UNIVERSITY OF HEIDELBERG
DEPARTMENT OF ECONOMICS

Promises and Opportunity Cost

Arjun Sengupta

Christoph Vanberg

AWI DISCUSSION PAPER SERIES NO. 692

October 2020

Promises and Opportunity Cost

Arjun Sengupta, Christoph Vanberg

Department of Economics, Heidelberg University

Abstract

This paper experimentally investigates the hypothesis that promise-keeping behavior is affected by the opportunities that a counterpart foregoes by relying on the promise. We present two motivational mechanisms that could drive such an effect. One is that people dislike causing harm through a promise, and the natural way to measure such harm is to take into account what the counterpart would have received had she not relied on the promise. The other is that people may dislike causing regret in another person. We test these ideas in the context of an experimental trust game. The main treatment variable is the payoff that the first mover forgoes if he "trusts". Consistent with our main hypothesis, we find that an increase in this foregone payoff increases promise-keeping behavior. The experiment is designed to rule out alternative explanations for such an effect. Our evidence suggests that the mechanism driving the effect may involve an aversion to causing regret in others.

1. Introduction

A large and active literature in experimental economics investigates motivational mechanisms involved in the fulfillment of promises (Charness and Dufwenberg, 2006; Vanberg, 2008; Ederer and Stremitzer, 2017). Most of the existing experimental work studies these motivations in the context of trust or investment games (Berg et al., 1995). Such games constitute a natural setting for the investigation of promises because they capture the strategic structure of many economic interactions in which promises are likely to play an important role. In

Preprint submitted to Elsevier *October 7, 2020*

these settings, a promise can serve to convince a potential partner to enter into a cooperative arrangement of some type, e.g. a simple trade in which that party is the first to deliver, or a more complex principal-agent relationship in which the agent is promising to deliver a service. Importantly, a promise can also serve to increase trustworthy behavior.

Within such settings, existing research suggests that at least two basic motivations contribute to explain the effectiveness of promises in increasing trustworthy behavior. The first is that promises create expectations, and people have the desire to fulfill expectations (Charness and Dufwenberg, 2006), or equivalently an aversion to disappointing others. The second is that promises create obligations, and people have a desire to live up to obligations (contractual or otherwise), or equivalently, find it psychologically costly to renege on such obligations (Ellingsen and Johannesson, 2004; Vanberg, 2008). A fundamental difference between these two explanations is that the first explanation involves a concern for the emotional state (or state of mind) of the promisee while the latter does not. Although these theories have often been presented as mutually exclusive alternatives, they are not. And indeed the existing evidence suggests that both types of motivations (expectation-based and obligations-based) contribute to the fulfillment of promises in simple laboratory settings (Ederer and Stremitzer, 2017; Bhattacharya and Sengupta, 2016; Di Bartolomeo et al., 2019).

In this paper, we wish to introduce and investigate an additional consideration which introspection suggests may be important for promise-keeping, but has so far been neglected. The idea is that people may, when considering whether to keep a promise, compare the outcome that they deliver if they do not keep their promise with the outcome that the counterpart could have obtained if the promise had never been made, or if she had not relied on the promise.

To illustrate, consider the following scenario. Bob dreams of opening a pub but needs a partner in order to stem the necessary investments. His friend Ann has substantial savings that she had so far intended to invest in an MSCI World ETF. When Bob tells her of his plans, she tells him she is worried that he won't work hard enough to make the pub profitable for both of them. Bob promises that he will work hard, and Ann agrees to invest in the pub. A few months later, the pub is open and running well. After a particularly long and exhausting night behind the bar, Bob asks himself whether perhaps he should take it easy, even if it means less profit than promised for Ann.

In this situation, existing theories of promise-keeping suggest that Bob would ask himself questions like "how much money will Ann lose if I take it easy rather than working hard like I promised?" and "how much money does Ann expect to earn, and how does this compare to what she will get if I take it easy?" To this, we wish to add questions like "what would Ann have done with her money if she had not invested in the pub? How much money would she have earned if she had not invested, and how does this compare to what she will receive if I take it easy?" Our main hypothesis is that the answer to this question is relevant to Bob, i.e. he will be more inclined to keep his promise, the larger are the opportunity costs that Ann incurred when she relied on his promise (These costs are sunk when Bob makes his decision.).

Within our example, a testable and substantively interesting implication of our theory is that Bob's willingness to work hard will depend on the value of the MSCI World ETF that Ann would have bought if she had not invested in Bob's bar. If the return that she would have earned on the ETF is very high, Bob will feel more obligated to fulfill his promise than if it was low or even negative. That is, Bob's motivation will depend on how the uncertainty regarding Ann's alternative investment is resolved after she has already decided not to invest in that alternative.

We conjecture that there are two (potentially compatible) mechanisms underlying this motivation, which differ in the extent to which they involve an amplificaton of a "direct" concern about the fulfillment of a contractual obligation, or of a concern about the other's state of mind. A "direct" moral cost of breaking a promise has been previously modeled as a fixed cost independent of the foregone payoff of the trustor (Ellingsen and Johannesson, 2004). Indeed, some authors have interpreted a "direct" preference for the fulfillment of obligations as somehow by definition being independent of consequences. However, this is not (or at least need not be) the key distinction. Instead, as noted above, the main distinguishing feature of this type of explanation is that the underlying concern is "direct" rather than working through a first-order concern about another individual's state of mind. However, how morally costly an individual finds breaking his promise may well depend on what the promisee could have done and would have received had she not followed his promise. If the foregone payoff was low, the promisor may find it less morally costly to break his promise compared to a high foregone payoff. A low foregone payoff may allow the promisor to rationalize his decision to break a promise by arguing that the promisee, by relying on his promise, will not lose much compared to her foregone option. But, if the foregone payoff was high, the loss by relying on the promise is larger. Thus one possibility that we wish to investigate is that the opportunity cost another individual has incurred by relying on a promise amplifies the promisor's "direct" moral obligation to fulfill that promise.

The second motivation is that the promisors may face second-order regret aversion. Regret theory suggests that the party who has relied on a promise (Ann in our example) will compare the realized outcome with the outcome that would have occurred, had they not relied on the promise. They experience regret when (and to the extent that)the foregone outcome is better (Bell, 1982; Loomes and Sugden, 1982; Zeelenberg et al., 1996). We conjecture that a promisor may take in account the regret that a promisee would face if a promise is broken. Such a concern might in turn be based on a combination of sympathy and the wish to avoid anger directed at oneself. If Bob breaks his promise to work hard, Ann is likely to regret her decision to rely on him. Moreover, she is likely to feel more anger towards Bob if she learns that the MSCI World ETF that she had been planning to buy performed well. If Bob cares about Ann's regret, he will keep his word more often if her alternative investment performed well. Note that the fundamental difference between these two explanations is again that the second explanation takes into account the emotional state of the promisee while the first does not. We will refer to the first explanation as *moral cost* explanation and the second as the *second-order regret-aversion* explanation.

We test out theory in the context of an experimental investment game. In our game, a first-mover (representing Ann in our example) chooses between an outside option and an investment (Bob's bar in our example). If she chooses the outside option, the interaction ends. If she invests, the second mover chooses between two options which we will refer to as 'Fair' and 'Unfair'. Choosing *Fair* results in an equal payoff to both. Choosing *Unfair* gives the second mover a large payoff and the first mover a small payoff. The first mover is better off investing if the second mover chooses *Fair*, but better off choosing the outside option if the second mover chooses *Unfair*. The second mover always does better if the first mover invests, but receives more money if he chooses *Unfair* over *Fair*.

We want to use this type of game to test i) whether promise-keeping is affected by the payoff that the first mover foregoes by investing, and ii) whether our two motivational mechanisms can explain this behavior. Note that as the foregone payoff varies, a change in promise-keeping behavior could also be explained by an expectations-based account a la Charness and Dufwenberg (2006). Stone and Stremitzer (2017) showed that an increase in investment (and hence, an increase in foregone payoff) leads to a change in higher-order expectations and hence a change in promise-keeping behavior.[1] In order to exclude this channel, our experiment is designed to keep expectations constant even as the foregone payoff varies. More generally, we will argue that the implication of our theory in our experimental design distinguishes it from other (complementary) theories such as distributional preferences, reciprocity, and guilt aversion.

In order to keep expectations constant while varying the outside payoff, we

[1] Charness and Dufwenberg (2006) vary the outside payment in their experiment. However, since their focus was on showing that promises increase trustworthy behavior, they do not compare promise-keeping rates across the two investment games. Furthermore, the outside payoff of the second-mover also varies, creating additional confounds.

modify the investment game in two important ways. First, Ann's outside payoff is unknown to her when she chooses whether or not to invest (like the return to the MSCI world ETF in our example). Specifically, in our investment game, the outside option is a lottery giving either a low or a high payoff. Second, the uncertainty of the lottery is resolved, and Bob learns the outside payoff, before he chooses between the fair and the unfair division.[2] In the design section we will argue that these modifications of the investment game allows us to keep the expectations constant while varying the foregone payoff.

We use a 2x2 design where we vary whether i) Bob can send a one-way message to Ann before she makes a choice (Communication), and ii) whether Ann learns what she has foregone at the end of the investment game after all choices have been made (Feedback). The variation in communication allows us to test whether promise-keeping itself is affected by the foregone payoff. The variation in feedback allows us to identify the motivations behind variation in promise-keeping when the foregone payoff varies. When Ann never learns the foregone outside payoff, a broken promise may induce regret, but the intensity of regret would not depend on the realized outside payoff. Whereas, when she learns about her outside payoff, her regret would depend on what is the realized foregone payoff. This allows us to disentangle whether the variation in outside payoff changes promise-keeping behavior because of change in anticipated regret or through a change in obligation to keep a promise, or both.

We find that the foregone payoff affects the promisor's motivation to choose the *Fair* option, but only when both the parties know the realized foregone payoff. Under the condition where Bob can promise and Ann learns the foregone payoff later, the proportion of *Fair* choices is higher when the foregone payoff is high as compared to when it is low. Moreover, when Ann learns the foregone payoff, the difference in the choice of *Fair* between the high and the low foregone payoff

[2]In general, outside option in investment games represent alternative investment opportunities. However, the alternative investment opportunities are also generally inherently risky, as in the case of our example.

