U_A(R,R,M)=6.75,\ U_B(R,M,R)=2.25,\ U_B(M,M,R)=5.75$

Outcome M,R

$f_A(R,M)=-1,\ f_A(M,M)=0,\ \tilde{f}_B(M,R)=-1,\ \tilde{f}_B(R,R)=0,$
$U_A(R,M,R)=3,\ U_A(M,M,R)=5,\ U_B(M,R,M)=1,\ U_B(R,R,M)=11$

Outcome M,M

$f_A(M,M)=0,\ f_A(R,M)=-1,\ \tilde{f}_B(M,M)=-0.5,\ \tilde{f}_B(R,M)=0.5$
$U_A(M,M,M)=5.5,\ U_A(R,M,M)=3,\ U_B(M,M,M)=6,\ U_B(R,M,M)=3$

The outcome (M, M) is the only fairness equilibrium.

2. ***Predictions for the Equal-Unkind Treatment:***

The payoff matrix and the utility matrix of the equal-unkind treatment under the condition that beliefs are correct are as follows:

		Player A	
		R	M
Player B	R	6 12	7 7
	M	3 1	6 6

		Player A	
		R	M
Player B	R	6 12.5	7 6.5
	M	3 1	5 6

<u>Outcome R,R</u>

$$f_A(R,R)= 0.5,\ f_A(M,R)= -0.5,\ \tilde{f}_B(R,R)= 0,\ \tilde{f}_B(M,R)= -1$$
$$U_A(R,R,R)= 6,\ U_A(M,R,R)= 7,\ U_B(R,R,R)= 12.5,\ U_B(M,R,R)= 1$$

<u>Outcome R,M</u>

$$f_A(M,R)=-0.5,\ f_A(R,R)=0.5,\ \tilde{f}_B(R,M)=0,\ \tilde{f}_B(M,M)=-1,$$
$$U_A(M,R,M)=7,\ U_A(R,R,M)=6,\ U_B(R,M,R)=6.5,\ U_B(M,M,R)=6$$

<u>Outcome M,R</u>

$$f_A(R,M)=-1,\ f_A(M,M)=0,\ \tilde{f}_B(M,R)=-1,\ \tilde{f}_B(R,R)=0$$
$$U_A(R,M,R)=3,\ U_A(M,M,R)=5,\ U_B(M,R,M)=1,\ U_B(R,R,M)=11$$

<u>Outcome M,M</u>

$$f_A(M,M)=0,\ f_A(R,M)=-1,\ \tilde{f}_B(M,M)=-1,\ \tilde{f}_B(R,M)=0,$$
$$U_A(M,M,M)=5,\ U_A(R,M,M)=3,\ U_B(M,M,M)=6,\ U_B(R,M,M)=7,$$

The outcome (R,M) is the only fairness equilibrium.

3. *Predictions for the Unequal-Neutral Treatment:*

The payoff matrix and the utility matrix of the unequal-neutral treatment under the condition that beliefs are correct are as follows:

		Player A	
		R	M
Player B	R	6 12	7 3
	M	3 1	9 6

		Player A	
		R	M
Player B	R	6 12.5	6.5 3
	M	6 1	9 6

Outcome R,R

$$f_A(R,R) = 0.5,\ f_A(M,R) = -0.5,\ \tilde{f}_B(R,R) = 0,\ \tilde{f}_B(M,R) = -1,$$
$$U_A(R,R,R) = 6,\ U_A(M,R,R) = 7,\ U_B(R,R,R) = 12.5,\ U_B(M,R,R) = 1,$$

Outcome R,M

$$f_A(M,R) = -0.5,\ f_A(R,R) = 0.5,\ \tilde{f}_B(R,M) = -1,\ \tilde{f}_B(M,M) = 0,$$
$$U_A(M,R,M) = 6.5,\ U_A(R,R,M) = 4.5,\ U_B(R,M,R) = 3,\ U_B(M,M,R) = 5.5,$$

