Exp Econ (2019) 22:530–551
https://doi.org/10.1007/s10683-018-9567-2

ORIGINAL PAPER

Why do people keep their promises? A further investigation

Steven Schwartz[1] · **Eric Spires**[2] · **Rick Young**[2]

Received: 29 October 2015 / Revised: 12 March 2018 / Accepted: 19 March 2018 /
Published online: 31 March 2018

Abstract Two rationales have emerged for why individuals keep their promises: (a) an emotional commitment to keep actions and words consistent, a *commitment rationale* and (b) avoidance of guilt due to not meeting the expectations of the promisee, an *expectations rationale*. We propose a new dichotomy with clearer distinctions between rationales: (1) an *internal consistency* rationale, which is the desire to keep actions and words consistent regardless of others' awareness of the promise and (2) a *communication* rationale, which captures all aspects of promise keeping that are associated with the promisee having learned of the promise, including but not limited to promisee expectations. Using an experiment that manipulates whether promises are delivered, we find no support for the internal consistency rationale; only delivered promises are relevant. In a second experiment designed to better understand what aspect of promise delivery influences promisor behavior, we manipulate whether the promise is delivered before or after the promisee is able to take a trusting action. We find late-arriving promises are relevant though not as relevant as promises delivered before the promisee chooses whether to take the trusting action. We conclude that implicit contracting does not fully explain promise keeping, because had it done so, late-arriving promises would also be irrelevant.

Keywords Promises · Commitment · Expectations · Trust game

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10683-018-9567-2) contains supplementary material, which is available to authorized users.

✉ Steven Schwartz
sschwart@binghamton.edu

[1] School of Management, Binghamton University, Binghamton, NY 13902, USA

[2] Fisher College of Business, Ohio State University, Columbus, OH 43210, USA

JEL Classification C91 · C70

1 Introduction

Considerable evidence suggests that individuals keep promises even when it is contrary to their financial self-interest. The natural question that arises is *why* individuals keep such promises. Two suggestions have emerged in the literature. The first is that, in the event a promise affects another individual's expectations, promisors do not wish to disappoint the promisee, which we refer to as the *expectations rationale*. The second is that individuals have a preference for promise keeping per se, referred to as the *commitment rationale*.

The expectations rationale for promise keeping was first suggested by Charness and Dufwenberg (2006) (CD), who build on an earlier study of Dufwenberg and Gneezy (2000) that examines expectations but does not include promises. CD posit that promisors wish to avoid the guilt associated with breaking their promises, or more literally, promisors wish to avoid violating what they believed to be the expectations of the promisee. Using a trust game and measurements of promisors' beliefs regarding promisees' beliefs (second-order beliefs), CD find evidence consistent with the expectations rationale. However, there are other explanations consistent with CD's results, in particular, false consensus bias, which is the tendency for individuals to think others are more like them than they actually are.

Several studies follow up on CD. Kawagoe and Narita (2014) elicit the trustor's first-order beliefs and communicate them to the trustees but find little support for the expectations rationale.[1] In contrast, Ederer and Stremitzer (2016) manipulate second-order beliefs by administering games that differ in the likelihood that promisors can keep their promise. They find that first- and second-order beliefs and promisors' willingness to keep their promise (conditional on the ability to do so) are all higher if the probability of being able to keep one's promise is higher. They interpret this as evidence of the relevance of others' expectations in promise keeping.

Vanberg (2008) also conducts an experiment similar to CD, but designed to eliminate false consensus bias, and concludes that second-order beliefs are not sufficient to explain promise keeping.[2] Vanberg introduces an alternative explanation for promise keeping stating that "people have a preference for promise keeping per se" (p. 1467), which he refers to as a *commitment rationale* for promise keeping. In his experiment half of the promisors were switched to different promisees. It was found that promises made by former partners do not affect current partners'

[1] Kawagoe and Narita (2014) do not inform trustors that their first-order beliefs will be communicated to trustees.

[2] Ellingsen et al. (2010) eliminate the potential for false consensus bias by directly informing participants of others' beliefs. They find no evidence that the expectations of others matter. However, the design used by Ellingsen et al. (2010) does not include promises. Further, Khalmetski et al. (2015), using within-subject data, call into question the conclusions of Ellingsen et al. (2010), finding that individuals do condition their responses on the expectations of others, but that variation between individuals masks important behavior if analyzed only on an aggregate level. In neither study are trustors informed their beliefs will be communicated to trustees.

behavior. Also a promisor who made a promise to a former partner does not appear to adhere to the promise when dealing with a new partner. However, questions on the role of expectations remain. As Bicchieri (2006, p. 25) notes when discussing social norms, "Guilt… presupposes the violation of expectations we consider *legitimate*" [emphasis in the original]. We suggest there is no particular reason why an individual would necessarily consider as legitimate the expectations of another if those expectations were not conditioned on that same individual's promise.