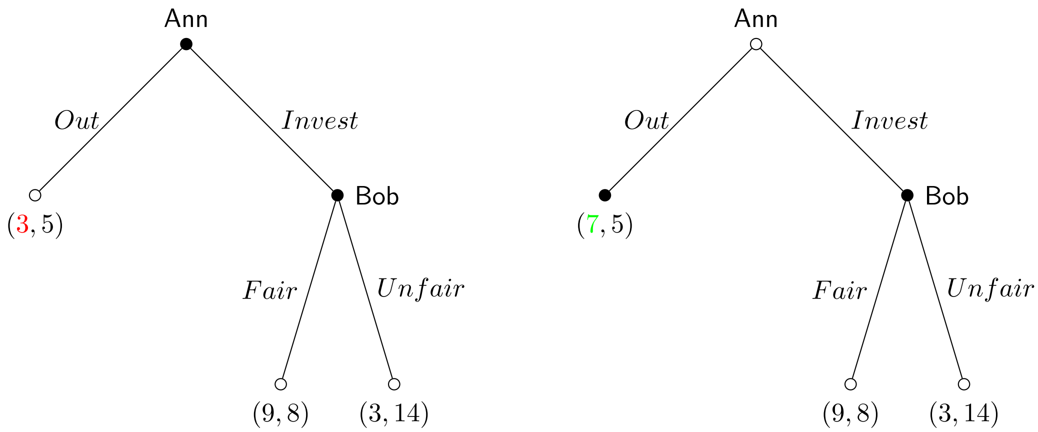
is significantly greater when Bob can promise than when Bob cannot promise. When Ann learns the foregone payoff, a change in promise-keeping behavior with the foregone payoff can be attributed to both second-order regret aversion and variation of moral cost . However, we find that when Ann never learns the foregone outside payoff, the proportion of fair choice is statistically identical between a high and low foregone payoff, both when Bob can promise and cannot promise. These observations suggest that the "direct" moral cost associated with breaking a promise is independent of the realized foregone payoff, and that second-order regret-aversion is an important reason behind promise keeping

The rest of the paper is arranged as follows. In section 2, we introduce the design, lay down the predictions of our theories in terms of our design, and explain why other social preference models and expectation based guilt cannot explain a difference in promise-keeping behavior when the outside payoff varies. In section 2.3, we introduce the experimental procedure. In section 3, we present the results of the experiment. And finally, section 4 concludes.

2. Design and Procedure

2.1. Design

Figure 1 presents two versions of an investment game that will play a role in our design. Ann, who is an investor, moves first and decides whether to *Invest* in Bob's project or choose *Out* (*Out* represents all other investment opportunities.). Choosing to invest increases the total available amount to be shared between Bob and Ann. However, if Ann decides to invest, Bob decides the amount he wants to return. Bob can return a *Fair* amount, in which case both players receive similar payoffs, or, Bob can choose *Unfair*, in which case he keeps most of the money. In the investment game described, Bob has both an incentive to promise to choose *Fair* if he can communicate with Ann, and a monetary incentive to choose the *Unfair* amount if Ann invests. This, in turn, creates a disincentive for Ann to invest.



(a) Low Outside Payoff

(b) High Outside Payoff

Figure 1: Investment Games

Note that the games presented in Figure 1 differ only in the payoff that Ann would receive if she chooses *Out.* An implication of our theory is that Bob's motivation to keep a promise to choose *Fair* will be affected by this outside payoff. Specifically, the theory predicts that, ceteris paribus, Bob will be more likely to keep a promise to choose *Fair* in the game on the right, where Ann has foregone a payoff of 7, than in the game on the left, where she has foregone only 3. The underlying motivations, recall, are that (i) Bob's moral cost of breaking a promise may depend on the extent to which a broken promise harms Ann relative to what she would have received had she not relied on his promise and/or (ii) Bob may be concerned about the intensity of the regret and anger that Ann feels when he breaks his promise.

Although we do not intend to present a fully developed formalization of our theory, it may help to provide a sketch of such a theory in a form that is, for the moment, specific to our setting. Denote Ann's payoff from the outside option by α^k, where $k = H$ or L and $\alpha^H = 7$ and $\alpha^L = 3$. Then if Bob has promised to choose *Fair*, his payoff from choosing *Fair* is $u_B(9, 8)$ where $u_B(x_A, x_B)$ is a purely consequentialist value that Bob attaches to the distribution (x_A, x_B)

(including 'non-standard' distributional preferences). And his payoff from breaking his promise by choosing *Unfair* is $u_B(3,14) - \eta(\alpha^K - 3) - \mu(\alpha^K - 3)$, where η represents the moral cost of breaking a promise while μ represents the effect of second-order regret aversion. Note that both η and μ are increasing functions representing Bob's disutility from the 'harm' he is causing *relative to Ann's outside option.*[3] Then Bob will choose to keep his promise if $\eta(\alpha^K - 3) + \mu(\alpha^K - 3) > u_B(3,14) - u_B(9,8)$. For a given α^K, the truth value of this condition will depend on Bob's distributional preferences with respect to the possible outcomes. If Bob is sufficiently inequity averse or altruistic, the right hand side of the inequality may be negative, so that he would keep his promise even if $\eta(\alpha^K - 3) + \mu(\alpha^K - 3) = 0$. If he is purely selfish, he will do so only if $\eta(\alpha^K - 3) + \mu(\alpha^K - 3)$ is sufficiently large. In practice, we would expect there to be some heterogeneity, both with respect to the relevant distributional preferences, as well as with respect to the motivation we are discussing. Therefore the theory does not provide a specific prediction about an isolated individual choice. However, it does make the comparative static prediction that an increase in α^K should lead to an increase in the proportion of individuals for whom the condition is satisfied, and therefore in the proportion of subjects who choose to keep their promise.

A comparison of promise-keeping behavior across investment games in figure 1 which only differs in the outside payoff is not sufficient to test our theory, since other theories would also predict a difference in promise keeping rates. For example, Ann's choice of *Invest* is likely to induce higher second-order beliefs when the outside payoff is high, such that the theory of expectations-based guilt aversion predicts the same effect as our theory.[4] The modified investment game

[3]Specifically, η and μ are functions of a distance metric, $d(\alpha^K, x_A)$ which itself is a function of α and x_A, $\eta(d(\alpha^K, x_A))$ and $\mu(d(\alpha^K, x_A))$. Both η and μ are increasing function of the distance metric $d(\alpha^K, x_A)$. To keep our exposition simple, in our example we use $d(\alpha, x_A) = \alpha^K - x_A$.

[4]Ann's choice to invest acts as a signal of Ann's minimum expected return from Bob, since otherwise she would not have invested. If Ann invests, the minimum expected returns from

we present not only keeps the expectations constant while varying the outside payoff, but also allows us to disentangle the two motivations we suggest can explain a difference in promise-keeping behavior.

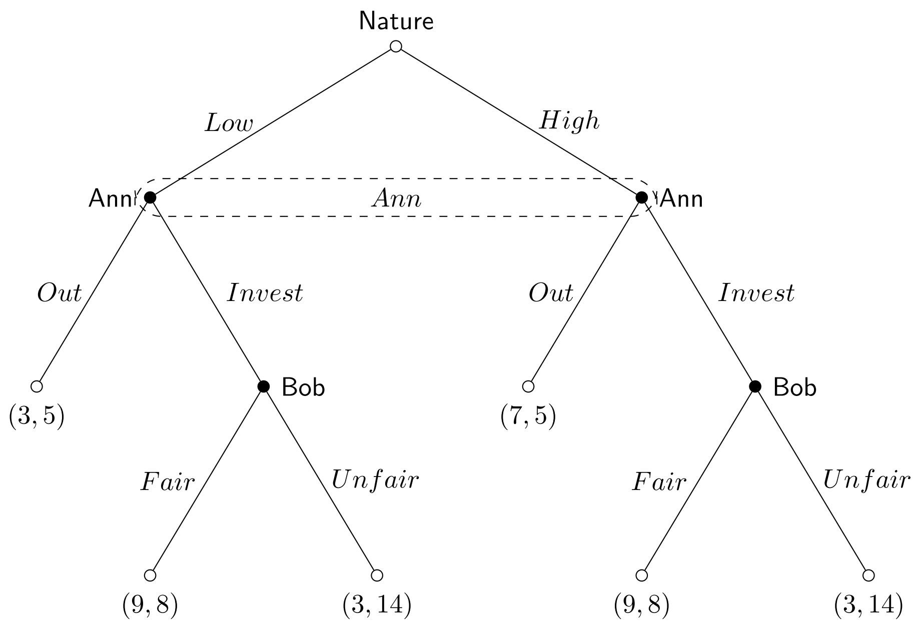


Figure 2: Modified Investment Game

Figure 2 represents our modified investment game, which involves uncertainty in the outside payoff. Nature moves first and determines the payoff Ann would receive if she chooses *Out*. Nature draws either a *High* or a *Low* outside payoff with equal probability. After Nature draws the outside payoff, Ann decides whether to *Invest* or stay *Out* without learning Nature's draw. If Ann chooses *Out*, she receives a payoff of €3 when the outside option is *Low*, and she receives €7 when it is *High*. Bob's payoff if Ann chooses *Out* is €5 in both states (equal to the expected payoff of Ann's outside option). After Ann makes a decision,

investment is €3 in figure 1a and €7 in figure 1b. Expectations-based explanation of promise-keeping would predict that promises will be kept more often when the foregone payoff is €7 than when it is €3.

Bob learns the foregone payoff of Ann (i.e. Bob learns whether the outside option is *High* or *Low*). If Ann chose to invest, he chooses either a *Fair* or an *Unfair* division of the total available amount of €17. If Bob Chooses *Fair*, then Ann receives €9 and Bob receives €8. If Bob Chooses *Unfair*, then Ann receives €3 and Bob receives €14.

Since the uncertainty regarding Nature's move is resolved only after Ann makes her choice, Ann's *expectation* concerning Bob's behavior is unaffacted by the realized value of the outside option. Bob learns the foregone payoff before he makes a choice, but he knows that Ann was not aware of it when she made her choice. Therefore we would expect his *second-order expectation* to be independent of the realized foregone payoff. Thus we expect both first- and second-order expectations to be independent of the realized foregone payoff.