Outcome M,R

$$f_A(R,M) = -1,\ f_A(M,M) = 0,\ \tilde{f}_B(M,R) = -1,\ \tilde{f}_B(R,R) = 0,$$
$$U_A(R,M,R) = 6,\ U_A(M,M,R) = 8,\ U_B(M,R,M) = 1,\ U_B(R,R,M) = 11,$$

Outcome M,M

$$f_A(M,M) = 0,\ f_A(R,M) = -1,\ \tilde{f}_B(M,M) = 0,\ \tilde{f}_B(R,M) = -1,$$
$$U_A(M,M,M) = 9,\ U_A(R,M,M) = 3,\ U_B(M,M,M) = 6,\ U_B(R,M,M) = 3,$$

The outcome (M,M) is the only fairness equilibrium.

4. *Predictions for the Unequal-Unkind Treatment:*

The payoff matrix and the utility matrix of the unequal-unkind treatment under the condition that beliefs are correct are as follows:

		Player A	
		R	M
Player B	R	6 12	7 3
	M	3 1	6 2

		Player A	
		R	M
Player B	R	6 12.5	7 2.5
	M	3 1	5 2

<u>Outcome R,R</u>

$$f_A(R,R)=0.5,\ f_A(M,R)=-0.5,\ \tilde{f}_B(R,R)=0,\ \tilde{f}_B(M,R)=-1$$
$$U_A(R,R,R)=6,\ U_A(M,R,R)=7, U_B(R,R,R)=12.5,\ U_B(M,R,R)=1$$

<u>Outcome R,M</u>

$$f_A(M,R)=-0.5,\ f_A(R,R)=0.5,\ \tilde{f}_B(R,M)=0,\ \tilde{f}_B(M,M)=-1$$
$$U_A(M,R,M)=7,\ U_A(R,R,M)=6,\ U_B(R,M,R)=2.5,\ U_B(M,M,R)=2$$

<u>Outcome M,R</u>

$$f_A(R,M)=-1,\ f_A(M,M)=0,\ \tilde{f}_B(M,R)=-1,\ \tilde{f}_B(R,R)=0$$
$$U_A(R,M,R)=3,\ U_A(M,M,R)=5,\ U_B(M,R,M)=1,\ U_B(R,R,M)=11$$

<u>Outcome M,M</u>

$$f_A(M,M)=0,\ f_A(R,M)=-1,\ \tilde{f}_B(M,M)=-1,\ \tilde{f}_B(R,M)=0,$$
$$U_A(M,M,M)=5,\ U_A(R,M,M)=3,\ U_B(M,M,M)=2,\ U_B(R,M,M)=3$$

The outcome(R,M) is the only fairness equilibrium.

B3. Study 2b. Calculation of equilibrium of the inequality averse preferences

In Fehr and Schmidt's model (1999) players care about their payoff and the payoff of others as well. Inequity aversion plays a role and an individual has disutility either if his payoff exceeds or is less than his counterpart's payoff. Moreover, the disutility of someone else earning more is higher than the disutility from earning more than others. Individuals maximize their utility taking into consideration inequity issues. Specifically, consider a setting with individuals $\{1,2,\ldots,n\}$ who receive payoff outcomes x_i. Then the utility to person i is given by

$$U_i\left(\left\{x_i, x_j\right\}\right) = x_i - \frac{a_i}{n-1} \times \sum \max\left(x_j - x_i, 0\right) - \frac{\beta_i}{n-1} \times \sum \max\left(x_i - x_j, 0\right)$$

and for a 2 person's game (n=2) the utility function simplifies to the form:

$$U_i\left(x_i,\ x_j\right) = \ x_i - \alpha_i max\ \left\{x_j - x_i,\ 0\right\} - \beta_i max\ \left\{x_i - x_j,\ 0\right\},\ i \neq j, \alpha_i \geq \beta_i, \text{ and } 0 \leq \beta_i < 1$$

Next, we calculate the equilibrium using the above utility function for players in each outcome.