We propose a new, clearer dichotomy that consists of an *internal consistency rationale* and a *communication rationale*. The former, first proposed by Ismayilov and Potters (2016) (I&P), is a narrower version of the commitment rationale. The substance of this rationale is as follows: the mere fact that an individual explicitly states his or her intentions carries weight, even if the promisee is unaware of the promise. There is no role for expectations, as undelivered or unheard promises cannot affect first- or second-order beliefs. The latter element of the dichotomy, the *communication rationale*, is the non-overlapping alternative to the internal consistency rationale that comprises all aspects of the promisee learning of the promise, including, but not limited to the effect of the promise on the expectations of the promisee.

In a similar vein, I&P conduct an experiment designed to test the validity of the *internal consistency rationale*. Their experiment is similar to CD's except that 50% of the messages by trustees are never delivered. Trustees know the delivery status of their message when making their decision. I&P find evidence for the internal consistency explanation for promise keeping: promises were associated with more trustworthy behavior than non-promises and delivery status did not affect trustworthiness differently for promises and non-promises. They found no support for the *social obligation rationale*, which is their alternative to the internal consistency rationale.[3] However, using data from a second experiment that does not include promises, I&P conclude that people who make promises are *inherently* more trustworthy, and importantly, *that promises themselves play no role*.

We examine the role of both the internal consistency and the communication rationales for promise keeping with an experiment that differs from I&P's in several ways. In particular, our experiment employs (1) a wider range of trustworthy behavior (as opposed to a binary return trust/do not return trust choice),[4] (2) a within-participant rather than between-participant measurement of the effect of message delivery, and (3) a post-experimental questionnaire. The results of our experiment also differ from I&P. Whereas they find that the delivery of *any* sort of message is relevant to the behavior of the trustee, *regardless* of whether it contains a promise, we find that the delivery of the message is relevant *only* if it contains a

[3] We prefer the term "communication" to "social obligation" as social obligation appears to make an a priori assumption regarding the promisor's emotions.

[4] If the promisor is given a binary choice on whether or not to return the trust, the communication rationale would be evidenced by finding promisors who would not break their promise to return the trust if their promise were delivered, but would not otherwise act in a trustworthy manner. It is not unreasonable to conjecture those who keep their promises would also be those who would act trustworthy even without a promise. In our design, a promisor can act *less* kindly if the message is not delivered than if it is delivered, but still not take maximum advantage of the other individual.

promise. Our results show the amount returned by the trustee is 60% higher when the promise is delivered than when it is not delivered. Further, messages that do not contain a promise (delivered or not delivered), as well as undelivered messages that do contain a promise, are characterized by similar returned amounts. The rates are roughly equivalent to those found in similar experiments without promises. Thus, our results support the communication rationale. We find no support for the internal consistency rationale.

After analyzing the results of the experiment and observing that simply writing an undelivered promise led to approximately the same amount of trustworthiness as similar experiments without promises, we designed a second experiment to look more closely at the factors that enhance promise keeping. Our second experiment is identical to the first, except the trustee's message is either delivered before or after the trustor makes her decision. The experiment is intended to help identify which aspects of the promisee learning of the promise create legitimate expectations that the promise will be kept. Specifically, is the legitimizing aspect of the promise that it induces an action by the promisee? If so, late delivery would not be enough to legitimize the promise. Or is the legitimizing aspect simply that the promisee becomes aware of the promise, in which case the timing of the delivery is irrelevant? The question is particularly important, given the suggested relevance of promise keeping to contracts.

The results of the second experiment reveal that promisors condition their response on late delivery versus on-time delivery, but significantly less so than on never delivered versus on-time delivery. Specifically, the difference in mean levels of trustworthiness between the two delivery protocols for promises is only about 40% as large in the second experiment as it is in the first experiment. Put differently, timely promises matter more than late promises, but late promises matter more than never delivered promises. The implication is that while implicit contracting may be part of the reason individuals keep their promises it is not the whole story. Had it been so, late arriving promises would be treated the same as non-delivered promises (as neither can be part of an implicit contract), and only on-time delivered promises would be relevant.