If Bob could communicate, expectations will increase, however for the same logic, the first and the second-order expectations will remain unaffected by the realized foregone payoff. Moreover, note that Bob communicates to Ann before she makes a choice and hence before the uncertainty is resolved. This additionally allows us to keep the motivation to promise independent of the realized foregone payoff.[5]

To understand the effect of the foregone payoff on promise-keeping, and to disentangle the two motivations which can explain changes in promise-keeping behavior, we vary whether (i) Bob can send a message to Ann before Ann makes an investment decision (Communication) or not (No Communication) and (ii) whether Ann learns her realized foregone outside payoff at the end of the interaction (Feedback) or not (No Feedback). Table 1 shows all 4 treatments. Our main outcome variable of interest is Bob's choice of *Fair*. In Table 1, F represents the proportion of *Fair* choice. The superscript indicates whether the foregone payoff is *H*igh or *L*ow. The subscripts represent the treatment (e.g. 'ncf' = No Communication and Feedback). We now layout our hypotheses in terms of our

[5]Knowledge of the exact foregone payoff before Bob makes a promise could affect both his motivation to make a promise and the content of the message.

main variable of interest, proportion of *Fair (F)* choice.

Table 1: Treatments

	No Feedback	Feedback
No Communication	F^H_{ncnf}, F^L_{ncnf}	F^H_{ncf}, F^L_{ncf}
Communication	F^H_{cnf}, F^L_{cnf}	F^H_{cf}, F^L_{cf}

ncnf = no communication, no feedback; ncf = no communication feedback; cnf = communication no feedback; cf = communication feedback. For example, F^H_{cf} represents the proportion of Fair choice when the foregone payoff was high in the Communication Feedback treatment.

2.2. Hypotheses

Our first hypothesis concerns the effects of the outside option on promise-keeping in a context where the second mover (Ann) learns the value of the outside option following the interaction. Recall that in this case, we conjecture that the outside option may affect promise keeping for two reasons. First, Bob may be concerned about the intensity of regret and anger that Ann will experience if he breaks a promise to choose *Fair*. This intensity, $\mu(\alpha^K - 3)$, is likely to be increasing in the value of the outside option. Second, Bob may be concerned about the harm that he imposes on Ann, and we conjecture that Bob measures "harm" by comparing Ann's payoff to her foregone payoff, $\eta(\alpha^K - 3)$. Therefore, the utility of Bob when he breaks his promise can be written as $u_B(3, 14) - \eta(\alpha^K - 3) - \mu(\alpha^K - 3)$. Note that both these motivations work in the same direction when Ann learns the foregone payoff (and has therefore a reason to regret her decision differently based on the foregone payoff.).

Naturally, it is possible that Bob entertains similar thoughts even in the absence of a promise. That is, Bob may be more likely to chose *Fair* when the foregone payoff is large, even in our "No Communication" condition. To test the idea that the foregone payoff increases the effect of promises, we test the Hypothesis that the difference in the proportion of *Fair* choices between the low and the high foregone payoff will be larger when individuals can communicate (and presumably promises) than when individuals cannot communicate.

Hypothesis 1: *The difference in the proportion of fair choice between the high and the low foregone payoff in the Feedback Communication treatment will be larger than the difference in the Feedback No Communication treatment.* $H_a : [(F^H_{cf} - F^L_{cf}) - (F^H_{ncf} - F^L_{ncf})] > 0$

Without feedback, Ann never learns what she had foregone when she invested in Bob. Thus, when she observes a broken promise, she may regret her decision, but her regret (and anger towards Bob) would not depend on the foregone payoff. Consequently, Bob's second-order regret should not be dependent on the realized foregone payoff. This means that Bob's utility function is represented by $u_B(3, 14) - \eta(\alpha^K - 3) - [\frac{1}{2}\mu(\alpha^k - 3) + \frac{1}{2}\mu(\alpha^J - 3)]$ and for a given α^K he will choose to keep his promise if $\eta(\alpha^K - 3) + [\frac{1}{2}\mu(\alpha^k - 3) + \frac{1}{2}\mu(\alpha^J - 3)] > u_B(9, 8) - u_B(3, 14)$. However, note that the difference in promise keeping rate in the No Feedback condition when the foregone payoff varies is only predicted by a change in moral cost, $\eta(\alpha^K - x_A)$, as the level of second order regret, $[\frac{1}{2}\mu(\alpha^k - 3) + \frac{1}{2}\mu(\alpha^J - 3)]$, remains the same regardless of whether the foregone payoff was low or high. Therefore, any variation in promise-keeping due to foregone payoff would be explained only by Bob being *directly* concerned about the harm he is causing as compared to Ann's outside option (moral cost mechanism). Therefore, if Bob's cost of breaking a promise is directly affected by the foregone payoff, we should find support for the following Hypothesis.

Hypothesis 2: *The difference in the proportion of fair choice between the high and the low foregone payoff in the No Feedback Communication treatment will be larger than the difference in the No Feedback No Communication treatment.* $H_a : [(F^H_{cnf} - F^L_{cnf}) - (F^H_{ncnf} - F^L_{ncnf})] > 0$

In the Feedback condition, the difference in the promise-keeping rate could be driven by a change in the moral cost of breaking a promise or a change in second-order regret when the foregone payoff varies. Without feedback, only a change in the moral cost of breaking a promise can explain a difference in promise-keeping. This allows us to decompose the effect of foregone payoff on promise-keeping due to variation in the cost of breaking a promise and second-order regret aversion. The additional difference in promise-keeping rate between the high and the low foregone payoff in the Feedback condition over and above the difference in promise-keeping in the No Feedback condition would capture the effect of second-order regret aversion on promise-keeping. Thus, if second-order regret aversion affects promise keeping, we expect support for the following Hypothesis.

Hypothesis 3: *The difference in difference of the proportion of choice of Fair between the Communication Feedback treatment and the No Communication Feedback treatment will be larger than the difference in difference between the Communication Feedback treatment and the No Communication Feedback treatment.* $H_a : [(F^H_{cf} - F^L_{cf}) - (F^H_{ncf} - F^L_{ncf})] - [(F^H_{cnf} - F^L_{cnf}) - (F^H_{ncnf} - F^L_{ncnf})] > 0$

Our design rules out other social preference models, along with expectation based guilt aversion, as an explanation for the variation in promise-keeping. Outcome-based social preference models are based on comparisons of the final distribution of payoffs (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000). In our design, if Ann chooses *In*, the final distribution of payoffs is unaffected by the realized *Out* payoff. Thus, outcome-based social preference models would predict the same promise-keeping behavior regardless of the realized foregone payoff.

Intention based reciprocity models cannot explain differences in Bob's behavior based on foregone payoff (Levine, 1998; Falk and Fischbacher, 2006; Battigalli and Dufwenberg, 2009). The standard theory of intention-based reciprocity has a counter-intuitive prediction. It predicts negative reciprocity from Bob if Ann chooses *Invest* because Ann's decision to invest can be interpreted by Bob as Ann wanting him to choose *Fair*. Since choosing Fair gives lower payoff to Bob than choosing *Unfair*, Ann's act of investment can be viewed as an unkind act by Bob. However, the negative reciprocity is independent of the realized foregone payoff. It is also possible that Bob views an act of investment as a kind action as it allows him to receive a higher payoff compared to the outside option. Even so, Bob should not reciprocate differently to Ann's kindness based on the foregone outside payoff as Bob's payoff is the same across the high and the low foregone options.

Finally, the design also rules out image concern as a motivation for differential promise-keeping behavior. Image concern theory posits that individuals care about how they appear to others and themselves (Bénabou and Tirole, 2006). Deviating from socially acceptable norms of behavior could lead to both a fall in social and self-image. The norm of promise-keeping suggests that breaking a promise could lead to a loss in both social and self-image (Schütte and Thoma, 2014; Grubiak et al., 2019). If image concern matters, then, to avoid a negative image from being branded as a liar, Bob would keep his promise. However, the loss of image is dependent on the promised choice and the final choice, but not on the foregone payoff. Moreover, in our experiment, if Ann chose *In*, there is no uncertainty in Bob's choice. A broken promise is always detectable.

2.3. Experimental Procedure

The experiment was conducted at the Alfred-Weber Institute laboratory at Heidelberg University. Subjects were students of the university and were recruited using Hroot and Sona. The design was between subjects, each subject participated only in one of the four different treatments. The experiment was programmed using Ztree (Fischbacher, 2007). We explain in detail the procedure for the Communication condition with Feedback.[6] The procedure for the other three treatments was identical except for the treatment variation.

Subjects were randomly assigned to a computer terminal in the lab. The instruction for the experiment was displayed to the subjects on the computer screen. The instruction appeared over several pages and the subjects could move forward and backward between the pages while reading the instruction. They were also handed out written instruction. When all subjects indicated that they had finished reading the instruction by pressing the "done reading" button which

[6] The subjects were given a German version of the instruction. Appendix A contains the English version of the instruction.

appeared on the last page of the instruction, the experiment proceeded to the main stage.

At the beginning of the experiment, each subject was randomly assigned to a matching group of size six. All interactions took place within the matching group. Half of the subjects in a matching group were assigned the role of Bob and the other half to the role of Ann. Their roles remained the same for the entire experiment. There were nine rounds. In each round, Ann and Bob were randomly matched within their matching group and they played the game represented in Figure 2.[7]

Within a round, Bob first had an opportunity to send a free form message to Ann. If he did not want to send a message, he could leave the message box blank. After Ann observed Bob's message, she decided whether to *Invest* or stay *Out* without knowing the realized outside payoff. We used the strategy method to elicit Bob's choice. Bob had to decide whether to choose the *Fair* or the *Unfair* option assuming that Ann has invested.[8] Before he made his choice, the uncertainty about the foregone payoff was resolved. Bob flipped a virtual coin that had an equal probability of landing Head or Tails and observed the outcome of the coin flip. Had Ann chosen *Out*, if the outcome of the coin flip was Heads, Ann would have received the low payoff, and if it was Tails, she would have received the high payoff. The outcome of the coin flip did not affect Bob's payoff had Ann chosen *Out*. Bob then chose between the *Fair* and the *Unfair* option. After they made their choice, they stated their beliefs. Ann stated on a scale of 1-5 how likely Bob is going to choose the *Fair* option, with 1 indicating that she is sure that Bob would choose the *Unfair* option and 5 indicating that she is sure that Bob would choose the *Fair* option. Bob was asked to guess the

[7]Since each matching group had six subjects, and the game was repeated nine times, each individual in the role of Ann and Bob interacted more than once. The pairing was constructed such that no same individuals met twice in a row. Learning is limited in our experiment as neither players received any information about others choice until the end of the experiment.