1. ***Predictions for the Equal-Neutral Treatment:***

The payoff matrix and the utility matrix of the equal-neutral treatment are as follows:

		Player A	
		R	M
Player B	R	12, 6	3, 7
	M	1, 3	6, 6

Payoff matrix

		Player A	
		R	M
Player B	R	$12-6\beta_A$, $6-6\alpha_B$	$3-4\alpha_A$, $7-4\beta_B$
	M	$1-2\alpha_A$, $3-2\beta_B$	6, 6

Utility matrix

(i) <u>Outcome R,R</u>

For Player A, playing M gives a higher utility than playing R when player B plays R because $U_A(R,M) \geq U_A(R,R) \Rightarrow 7-4\beta_A \geq 6-6\alpha_A \Rightarrow 6\alpha_A - 4\beta_A \geq -1$. The last inequality holds since $\alpha_i \geq \beta_i$.

(ii) <u>Outcome R,M</u>

For Player B, playing M gives a higher utility than playing R when player A plays M, because $U_B(M,M) \geq U_B(R,M) \Rightarrow 6 \geq 3-4\alpha_B \Rightarrow \alpha_B \geq -\frac{3}{4}$. This inequality holds since $\alpha_i \geq 0$.

(iii) <u>Outcome M,R</u>

For Player B, playing R gives a higher utility than playing M when player A plays R because $U_B(R,R) \geq U_B(M,R) \Rightarrow 12-6\beta_B \geq 1-2\alpha_B \Rightarrow 11 \geq 6\beta_B - 2\alpha_B$. The last inequality holds since $\alpha_i \geq 0$ and $\beta_i < 1$.

(iv) <u>Outcome M,M (the only equilibrium using inequality aversion)</u>

For Player A, playing M gives a higher utility than playing R when player B plays M. $U_A(M,M) \geq U_A(M,R) \Rightarrow 6 \geq 3-2\beta_A \Rightarrow \beta_A \geq -\frac{3}{2}$. This holds since $\beta_i \geq 0$.

See (ii) for Player B.

2. ***Predictions for the Equal-Unkind Treatment:***

		Player A	
		R	M
Player B	R	6 12	7 7
	M	3 1	6 6

		Player A	
		R	M
Player B	R	$6\text{-}6\alpha_B$ $12\text{-}6\beta_A$	7 7
	M	$3\text{-}2\beta_B$ $1\text{-}2\alpha_A$	6 6

(i) <u>Outcome R,R</u>

For Player A, playing M gives a higher utility than playing R when player B plays R because $U_A\left(R,M\right) \geq U_A\left(R,R\right) \Rightarrow 7 \geq 6-6\alpha_A \Rightarrow \alpha_A \geq -\frac{1}{6}$. The last inequality holds since $\alpha_i \geq 0$.

(ii) <u>Outcome R,M (the only equilibrium using inequality aversion)</u>

See (i) for Player A.

For Player B, playing R gives a higher utility than playing M when player A plays M, because: $U_B\left(R,M\right) \geq U_B\left(M,M\right) \Rightarrow 7 \geq 6$. This inequality is always true.

(iii) <u>Outcome M,R</u>

For Player B playing R gives a higher utility than playing M when player A plays R, because:$U_A\left(R,R\right) \geq U_A\left(M,R\right) \Rightarrow 12-6\beta_B \geq 1-2\alpha_B \geq\Rightarrow 2\alpha_B - 6\beta_B \geq -11$. This inequality holds since $\alpha_i \geq \beta_i$ and $\beta_i < 1$.

(iv) <u>Outcome M,M</u>

See (ii) for Player B's decision.