2 Hypotheses

The setting we use to test our hypotheses is a trust game similar to Berg et al. (1995). The game consists of two people, a trustor and a trustee. Aside from show-up fees, the trustor is endowed with an amount I, where $I > 0$, which may be either retained or invested with the trustee. If retained, the trustor earns I and the trustee zero. If invested with the trustee, the investment becomes $Y > I$. The trustee may return an amount R, $R \leq Y$, to the trustor, keeping $Y - R$ for himself. Because in the final move the trustee's best response is to choose $R = 0$, in the unique subgame perfect equilibrium assuming only selfish motives, the trustor chooses not to invest.

Before the trustor makes her investment decision, the trustee may send a short written message to the trustor. The message may have any content (including being empty) other than offensive or threatening language. In the event the message

contains a promised amount to be returned, we denote it by P. We use free-form messages, rather than simply having trustees state how much they intend to return, because free-form messages are more likely to be followed than parsimonious messages (Charness and Dufwenberg 2010; Lundquist et al. 2009), thereby increasing the power of our tests. It is common information that 50% of the messages would be delivered and 50% would not be delivered to the trustors.

There is substantial evidence that people keep their promises even at personal cost (Klein and O'Flaherty 1993; Ellingsen and Johannesson 2004a, b; Charness and Dufwenberg 2006; Douthit et al. 2012). If trustees make promises, we would expect them to frequently promise to return something to trustors, and if they are inclined to keep their promises they will tend to behave in a trustworthy manner. Therefore, *conditional* on trustees' messages having been received by trustors, we expect R to be increasing in the promised return P:

Hypothesis 1 Conditional on the message being delivered, the amount returned by trustees is increasing in the amount they promised to return.

If trustors anticipate the hypothesized behavior in Hypothesis 1, their expectations about how much will be returned would be increasing in P. Therefore, we expect the frequency of investment by trustors would be increasing in P:

Hypothesis 2 Conditional on the message being delivered, the frequency of investment is increasing in the amount promised to be returned.

We now arrive at our primary research question, which is *why* people keep their promises. We focus on two non-overlapping rationales. The internal consistency rationale follows from a desire to keep words and deeds consistent. The awareness of the promise by the promisee is not relevant. The communication rationale encompasses any motivation for promise keeping that occurs due to the promisee learning of the promise.

The communication rationale implies promisors care about delivered promises; hence, delivered promises increase the amount returned, relative to non-delivered promises or delivered messages containing no promise.[5] The internal consistency rationale assumes the promise is relevant irrespective of delivery. This discussion leads to the following hypotheses:

Hypothesis 3A (*Communication rationale*) Delivered promises are more likely to be kept than non-delivered promises, and non-delivered promises are associated with the same degree of trustworthiness as non-delivered messages that contain no promise.

Hypothesis 3B (*Internal consistency rationale*) Promises are associated with a higher degree of trustworthiness than messages that contain no promise, regardless of delivery status.

Of course, both rationales could explain promise keeping, which we address in the analysis of the results.

[5] It is conceivable that delivered non-promises could *decrease* trustworthiness relative to undelivered non-promises. If a blank piece of paper or an irrelevant message were interpreted by the recipient as "don't count on me," it could ease trustees' guilt for not acting in a trustworthy manner.

3 Design

The experiments were conducted at a large Midwestern university. The parameters used were as follows. The amount the trustor was endowed with was $I = \$5$. If invested, this amount became $Y = \$19$. The amount, R, the trustee could return to the trustor was required to be in whole dollars, $R \in \{0, 1, 2, 3,\ldots, 19\}$. Using an odd number for Y removes the focal 50/50 split and makes the promises even more important for forming expectations. The first experiment included four sessions with 20, 22, 18 and 22 student-participants. All participants received a show-up fee of \$10.

All aspects of the game were common information. Trustees, referred to as Player B in the administration of the experiment, had the opportunity to send a message to the trustor, referred to as Player A. It was made common information to the participants that only 50% of the messages were to be delivered, and that a randomization device would determine whether a particular message would be received by trustors. If the trustee-trustor pair was chosen for interception, the trustor received a note from the experimenter that the message was not delivered. In the case where the pair was chosen for delivery but the trustee chose not to send a message, a blank sheet was given to the trustor, so the trustor did not erroneously think a message was sent and intercepted. Otherwise, the trustor received the message written by the trustee.

In addition to writing a message if they so desired, trustees made their return choices using the strategy method. That is, they chose R conditional on the message being delivered and conditional on the message not being delivered. Clearly, if there is no investment R is irrelevant. The trustor made her investment decision after reading the message of the trustee or after learning the message was not delivered.