[8]Action labels used in the experiment were neutral. *Unfair* choice was labeled as Option 1 and *Fair* choice was labeled as Option 2.

number Ann had stated. The belief procedure was incentivized using a modified scoring rule used in Vanberg (2008). After the beliefs were stated, the round ended, and a new round began. No information about others choices was given to players between rounds. Choice information and outcomes were displayed for all nine rounds at the end of the experiment. Furthermore, in the feedback condition, Ann also learned her foregone outside payoff.[9]

In our experiment, subjects received a show-up fee of €3. If Ann chose *Out*, then Ann received a payoff of €3 or €5 and Bob received €5. If Ann chose *In* and Bob chose *Fair*, Ann received €9 and Bob received €8. We used three different payoff constellations for Ann and Bob if Ann chose *In* and Bob chose *Unfair*. Bob received either €13 or €14 or €15 and Ann received the remaining of the €17 that Bob was splitting. In a round, only one of the three *Unfair* payoffs were used, but the *Unfair* payoff varied across rounds. The exact payoff structures were common knowledge before Ann and Bob made any decisions. They also knew that *Unfair* choice payoff would change across rounds. The ordering of the *Unfair* choice payoffs over rounds were randomized across sessions.

3. Results

In this section, we present the results from the experiment. Our main variables of interest are (i) the proportion of choice of *Fair* and (ii) the difference in the proportion of choice of *Fair* between the high and the low foregone payoff. In section 3.1.1, we present results from the Feedback treatments and analyze whether variation in foregone payoff affects promise-keeping behavior. In section 3.1.2, we investigate whether the variation in promise-keeping is due to variation in preference for promise-keeping when foregone payoff varies. In section 3.1.3,

[9] In the Feedback condition, Ann always learned the outside payoff, regardless of her choice of *In* or *Out*. However, if she chose *Out*, she did not learn what she would have received had she chosen *In*. Though we have information about Bob's choice because of the strategy method elicitation, this information structure is consistent with the sequential nature of the game. Bob would not have made a chosen if Ann chose *Out*. Hence, there is no counter-factual information regarding Bob's choice if Ann chose *Out*.

Table 2: Messages

	Strong Promise	Weak Promise	Promise	Empty
No Feedback	42.39%(103/243)	19.75%(48/243)	62.14%(151/243)	37.86%(92/151)
Feedback	63.79% (155/243)	11.11% (27/243)	74.90%(182/243)	25.10% (61/243)

we decompose the effect of foregone payoff on promise-keeping due to variation in second-order regret and variation in the cost of breaking a promise.[10][11]

We use non-parametric tests and linear probability regression model to test our hypotheses. All our statistical analyses are done at the matching group level, i.e., each independent observation is the proportion of fair choice at the matching group level conditional on whether the foregone payoff was low or high.[12] We had eighteen participants in each session. We ran three sessions for each treatment. Since we analyze data at the matching group level, for each treatment, we have nine independent observations each for Ann and Bob.

3.1. Bob's Choice

3.1.1. Feedback

In the communication condition, Bob could send a free-form message to Ann before she made a decision. The messages were coded into three categories. They were coded as Strong Promise if Bob clearly expressed his intent to choose the fair option. They were coded as Weak Promise if the message suggests a choice

[10] Since our main interest is in Bob's promise keeping behavior, the statistical analysis of Bob's second-order beliefs and Ann's choices and beliefs are presented in Appendix B and Appendix C, respectively.

[11] The statistical analysis is done by pooling choices across all rounds. The *Unfair* payoff for Ann and Bob varied across rounds. As the *Unfair* payoff increases, Bob has a stronger incentive to renege on his promise. However, foregone payoff should still affect promise-keeping behavior differently. We varied the *Unfair* to pick up switch points for Bob, but given the size of the data, we are under-powered to analyze behavior based on the unfair payoff. We present the averages in Tables E.18 and E.19 in the appendix.

[12] Matching groups were groups of size 6 that were randomly formed before the start of the experiment in a session. In a matching group, three of them were randomly assigned the role of Ann and the other three the role of Bob. The investment game pairs were formed by randomly pairing players in the role of Ann and Bob within a matching group.

of the fair option. The rest of the messages were classified as Empty Talk.[13] Table 2 shows the number of messages in each category. For our statistical analyses, we consider both a weak and a strong promise as a promise.[14] We first look at the proportion of fair choice and its difference in the communication condition for the complete data and then we look at the proportion of fair choice and its difference conditional on a promise being sent.

Table 3: Proportion of *Fair* Choice(Feedback)

	Low	High	Difference
No Communication	27.19%	18.73%	-8.46 ($p = 0.02$)
Communication	36.09%	54.64%	18.55 ($p = 0.02$)
Promise	43.69%	59.63%	15.94 ($p = 0.06$)

Numbers in the table represent the proportion of *Fair* choice. p- values reported are from Sign-Rank test.

Table 3 represents the proportion of fair option chosen by Bob in the Feedback conditions. Under the communication condition, the proportion of fair choice is 36.09% and 54.64% under the low and the high foregone payoff respectively. The difference in the proportion of the choice of *Fair* between the high and the low foregone payoff is 18.55%. Figure 3 shows the difference in the proportion of *Fair* choice at the matching group level between the high and the low foregone payoff. The observations are arranged in ascending order and the rank of observation is signed according to the sign of the observation. If the differences were randomly distributed, then the observations in Figure 3 would have been equally distributed around zero. As can be seen, in the communication condition, the difference in choice is positive for most of the matches. The median value of the difference is 14.29% in the Communication condition (vertical blue line) and is significantly

[13]Messages were coded independently by one of the authors and a research assistant. The coders disagreed on 22 out of 486 observations, mostly on Weak Promises. The discrepancies were resolved together.

[14]Our results do not change if we only restrict our data to Strong Promise.

different from zero (Sign-Rank, $p = 0.02$).

In the Feedback condition with communication, 74.90% (182/243) messages were coded as a promise. If we only consider observations where a promise was sent, the proportion of fair choice is 43.69% and 59.63% under the low and the high foregone payoff respectively. The matching group level difference in choice of fair option between the high and the low outside option is 15.94%. The median value of the distribution of differences is 16.66% and it is significantly different from 0 (Sign-Rank, $p = 0.06$).

Result 1a: *We find that when individuals can promise and Ann learns the foregone payoff, the proportion of choice of fair is higher under the high compared to the low foregone payoff.*

These observations are consistent with the conjecture that foregone payoff affects promise-keeping behavior, but, to conclude that foregone payoff indeed affects behavior due to the interaction with communication, we need to rule out a similar behavioral pattern in the No Communication condition.

When Bob could not communicate, the proportion of *fair* choice was 27.19% and 18.73% under the low and the high outside option respectively. There is an 8.46% point drop in the choice of *Fair* moving from the low to the high foregone payoff. As can be seen from Figure 3, the difference in choice is negative for most of the matches. The median of the difference in fair choice is significantly different from 0 and is negative (Sign-Rank, $p = 0.02$) contrary to our expectations of a positive or zero difference between the high and the low outcome. This suggests that Bobs who cannot promise were likely to choose the fair option less when the foregone payoff was high compared to low.

This result was unexpected. We conjecture that since Bob chooses without knowing exactly what Ann has chosen, he might think of his choice as a choice between lotteries. When he observes a high outside option payoff instead of a low outside option payoff, he might feel that Ann has a possibility to earn a higher amount if she has chosen *Out*, thus making it easier for him to choose the lottery which implements the *Unfair* choice if *Invest* is chosen. Holding every

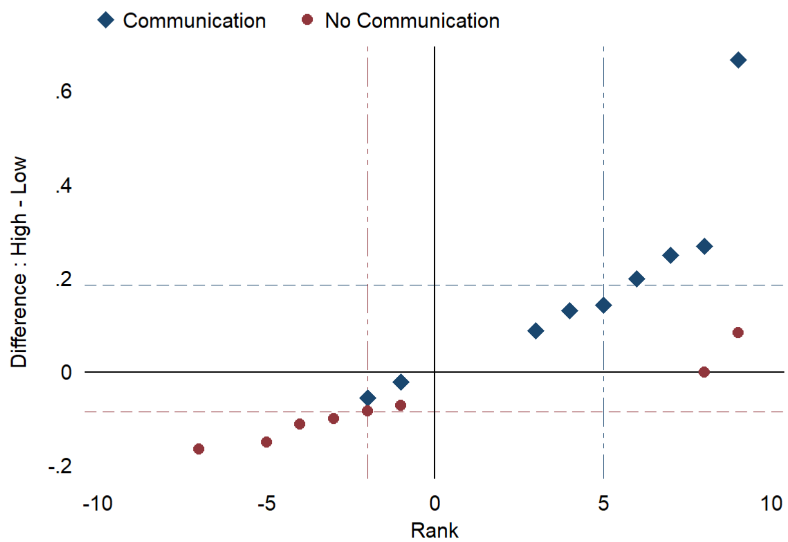


Figure 3: Distribution of differences (Feedback)

Note: Each observation represents a single matching group. The vertical axis measures the difference in the proportion of fair choice between high and low foregone payoff. The observations are ranked in ascending order within a treatment. The horizontal position of each observation corresponds to the signed rank of that observation, where the sign of the rank is positive if the difference is positive and negative if the difference is negative. The dashed horizontal lines indicate the average of the observations. The dashed vertical lines mark the median observations.

other motivation constant, this would mean that Bob would be more likely to choose the *Fair* option under the low than the high foregone payoff. As we use the strategy method in our experiment, this motivation is present in all our treatments. Our difference in difference analysis therefore takes care that this motivation does not affect our results.

Result 1b: *We find that when individuals cannot communicate and Ann learns the foregone payoff, the proportion of the choice of fair is lower under the high compared to the low foregone payoff.*

To understand the effect of foregone payoff on promises, we need to compare the difference in the proportion of fair choice across the Communication and the No Communication condition. The average difference in difference between the Communication and No Communication condition is 27.03% points. Figure 3 shows that while the matching group level differences in the Communication condition are mostly positive, in the No Communication condition, they are mostly negative. The distributions of the differences are significantly different between the Communication and the No communication condition (Rank Sum, $p < 0.001$). If we restrict our data only to promisors, then the difference in difference between the Communication and No Communication condition is 24.41% points and the distributions of the differences are also significantly different between the two (Rank Sum, $p = 0.004$). This suggests that the positive differences observed in the communication condition are indeed related to an interaction effect of promises and foregone payoff.