3. *Predictions for the Unequal-Neutral Treatment:*

The payoff matrix and the utility matrix of the unequal-neutral treatment are as follows:

		Player A	
		R	M
Player B	R	6 12	7 3
	M	3 1	9 6

		Player A	
		R	M
Player B	R	$6-6\alpha_B$ $12-6\beta_A$	$7-4\beta_B$ $3-4\alpha_A$
	M	$3-2\beta_B$ $1-2\alpha_A$	$9-3\beta_B$ $6-3\alpha_A$

(i) Outcome R,R

For Player A playing M gives a higher utility than playing R when player B plays R, because $U_A(R,M) \geq U_A(R,R) \Rightarrow 7-4\beta_A \geq 6-6\alpha_A \Rightarrow 6\alpha_A - 4\beta_A \geq -1$. This inequality holds since $\alpha_i \geq \beta_i$.

(ii) Outcome R,M

For Player B playing M gives a higher utility than playing R when player A plays M, because $U_B(M,M) \geq U_B(R,M) \Rightarrow 6-3\alpha_B \geq 3-4\alpha_B \Rightarrow \alpha_B \geq -3$. This is true since $\alpha_i \geq 0$.

(iii) Outcome M,R

For Player A playing M gives a higher utility than playing R when player B plays M, because $U_A(M,M) \geq U_A(M,R) \Rightarrow 9-3\beta_A \geq 3-2\beta_A \Rightarrow \beta_A \leq 6$. This inequality holds since $\beta_i < 1$.

For Player B playing R gives a higher utility than playing M when player A plays R, because $U_B(R,R) \geq U_B(M,R) \Rightarrow 12-6\beta_B \geq 1-2\alpha_B \Rightarrow 2\alpha_B - 6\beta_B \geq -11$. This inequality holds since $\alpha_i \geq \beta_i$ and $\beta_i < 1$.

(iv) Outcome M,M (the only equilibrium using inequality aversion)

See (iii) for Player A, and (ii) for Player B.

4. *Predictions for the Unequal-Unkind Treatment:*

The payoff matrix and the utility matrix of the unequal-unkind treatment are as follows:

		Player A	
		R	M
Player B	R	6 12	7 3
	M	3 1	6 2

		Player A	
		R	M
Player B	R	$6-6\alpha_B$ $12-6\beta_A$	$7-4\beta_B$ $3-4\alpha_A$
	M	$3-2\beta_B$ $1-2\alpha_A$	$6-4\beta_B$ $2-4\alpha_A$

(i) Outcome R,R

For Player A playing M gives a higher utility than playing R when player B plays R, because $U_A(R,M) \geq U_A(R,R) \Rightarrow 7-4\beta_A \geq 6-6\alpha_A \Rightarrow 6\alpha_A - 4\beta_A \geq -1$. This inequality holds since $\alpha_i \geq \beta_i$.

(ii) Outcome R,M (the only equilibrium using inequality aversion)

For Player A playing M gives a higher utility than playing R when player B plays R, because $U_A(R,M) \geq U_A(R,R) \Rightarrow 7-4\beta_A \geq 6-6\alpha_A \Rightarrow 4\beta_A - 6\alpha_A \leq 1$. This inequality holds since $\alpha_i \geq \beta_i$.

For Player B playing R gives a higher utility than playing M when player A plays M, because $U_B(R,M) \geq U_B(M,M) \Rightarrow 3-4\alpha_B \geq 2-4\alpha_B \Rightarrow 3 \geq 2$.

(iii) Outcome M,R

For Player B playing R gives a higher utility than playing M when player A plays R, because $U_B(R,R) \geq U_B(M,R) \Rightarrow 12-6\beta_B \geq 1-2\alpha_B \Rightarrow 2\alpha_B - 6\beta_B \geq -11$. This inequality holds since $\alpha_i \geq \beta_i$ and $\beta_i < 1$.

(iv) Outcome M,M

See (ii) above that playing R gives a higher utility to Player B.