Upon arrival participants were handed instructions and given time to read. Instructions were then read aloud by an experimenter. Participants next took a quiz; they did not proceed until they correctly completed the quiz. Afterwards, slips of paper were shuffled and handed out, face down. These slips informed the participants of their roles: Player A or B. A-Players were then relocated to another room. All decisions were made on paper. After all decisions were made, but before participants were informed of their earnings, they were asked to complete a short questionnaire. Payments were made privately, so that no participants were aware of other participants' earnings. Sessions lasted about 75 min and average earnings were \$18, including show-up fees.

One of the authors and an unaffiliated Ph.D. student independently coded each message as a promise or no promise. Considering both experiments, there was only one disagreement (observation 32 in experiment 2). Discussion about this observation with several others yielded mixed opinions, so the coding of the disinterested party (no promise) is used in the analyses.

4 Results

The raw data from the first experiment are presented in Table 1 and summary statistics are disclosed in Table 3. One of the most striking features is the trustees' high degree of trustworthiness. As shown in Table 3, the mean return conditional on a *delivered* message, Ret|Del, is 6.88 (SD 3.74), which is almost a 38% return to trustors ((6.88 − 5)/5). A t test shows the two-tailed p value for mean return not equal to 5 is 0.003. In contrast, Berg et al. (1995), who did not allow communication by the trustees, showed a small loss to trustors who transferred a positive amount.[6] This difference possibly indicates the effect of promises on trustworthiness. The mean return conditional on an *undelivered* message, Ret|Inter, is 4.83 (SD 4.17), which does not significantly differ from 5 (p value = 0.8). A large majority of trustors invested money with the trustee (30/41). The latter two results are similar to the findings of Berg et al. (1995) (Table 2).

4.1 Hypothesis tests

Hypothesis 1 predicts that if a promise is delivered to a trustor the amount returned by the trustee is increasing in the amount promised. To test the hypothesis, we selected those messages with unambiguous promises of the amount to be returned. We then compared the *promised* returns in the message to the trustee's *decision* regarding how much would be returned *conditional on delivery*. There were twenty messages where trustees promised to return an unambiguous amount.[7] The mean such promise was to return 9.48 (SD 1.22) (see Table 3). A comparison of the twenty promises and returns conditional on delivery of message yields no significant association (correlation coefficient $r = 0.07$). Looking more closely at the data, five of the twenty were promises of at least 9, but where the trustees actually chose to return 0. Excluding those five observations, we find a significant (positive) correlation, $r = 0.74$, between promises and returns (two-tailed $p = 0.005$).[8] One might therefore conclude that, conditional on trustees being at all trustworthy, their promises are informative about returns. Interestingly, using the same 15 observations and looking at promised amounts and returns *conditional on non-delivery*, $r = 0.18$ and does not significantly differ from 0.

A more inclusive test of Hypothesis 1 would include trustees who made some sort of promise, albeit possibly ambiguous, including promises such as returning "more than was invested." Using this alternative definition of a promise, we classify 32 trustees as having made a promise and nine as having made no promise (again, based on the raw data in Table 1). However, we cannot compute correlations

[6] In our experiment $I = 5$ dollars and $Y = 19$ dollars. Although the growth factor Y/I is somewhat higher in our experiment than in Berg et al. (1995) study, 3.8 versus 3, the meta-analysis of Johnson and Mislin (2011) suggests that once the growth factor becomes this large trustees tend to keep for themselves any profit due to the increased factor.

[7] In those instances where a promise of "half" was made, we coded as 9.5. Coding as either 9 or 10 would not change the nature of the results.

[8] The correlation between the promised return and the decision regarding how much would be returned if the message were *not* delivered is 0.13, not significantly different from zero.