Further evidence of Hypothesis 1 is provided by linear probability model in column I of Table 5.[15] The dependent variable is Bob's choice with 1 indicating a *Fair* choice and 0 otherwise. $High$ takes the value 1 if the outside payoff was high and 0 otherwise. $Communication$ takes the value 1 if Bob could send a message and 0 otherwise. $HighXCommunication$ is the interaction term between $High$ and $Communication$. The coefficient of $High$ is negative, indicating that when individuals cannot communicate, they choose the fair option more often under the low than under the high outside option. The coefficient on communication is also positive, indicating that there is an increase in the choice of *Fair* moving from No communication to Communication under the low foregone payoff, but, it is not significant. Finally, our variable of interest, the interaction term between $High$ and $Communication$, which measures the difference in difference in choice of fair option between high and low foregone payoff moving from No communication to Communication condition, is positive and significant, consistent with hypothesis 1. The qualitative results do not change if we restrict observations in the Communication condition to only those

[15] To check for robustness, we also use a logistic and probit model with clustering at the matching group level. The results are reported in Table D.14 and D.15 and are qualitatively similar to the results reported in the main text.

who have sent a promise (Table D.12).

Result 1: *In line with Hypothesis 1, we find that if an individual promises, then an increase in foregone payoff leads to an increase in proportion of fair choice.*

3.1.2. No Feedback

Table 4: Proportion of *Fair* choice (No Feedback)

	Low	High	Difference
No Communication	28.46%	26.37%	-2.09 ($p = 0.51$)
Communication	58.05%	50.5%	-7.55 ($p = 0.25$)
Promise	75.88%	71.02%	-4.84 ($p = 0.67$)

Numbers in the table represent the proportion of *Fair* choice. p- values reported are from Sign-Rank test.

In the last section, we found that higher foregone payoff leads to higher promise-keeping. This behavioral pattern can be explained by both *second-order regret-aversion* and a variation in the *moral cost* of breaking a promise. In this section, we analyze the data from the No Feedback conditions, where regret is no longer dependent on the foregone payoff and any variation in choice due to variation of the foregone payoff can be attributed to a change in the *moral cost* of breaking a promise. Table 4 represents the proportion of choice of *Fair*. When Bob could communicate, the proportion of the choice of fair option is 58.50% and 50.05% under the low and the high outcome respectively. There is a decrease of 7.55% point in the choice of *Fair* moving from low to high foregone payoff. However, as can be seen from Figure 4, in the communication condition, the matching group level differences are equally likely to be positive, negative or very close to zero. The median of the distribution of the differences does not differ significantly from 0 (Sign-Rank, $p = 0.51$).

In the No Feedback condition with communication, 62.14% (151/243) of the messages sent were a promise. If we restrict observations where only promises are sent, the proportion of choice of fair option is 75.88% and 71.02% under

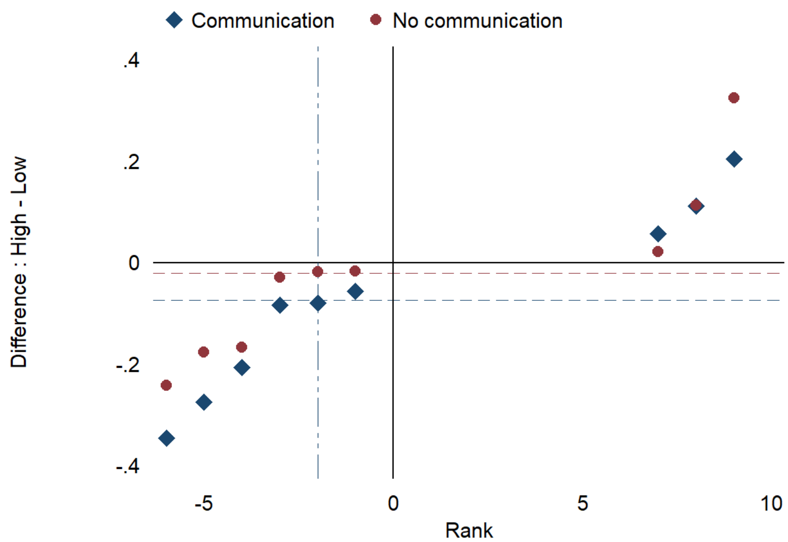


Figure 4: Distribution of differences (No Feedback)

the low and the high outcome respectively. The median of the distribution of the differences does not differ significantly from 0 (Sign-Rank, $p = 0.67$). The lack of any difference in choice suggests that the cost of breaking a promise is independent of the harm done relative to the foregone payoff.

Result 2a: *We find that proportion of fair choice does not differ between high and low outcomes when Bob can promise and Ann never learns the foregone payoff.*

To formally come to this conclusion, we now compare the difference in choice of *Fair* between the high and the low foregone payoff in the Communication condition to the No Communication condition. When Bob could not communicate, the *Fair* option is chosen 28.46% and 26.37% under the low and the high foregone payoff. There is a small decrease of 2.09% in choice of *Fair* moving from low to high foregone payoff, but the median of the distribution of the difference does not differ significantly from 0 (Sign-Rank, $p = 0.51$). Comparing the difference in choice of *Fair* across the high and low foregone payoff between the Communication and No Communication condition, we cannot reject the null that the distributions of the differences are identical across the two (Rank Sum, $p = 0.45$). If we restrict observations only to promises in the Communication condition, we still cannot reject the null that the distributions of the differences are identical across the Communication and the No Communication condition (Rank Sum, $p = 0.75$)

Result 2b: *We find that proportion of fair choice does not differ between the high and the low outcomes when Bob cannot communicate and Ann never learns the foregone payoff.*

Further evidence of this observation is provided by regression specification II in Table 5. The only coefficient that is significant and positive is the coefficient of $Communication$. The coefficient of $High$ and the interaction term between $High$ and $Communication$ is not significantly different from zero. The results are similar if we restrict observations in the communication condition to only those who have sent a promise (Table D.12). This observation tells us that promise-keeping behavior is unaffected by the foregone payoff if the investor never learns what she had foregone. This suggests that the *moral cost* of breaking a promise is independent of the harm done relative to the foregone payoff, consistent with the constant cost model of commitment by Ellingsen and Johannesson (2004).

Result 2: *Contrary to Hypothesis 2, we find that when Ann never learns the foregone payoff and Bob promises, then the proportion of fair choice is unaffected by the foregone payoff.*

3.1.3. Evidence For Second-Order Regret Aversion

In section 3.1.1 and 3.1.2, we found that the promise-keeping rate depends on the foregone payoff in the Feedback condition, but not in the No Feedback condition. This suggests that promise-keeping is driven by second-order regret and not a change in the moral cost of reneging on a promise. In this section, we present a formal statistical result where we decompose the effect of foregone payment on the *moral cost* of reneging on a promise and *second-order regret-*

Table 5: Linear Probability Model

	I Feedback	II No Feedback	III Combined
High	-0.087** (0.035)	-0.032 (0.058)	-0.032 (0.057)
Communication	0.084 (0.096)	0.306*** (0.075)	0.306*** (0.074)
High*Communication	0.248*** (0.069)	-0.064 (0.083)	-0.064 (0.082)
Feedback			-0.016 (0.076)
Communication*Feedback			-0.222*** (0.120)
High*Feedback			-0.055 (0.067)
High*Communication*Feedback			0.312*** (0.107)
Constant	0.271*** (0.060)	0.287*** (0.048)	0.287*** (0.047)
Number of observations	486	486	972
R^2	0.62	0.60	0.60

Standard errors in parentheses. All specifications are clustered at the matching group level. $^{*}p < .10$, $^{**}p < .05$, $^{***}p < .01$.

aversion.

Specification III in Table 5 reports results from the combined data. The variable $Feedback$ takes the value 1 when feedback was provided to Ann and 0 otherwise. The model includes interaction terms between $Feedback$ and $Communication$, $Feedback$ and $High$, and a triple interaction term between $High$, $Feedback$ and $Communication$. The interaction term of $High$ and $Communication$ is the effect of the foregone payoff on the promise-keeping rate in the No Feedback condition. Therefore, it captures variation in promise-keeping rate due to a change in the moral cost of reneging on a promise as the foregone payoff varies. The triple interaction term captures the variation in promise-keeping rate in the Feedback condition after controlling for the variation in promise-keeping under the No Feedback condition. Thus, it captures the effect of second-order regret on promise-keeping as foregone payoff varies.

We find that the coefficient of the triple interaction term is positive and significant. As expected, the size of the coefficient is similar to the size of the coefficient on $HighXCommunication$ in specification I.[16] Furthermore, the interaction term between $Communication$ and $High$ in specification III is not significant. This is consistent with our observation that the moral cost of breaking a promise does not vary with the foregone payoff. These two observations together confirm that when there is no feedback, promise-keeping is unaffected by the realized outside payoff, but significantly increases with an increase in foregone payoff when there is feedback. This suggests that any difference in promise-keeping rate due to variation in the foregone payoff is due to second-order regret aversion and not due to variation in the moral cost of reneging on a promise. Furthermore, the coefficient on the interaction term between Communication and Feedback is negative and significant. This is in line with our observation that promise-keeping falls with feedback if the foregone option was low. However, feedback does not affect the promise-keeping rate if the foregone payoff was high.[17] Qualitative results are the same if we restrict attention to only those who promised (Table D.12).

Result 3: *After accounting for the change in promise-keeping due to a change*

[16]Note that the coefficient of the interaction term $HighXCommunication$ in specification I captures both the average effect of the foregone payoff on promise-keeping due to a variation in second-order regret and a change in the cost of reneging on a promise.

[17]One may suggest that observing no difference in promise-keeping rate between Feedback and No Feedback under the high foregone payoff speaks against the theory of second-order regret aversion. However, this is not necessarily true. In the No Feedback condition, Ann never learns the foregone payoff. Therefore, how much regret Bob thinks Ann would suffer depends on what Bob thinks Ann believes about the foregone payoff. No difference in choice for the high foregone payoff is consistent with Bob believing that under No Feedback, Ann believes the foregone option was high.

in the foregone payoff from Low to High in the No Feedback condition, we find that the change in promise-keeping due to a change in the foregone payoff in the Feedback condition is still positive and significant, which is captured by the triple interaction term in our regression.

4. Discussion

Our paper contributes to the literature on communication and investment games in two ways. First, we propose and test the hypothesis that people will be more likely to keep a promise, the large the opportunity costs that their counterpart incurred by relying on it. We find support for this hypothesis. Our findings are consistent with the idea that the underlying mechanism involves second-order regret aversion as a motivational factor that drives people to keep their promise. To the best of our knowledge, our paper is the first to show that promise-keeping behavior is supported by second-order regret aversion. However, we do not find evidence that the moral cost of breaking a promise is itself affected by the harm that a broken promise imposes relative to the foregone payoff.

Second, we introduce a trust game which has an *uncertain* payoff from the outside option instead of a *certain* payoff. This is a relevant feature in many real-world investment transactions. In such situations, our data suggest that information about the foregone payoff is an important indicator of when promises will be kept. If only the promisor knows what the promisee has foregone to invest in him, the realized foregone payoff does not affect promise-keeping behavior. However, when both the promisor and the promisee know the realized foregone option, then a bad foregone option leads to lower promise-keeping compared to a good foregone option. This suggests that investors should strategically avoid information about what she has foregone if she has already invested. If the investor could choose to find out what the actual realized foregone payoff was, she is better off not revealing the information to herself if the choice of not revealing the information itself does not adversely affect promise-keeping behavior; or if she does reveal the information, she is better off if the promisor is unaware that she knows the foregone payoff. Furthermore, it also suggests that the promisor may want to strategically communicate the foregone payoff to the investor when the foregone payoff is low.

Our paper also relates to the literature on contract law. In contract law, a non-binding promise can be viewed as a binding promise enforceable by law if such a promise induced the promisee to take an action that the promisee would have otherwise not taken (Restatement (second) of contracts § 90(1) (1979)). Under promissory estoppel, generally two forms of penalties are imposed on the party which is guilty of breaking the promise.[18] The promisor could be asked to pay "expectation damage", which requires the promisee to be put in position she would have been in had the promise been kept. Or the promisor could be asked to pay "reliance damage", which requires the promisee to be restored in the position she would have been in had the promise never been made. Thus, in law, the amount of blame one faces and the penalty one has to pay is not just dependent on what the position of the promisee is when the promise is broken and what could have been had the promise been kept, but also what the position had been if the promisee did not rely on the promise in the first place. This in turn suggests that deterrence to break a promise when there is legal recourse is lower when the foregone payoff is low compared to when it is high. Our results show that the promise-keeping behavior in absence of legal punishment also follows the pattern which the promisor would have followed had there been a legal recourse like reliance damage. The promisor internalizes the regret and hence the blame that the promisee would attribute to him even in absence of legal enforcement.

References

Damages for breach of contract. URL https://www.law.nyu.edu/sites/default/files/ECM$_PRO_0 63763.pdf$.

[18] See https://www.law.nyu.edu/sites/default/files/ECM_PRO_063763.pdf for an exposition on the damages promisors pay when there is a breach in contract.

Pierpaolo Battigalli and Martin Dufwenberg. Dynamic psychological games. *Journal of Economic Theory*, 144(1):1–35, 2009.

David E Bell. Regret in decision making under uncertainty. *Operations research*, 30(5):961–981, 1982.

Roland Bénabou and Jean Tirole. Incentives and prosocial behavior. *American economic review*, 96(5):1652–1678, 2006.

Joyce Berg, John Dickhaut, and Kevin McCabe. Trust, reciprocity, and social history. *Games and economic behavior*, 10(1):122–142, 1995.

Puja Bhattacharya and Arjun Sengupta. Promises and guilt. *Available at SSRN 2904957*, 2016.

Gary E Bolton and Axel Ockenfels. Erc: A theory of equity, reciprocity, and competition. *American Economic Review*, pages 166–193, 2000.

Gary Charness and Martin Dufwenberg. Promises and partnership. *Econometrica*, 74(6):1579–1601, 2006.

Giovanni Di Bartolomeo, Martin Dufwenberg, Stefano Papa, and Francesco Passarelli. Promises, expectations & causation. *Games and Economic Behavior*, 113:137–146, 2019.

Florian Ederer and Alexander Stremitzer. Promises and expectations. *Games and Economic Behavior*, 106:161–178, 2017.

Tore Ellingsen and Magnus Johannesson. Promises, threats and fairness. *The Economic Journal*, 114(495):397–420, 2004.

Armin Falk and Urs Fischbacher. A theory of reciprocity. *Games and economic behavior*, 54(2):293–315, 2006.

Ernst Fehr and Klaus M Schmidt. A theory of fairness, competition, and cooperation. *Quarterly journal of Economics*, pages 817–868, 1999.

Urs Fischbacher. z-tree: Zurich toolbox for ready-made economic experiments. *Experimental economics*, 10(2):171–178, 2007.

Kevin Grubiak et al. Exploring image motivation in promise keeping â-"an experimental investigation. Technical report, School of Economics, University of East Anglia, Norwich, UK., 2019.

David K Levine. Modeling altruism and spitefulness in experiments. *Review of Economic Dynamics*, 1(3):593–622, 1998.

Graham Loomes and Robert Sugden. Regret theory: An alternative theory of rational choice under uncertainty. *The economic journal*, 92(368):805–824, 1982.

Lee Ross, David Greene, and Pamela House. The "false consensus effect": An egocentric bias in social perception and attribution processes. *Journal of experimental social psychology*, 13(3):279–301, 1977.

Miriam Schütte and Carmen Thoma. Promises and image concerns. Technical report, SFB/TR 15 Discussion Paper, 2014.

Rebecca Stone and Alexander Stremitzer. Promises, reliance, and psychological lock-in. 2017.

Christoph Vanberg. Why do people keep their promises? an experimental test of two explanations. *Econometrica*, 76(6):1467–1480, 2008.

Christoph Vanberg. A short note on the rationality of the false consensus effect. Technical report, Discussion Paper Series, 2019.

Marcel Zeelenberg, Jane Beattie, Joop Van der Pligt, and Nanne K De Vries. Consequences of regret aversion: Effects of expected feedback on risky decision making. *Organizational behavior and human decision processes*, 65(2):148–158, 1996.

Appendix A. Instructions: Communication and Feedback (Lab Language: German)

On screen instructions

Page 0: (Displayed when subjects enter the lab)

Thank you for participating in this experiment. Please read the following instructions carefully. If you have a question, please raise your hand quietly.

General rules.

- This experiment will take approximately 60 minutes. During this time, you should not leave your station.
- Please turn off and stow away your phone. Starting now, there should be nothing on your table. (A drink is okay.)
- Please remain quiet for the duration of the experiment, and do not speak to other participants.
- At the end of the experiment, please remain at your station until your number is called. You will then receive your payment and sign a receipt.
- You will receive further instructions once all participants have been seated.

Page 1: Rounds, Roles, and Groups. Today's experiment consists of **9 Rounds** which are conducted **independently of one another**. This means that your decisions in any given round have no influence on what will happen in other rounds. At the end of the experiment, **one round will be randomly chosen for payment**. Your payment will depend only on the decisions made in this randomly chosen round.

At the beginning of the experiment, a **"role"** will be assigned to each participant. Half of the participants will receive **"Role A"**, the other half **"Role B"**. Your role will remain the same **for all 9 rounds** of the experiment.

At the beginning of each round, groups will be **randomly** formed, each consisting of one participant A and one participant B. (You will never be matched with the same participant twice in a row. It is possible that you will be matched to the same participant in a later round. However this is not predictable and the participants remain *anonymous* in all rounds.)

Page 2: Process within a round. **In general:** The process within a round is basically the same in all rounds. First, participant A chooses one of the following options:

1. **Option "Coin flip":** If this option is chosen, participant B will subsequently toss a (virtual) coin to determine the payments to the participants. Participant A will receive **7 EUR** in case of "Heads" and **3 EUR** in case of "Tails". (Heads and Tails are equally likely.) Participant B will receive **5 EUR** in both cases.
2. **Option "Participant B decides":** If this option is chosen, participant B will subsequently decide how **17 EUR** will be split between the participants. He will chose between **two available divisions** (for example **9 for A and 8 for B** or **4 for A and 13 for B**). Which divisions will be available to chose from will vary from round to round. Both participants will be informed about the available choices at the *beginning* of each round.

Details: In order to better understand the participants' decisions, we will ask participant B to determine an outcome for **both of the options** that participant A may chose. While participant A makes his choice, participant B will first toss a (virtual) coin and then chose between the two available divisions of 17 EUR. *Whether the coin toss or the chosen division counts depends on participant A's choice.* (At this point, participant B will not know what choice participant A made.)

Page 3: Communication at the beginning of a round. Before the participants make their decisions, participant B will have the opportunity to send a **message (maximum 200 characters)** to participant A. Participant B may not reveal his identity or identifying characteristics (e.g. "I am the person with the blue

T-Shirt" or similar). Other than this, it is up to you to decide whether and how you use this opportunity.

Page 4: Bonus questions at the end of a round (Participant A). After the participants have made their decisions, they will have the opportunity to receive an additional payment by answering a question. Participant A will attempt to guess what decision participant B has made. And participant B will attempt to guess, how participant A answers that question. The concrete procedure works as follows:

- **Participant A** will report, on a scale of 1-5, how likely he feels it is that participant B chose the first or the second division. The bonus payment that participant A will receive for this task depends on his report and on participant B's actual choice. This is summarized in the table below (all numbers represent EUR.)

.

	(1) certainly division 1	(2) probably division 1	(3) unsure	(4) probably division 2	(5) certainly division 2
B's choice is division 1	0.65 EUR	0.60 EUR	0.50 EUR	0.35 EUR	0.15 EUR
B's choice is division 2	0.15 EUR	0.35 EUR	0.50 EUR	0.60 EUR	0.65 EUR

(For example, if participant A feels certain that participant B chose division 2, he should report "Certainly division 2", because if paricipant B actually chose division 2, he will then receive the largest payment. However, he will then receive the smallest payment if participant B chose division 1 after all. Therefore, if participant A feels uncertain, he should consider the other reports as well.)

Page 5: Bonus questions at the end of a round (Participant B).

- **Participant B** will attempt, at the same time, to guess what answer participant A will give to this question. That is, participant B will also chose a column in the table. If he guesses correctly, he will receive 1 EUR

.

	(1) certainly division 1	(2) probably division 1	(3) unsure	(4) probably division 2	(5) certainly division 2
B's choice is division 1	0.65 EUR	0.60 EUR	0.50 EUR	0.35 EUR	0.15 EUR
B's choice is division 2	0.15 EUR	0.35 EUR	0.50 EUR	0.60 EUR	0.65 EUR

(For example, if participant B chooses "(2) probably division 1", and if participant A actually chose "(2) probably division 1", then Participant B will receive 1 EUR. If participant A chose a different column, participant B will receive 0 EUR.)

At the end of the experiment, **one round** will be randomly chosen for payment of the bonus question. This round will be **different** from the round that is chosen for payment of the decisions.

Page 5: Summary. To summarize, the following steps are performed sequentially in each round:

1. Both participants learn what divisions will be available to participant B if participant A chooses the option "Participant B decides"
2. Participant B can send a message to participant A. If he does so, the message is immediately displayed to participant A.
3. Participant A chooses one of the options "Coin flip" or "Participant B decides". (B is not immediately informed of A's decision.) At the same time, participant B throws a (virtual) coin and chooses a division. (Whether the coin or his choice counts depends on A's decision.)
4. Both participants answer the bonus question.

After all groups have completed these steps, a new round will immediately begin. That is, you will not immediately be informed about the decisions made in your group, the result of the coin flip, or what payments were realized. You will receive this information only after all 9 rounds of the experiment have been completed. (See next page.)

Page 7.

Feedback at the end of the experiment. At the end of the experiment, you will receive a detailed summary of your decisions and the results in all 9 rounds.

- If participant A chose "coin toss", he will learn only the result of the coin toss, and not which division participant B would have chosen.
- If participant A chose "B choses", he will learn both the division that B chose and the result of the coin toss.

In addition, you will be informed about which round was chosen for payment of the decision and which round was chosen for payment of the bonus question.

Payment. Your payment (including 3 EUR show up fee) will be displayed on the final screen. Please enter this amount on your receipt and sign it. Then please wait quietly at your station until your number is called.

Appendix B. Beliefs

Table B.6: Beliefs

		Low	High	Difference
Feedback	No Communication	2.38	2.11	-0.27 ($p = 0.02$)
	Communication	3.58	3.84	0.26 ($p = 0.14$)
	Promise	4.32	4.24	-0.08 ($p = 0.63$)
No Feedback	No Communication	2.19	1.99	-0.20 ($p = 0.19$)
	Communication	3.63	3.52	-0.11 ($p = 0.37$)
	Promise	3.97	4.22	0.25 ($p = 0.12$)

p- values reported are from Sign-Rank test.

We designed the experiment to rule out expectation based guilt aversion as an explanation for the variation in promise-keeping. In our experiment, Ann indicated on a scale of 1 to 5 how likely she thinks Bob is going to choose *Fair*, 1 indicates that Ann is sure that Bob will not choose the fair option and 5 indicating that she is sure that Bob will choose the fair option. Bob guessed Ann's choice in the belief elicitation task, which serves as our measure of Bob's second-order belief. Since Bob knows that Ann is unaware of the outside option, we expect that his stated belief to be independent of the realized outside option.

Table B.6 reports the average beliefs reported at the matching group level. In the Feedback condition when Bob could not communicate, the average reported beliefs are 2.38 and 2.11 for the low and the high outcome respectively. The difference is statistically significant (Sign-Rank, $p = 0.02$). When Bob could communicate, the average second-order beliefs reported are higher than when he could not communicate, 3.58 and 3.84 for the low and the high outcome respectively, but the difference in belief between the high and low foregone payoff is not significant (Sign-Rank, $p = 0.14$). This observation holds true if we restrict our attention to only those who promised.

In the No Feedback condition, when Bob could not communicate, the average reported beliefs are 2.19 and 1.99 for the low and the high foregone payoff respectively. The difference is not statistically significant (Sign-Rank, $p = 0.19$).

When Bob could communicate, the average second-order beliefs reported are higher than the No Communication condition, 3.62 and 3.51 for the low and the high outcome respectively, but the difference between the two is not significant (Sign-Rank, $p = 0.37$).

Table B.7: Linear Probability Model: Controlling for Belief

	I (Feedback)	II (No Feedback)	III (Combined)
High	-0.058*	-0.001	-0.003
	(0.031)	(0.046)	(0.045)
Communication	-0.031	0.055	0.099
	(0.171)	(0.192)	(0.152)
High*Communication	0.183**	-0.066	-0.067
	(0.065)	(0.076)	(0.073)
Belief	0.136***	0.161***	0.149***
	(0.029)	(0.021)	(0.018)
Belief*Communication	-0.013	0.003	-0.004
	(0.052)	(0.042)	(0.034)
Feedback			-0.042
			(0.065)
Communication*Feedback			-0.177
			(0.119)
High*Feedback			-0.052
			(0.054)
High*Communication*Feedback			0.242**
			(0.098)
Constant	-0.050	-0.063	-0.037
	(0.051)	(0.066)	(0.062)
Number of observations	486	486	972
R^2	0.62	0.60	0.60

Standard errors in parentheses. All specifications are clustered at the matching group level. $^{*}p < .10$, $^{**}p < .05$, $^{***}p < .01$.

We find that the beliefs do not significantly differ between the high and the low foregone option, except for the No Communication Feedback treatment. In Table B.7 we control for second-order belief and its interaction term with

Communication in our regression analysis. As can be seen from the table, beliefs do not explain the differential choice of *Fair* between the high and the low foregone payoff. The coefficient of $HighXCommunication$ in specification I and the coefficient of $HighXCommunicationXFeedback$ in specification III are still positive and significant. This suggests that variation in promise-keeping due to variation in the foregone payoff is not explained by variation in beliefs.[19]

Appendix C. Ann's Choice

Table C.8: Ann's Choice

		Low	High	Difference
Feedback	No Communication	39.48%	46.41%	6.93% ($p = 0.21$)
	Communication	61.78%	68.96%	7.18% ($p = 0.21$)
	Promise	69.22%	74.02%	4.80% ($p = 0.55$)
No Feedback	No Communication	39.03%	44.56%	5.53% ($p = 0.21$)
	Communication	55.53%	60.83%	5.30% ($p = 0.59$)
	Promise	70.96%	80.50%	9.52% ($p = 0.20$)

p- values reported are from Sign-Rank test.

Table C.8 and C.9 shows average investment rates and average belief reported by Ann conditional on the foregone payoff. As Ann made her decisions without knowing the exact foregone payoff, we did not expect any systematic variation in choice of investment and beliefs between the high and the low foregone payoff. This is observed in our data. In the No Communication conditions, the investment rates were 39.48% and 46.41% with feedback and 39.03% and 44.56% without feedback under the low and the high outcome respectively. The difference in investment rate between the high and the low foregone payoff is not significant (Feedback, $p = 0.21$; No Feedback, $p = 0.21$). In the Communication

[19]Beliefs have a strong positive impact on the probability of *Fair* choice. However, this does not imply that change in belief causes change in the choice of Fair. Beliefs could follow choice instead of choice following beliefs(Ross et al., 1977; Vanberg, 2019).

conditions, the investment rates were higher compared to the No Communication conditions. The investment rates were 61.78% and 68.96% with feedback and 55.53% and 60.83% without feedback under the low and the high foregone payoff respectively. However, there is no significant difference in the investment rate between the high and the low foregone payoff (Feedback, $p = 0.21$; No Feedback, $p = 0.59$).

Table C.9: Ann's Belief

		Low	High	Difference
Feedback	No Communication	2.34	2.40	0.06 ($p = 0.81$)
	Communication	3.41	3.29	-0.11 ($p = 0.77$)
	Promise	3.80	3.93	0.13 ($p = 0.67$)
No Feedback	No Communication	2.43	2.41	-0.02 ($p = 1.00$)
	Communication	3.21	3.21	0.00 ($p = 0.86$)
	Promise	3.72	3.64	-0.08 ($p = 0.44$)

p- values reported are from Rank Sum test.

As Table C.9 shows, the average beliefs stated by Ann are also independent of the foregone payoff. In No Communication conditions, the average belief stated is 2.34 and 2.40 under Feedback and 2.43 and 2.41 under No Feedback for the low and the high outcome respectively. In both Feedback and No Feedback condition without communication, the beliefs stated does not depend on the foregone payoff (Feedback, $p = 0.81$; No Feedback, $p = 1.00$). When Bob could communicate, Ann stated higher beliefs. In the Communication condition, the average belief stated is 3.41 and 3.29 with Feedback and 3.21 and 3.21 without Feedback under both the low and the high outcome. The stated beliefs do not significantly vary with the foregone payoff (Feedback, $p = 0.21$; No Feedback, $p = 0.59$).

Table C.10 and C.11 shows the average investment rates and beliefs by treatments. Communication has a significant positive effect on investment and beliefs for both Feedback and No Feedback conditions. With Communication, the average investment rate significantly increased by 18.11% point in the No Feedback

Table C.10: Ann's Choice by Treatments

	No Communication	Communication	Promise	Column Difference
No Feedback	41.15%	59.26%	72.12%	18.11 ($p = 0.03$)
Feedback	43.21%	62.96%	71.86%	19.75 ($p = 0.06$)
Row Difference	2.06 ($p = 0.89$)	3.70 ($p = 0.63$)		

p- values reported are from Rank Sum test.

condition (Rank Sum, $p = 0.03$) and by 19.75% with Feedback (Rank Sum, $p = 0.03$). Similarly, the belief increased from 2.41 to 3.23 in the No Feedback condition (Rank Sum, $p = 0.008$) and from 2.37 to 3.38 in the Feedback condition (Rank Sum, $p = 0.008$).

Table C.11: Ann's Belief by Treatments

	No Communication	Communication	Promise	Column Difference
No Feedback	2.41	3.23	3.48	0.82 ($p = 0.008$)
Feedback	2.37	3.38	3.71	1.00 ($p = 0.001$)
Row Difference	.04 ($p = 0.96$)	0.15 ($p = 0.82$)		

p- values reported are from Rank Sum test.

Feedback has no impact on both beliefs, and choices Ann make. In the No Communication condition, the average investment rate is 41.15% under No Feedback and 43.21% under Feedback. With Communication, the average investment rate is 59.26% under No Feedback and 62.96% under Feedback. The effect of Feedback on choice of investment is insignificant (No Communication, $p = 0.89$; Communication, $p = 0.63$). For beliefs, in the No Communication condition, the average belief is 2.41 under No Feedback and 2.37 under Feedback. With Communication, the average belief is 3.23 under No Feedback and 3.38 under Feedback. The effect of Feedback on belief is also insignificant (No Communication, $p = 0.96$; Communication, $p = 0.82$).

Appendix D. Regression Results

Table D.12: Linear Probability Model: Promise

	I (Feedback)	II (No Feedback)	III (Combined)
high	-0.087** (0.035)	-0.032 (0.058)	-0.032 (0.057)
Promise	0.129 (0.105)	0.448*** (0.083)	0.448*** (0.082)
high*Promise	0.243*** (0.063)	-0.040 (0.096)	-0.040 (0.094)
Feedback			-0.016 (0.076)
Promise*Feedback			-0.320** (0.132)
high*Feedback			-0.055 (0.067)
high*Promise*Feedback			0.282*** (0.113)
Constant	0.271*** (0.060)	0.287*** (0.048)	0.287*** (0.047)
Number of observations	425	394	819
R^2	0.08	0.18	0.14

Standard errors in parentheses. All specifications are clustered at the matching group level. $^{*}p < .10$, $^{**}p < .05$, $^{***}p < .01$.

Table D.13: Linear Probability Model: Promise and Controlling for Belief

	I (Feedback)	II (No Feedback)	III (Combined)
high	-0.058*	-0.001	-0.003
	(0.031)	(0.046)	(0.045)
Promise	-0.139	-0.057	0.034
	(0.196)	(0.235)	(0.179)
high*Promise	0.170**	-0.027	-0.031
	(0.059)	(0.084)	(0.084)
Belief	0.136***	0.161***	0.149***
	(0.029)	(0.021)	(0.018)
Belief*Promise	0.014	0.038	0.023
	(0.058)	(0.050)	(0.039)
Feedback			-0.042
			(0.065)
Promise*Feedback			-0.225*
			(0.117)
high*Feedback			-0.052
			(0.054)
high*Promise*Feedback			0.193*
			(0.101)
Constant	-0.050	-0.063	-0.037
	(0.051)	(0.066)	(0.062)
Number of observations	420	394	814
R^2	0.62	0.60	0.60

Standard errors in parentheses. All specifications are clustered at the matching group level. $^{*}p < .10$, $^{**}p < .05$, $^{***}p < .01$.

Table D.14: Mixed Effect Logit Model

	I (Feedback)	II (No Feedback)	III (Combined)
high	-0.557*	-0.129	-0.126
	(0.337)	(0.302)	(0.304)
Communication	0.500	1.392***	1.420***
	(0.601)	(0.440)	(0.512)
high*Communication	1.385***	-0.246	-0.253
	(0.461)	(0.412)	(0.416)
Feedback			-0.174
			(0.517)
Communication*Feedback			-0.952
			(0.722)
high*Feedback			-0.422
			(0.451)
high*Communication*Feedback			1.606***
			(0.617)
Constant	1.204***	0.481**	0.778***
	(0.568)	(0.235)	(0.262)
Number of observations	486	486	972

Standard errors in parentheses. All specifications are clustered at the matching group level. $^{*}p < .10$, $^{**}p < .05$, $^{***}p < .01$.

Table D.15: Probit Model

	I (Feedback)	II (No Feedback)	III (Combined)
high	-0.291**	-0.098	-0.098
	(0.120)	(0.176)	(0.176)
Communication	0.238	0.797***	0.797***
	(0.269)	(0.204)	(0.201)
high*Communication	0.703***	-0.145	-0.145
	(0.194)	(0.230)	(0.227)
Feedback			-0.046
			(0.225)
Communication*Feedback			-0.559*
			(0.333)
high*Feedback			-0.193
			(0.210)
high*Communication*Feedback			0.849***
			(0.297)
Constant	-0.609***	-0.563***	0.778***
	(0.181)	(0.137)	(0.262)
Number of observations	486	486	972

Standard errors in parentheses. All specifications are clustered at the matching group level. $^{*}p < .10$, $^{**}p < .05$, $^{***}p < .01$.

Table D.16: Mixed Effect Logit Model: Promise

	I (Feedback)	II (No Feedback)	III (Combined)
high	-0.568*	-0.126	-0.122
	(0.340)	(0.304)	(0.306)
Promise	1.038	2.290***	2.388***
	(0.795)	(0.577)	(0.691)
high*Promise	1.419***	-0.193	-0.198
	(0.518)	(0.510)	(0.522)
Feedback			-0.231
			(0.636)
Promise*Feedback			-1.453
			(0.945)
high*Feedback			-0.437
			(0.455)
high*Promise*Feedback			1.575**
			(0.728)
Constant	2.231**	0.766**	1.365***
	(1.074)	(0.384)	(0.471)
Number of observations	425	394	819

Standard errors in parentheses. All specifications are clustered at the matching group level. $^{*}p < .10$, $^{**}p < .05$, $^{***}p < .01$.

Table D.17: Probit Model: Promise

	I (Feedback)	II (No Feedback)	III (Combined)
high	-0.291**	-0.098	-0.098
	(0.120)	(0.176)	(0.170)
Promise	0.355	1.192***	1.192***
	(0.287)	(0.248)	(0.245)
high*Promise	0.684***	-0.111	-0.111
	(0.177)	(0.264)	(0.261)
Feedback			-0.046
			(0.225)
Promise*Feedback			-0.836**
			(0.374)
high*Feedback			-0.193
			(0.210)
high*Promise*Feedback			0.795**
			(0.314)
Constant	-0.609***	-0.563***	-0.563***
	(0.181)	(0.137)	(0.137)
Number of observations	486	486	972

Standard errors in parentheses. All specifications are clustered at the matching group level. $^{*}p < .10$, $^{**}p < .05$, $^{***}p < .01$.

Appendix E. Variation of Unfair Payoff

Table E.18: Feedback: Variation of Unfair Payoff

		low	high	Difference = high - low
Unfair	Communication	42.53%	60.55%	18.01 ($p = 0.02$)
(4,13)	No Communication	25%	19.25%	-5.74 ($p = 0.02$)
Unfair	Communication	34.37%	41.25%	6.87 ($p = 0.02$)
(3,14)	No Communication	16.50%	14.81%	-1.69 ($p = 0.02$)
Unfair	Communication	26.81%	48.95%	22.54 ($p = 0.02$)
(2,15)	No Communication	34.20%	24.81%	-9.49 ($p = 0.02$)

Table E.19: No Feedback: Variation of Unfair Payoff

		low	high	Difference = high - low
Unfair	Communication	54.33%	53.33%	-1.0 ($p = 0.905$)
(4,13)	No Communication	25.55%	25.18%	-0.3 ($p = 0.81$)
Unfair	Communication	66.34%%	53.80%	-12.53 ($p = 0.44$)
(3,14)	No Communication	32.26%	17.51%	-14.74 ($p = 0.12$)
Unfair	Communication	50.18%	43.09%	-7.08 ($p = 0.44$)
(2,15)	No Communication	27.40%	32.77%	0.05($p = 0.67$)

Appendix F. Experiment Screen Shots

Figure F.5: B's Message Stage

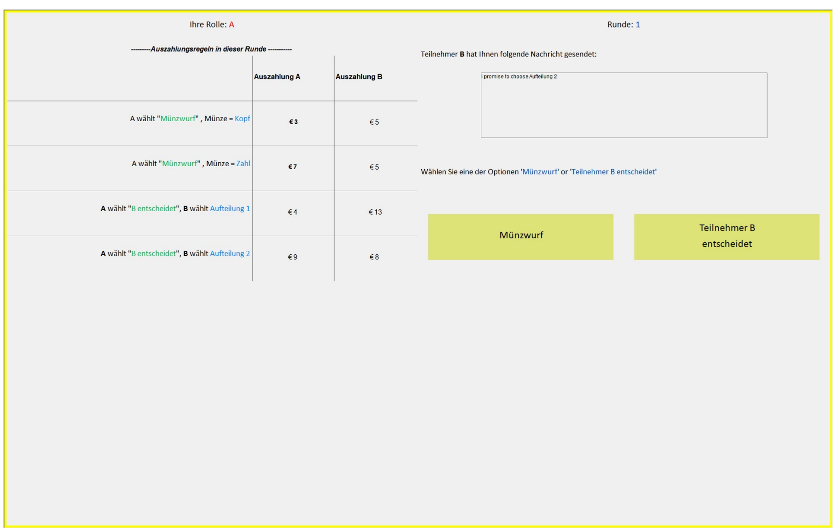


Figure F.6: A's Message and Choice Stage

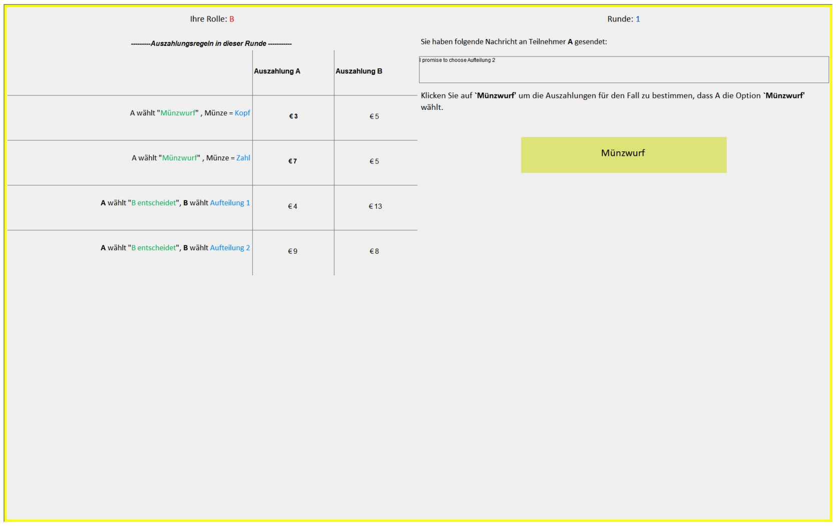


Figure F.7: B's Coin Flip Stage

Figure F.8: B's Choice Stage

Figure F.9: A's Belief Stage

Figure F.10: B's Belief Stage