Table 1 Raw data from Experiment 1

Player #	Delivered	Invest	Ret|Del	Ret|Inter	Message	Prom	Prom Amt
1	N	Y	9	7	Make the transfer and we'll both make more money	Y	
2	Y	Y	10	9	Thanks for your investing. Your return will be greater than 40%	Y	
3	N	Y	9	9	Your investment will return more money. Have a good day!	Y	
4	Y	Y	0	0	Please transfer. If you do chose to transfer I have written that we will split the funds as evenly as possible with you getting 10 and me getting 9. This will benefit both of us!	Y	10
5	N	Y	9	8	Transfer so we both win more than $5	Y	
6	Y	Y	6	8	Good deal	N	
7	N	Y	0	0	Let's split the $19, everybody wins. $10 You. $9 Me	Y	10
8	Y	Y	9	5	I will give you $9 if this message is received, which is basically half of what we could earn fully. You would lose $4 if you declined. Beer money ☺	Y	9
9	N	N	9	9	Transfer!	N	
10	Y	Y	9	0	You will get at least half ($10) if you transfer the money	Y	10
11	N	Y	0	5	Wonderful day!	N	
12	Y	Y	11	9	The sky is blue	N	
13	N	Y	7	6	If you invest the $5 with me, I can promise you more than $5 will come back to you. ☺ Go Bucks!	Y	
14	Y	Y	8	6	If you transfer, we get win–win. I wrote transferring 11 to you	Y	11
15	N	Y	8	6	You will be receiving $8.00 if you decide to transfer to me	Y	8
16	Y	Y	4	4	Hi! hook a brotha up	N	
17	N	N	9	7	Do transfer. I'll be fair, I promise ☺	Y	
18	Y	N	8	0	Hey. I'll be giving you $8 if the money is transferred over. ☺ Thanks!	Y	8
19	N	Y	10	9	Dollar ten will be given	Y	10
20	Y	Y	12	13	(What you can walk out with) = 14/2 of us = 7. You get 12, I get 7. Let's split the profit you get by investing with me. You'll be rewarded fairly	Y	12
21	N	Y	10	10	Hope you're will to split the money like I do!	Y	9.5

Table 1 continued

Player #	Delivered	Invest	RetIDel	RetIInter	Message	Prom	Prom Amt
22	N	Y	0	0	You make more money, I make money, let's do this. 50–50 seems most beneficial to both of us	Y	9.5
23	Y	N	0	0	<Blank Message>	N	
24	N	Y	6	0	If you transfer, you will receive more than 5	Y	
25	Y	Y	9	9	If you transfer I will return 9 to you	Y	9
26	N	Y	6	0	Invest—we will both be better off!	Y	
27	Y	Y	0	0	IF THIS IS RETURNED, AFTER THE STUDY I WILL GIVE YOU HALF OF THE $19. THERE IS MORE OF A CHANCE OF THE MONEY BEING TAKEN BACK COMPLETELY IF I SHIP ANYTHING BACK. I CANNOT IDENTIFY MYSELF BUT YOU WILL KNOW	N	
28	N	N	0	0	Please invest the five dollars. There will be more money for both of us. I will give you 9 dollars, I will keep 10. Do the math 9 > 5. I promise I will not rob you. It simply would not make sense to rob you. I want to make US money	Y	9
29	Y	N	9	9	I'd have to be terrible to give you less than $5, so it's worth it to transfer	Y	
30	N	Y	0	0	Okay, so here's the bottom line: If you transfer the $5, we will both be walking out of here with more than the $10 show-up payment. If you transfer, I will give you $10 and keep $9 for myself. I want to keep it as fair as possible, but I am willing to give you the extra $1 as an extra incentive to transfer the $5. You can honestly trust me. If you transfer the money, you and I will both be walking out of here happy	Y	10
31	Y	Y	6	0	Choose transfer. You'll get more money than choosing not to transfer. Have a good weekend ☺	Y	
32	N	N	9	1	I will return 9	Y	9
33	Y	Y	9	0	You will get 10, and I will keep 9 100% return on investment	Y	10

Table 1 continued

Player #	Delivered	Invest	Ret|Del	Ret|Inter	Message	Prom	Prom Amt
34	N	N	9	9	Hi (Don't know what message to write.) I like cats, can you tell? [picture of two cats]	N	
35	Y	N	9	6	Hello ☺ I hope you choose to transfer the money so we can both walk away with some extra cash for the weekend. I plan to keep $9 for myself while transferring $10 back to you. Have a great weekend!	Y	10
36	N	Y	7	7	<Blank Message>	N	
37	Y	Y	10	10	I hope you invest your $5, as I would like to see that we can both make a good gain from this! ☺	Y	
38	N	N	9	10	I have tried to be as fair as I could with dividing the payment. I hope you use your money wisely. Go Bucks!	Y	
39	Y	N	10	0	If you read this I will return half to you	Y	9.5
40	N	Y	10	0	If you transfer the $, we can both walk out of here with more than $10. I will give you $10 & keep $9 for myself just to show how glad I am that you invested! INVEST IT!	Y	10
41	Y	Y	7	7	If you transfer, I will give you $6	Y	6

Player # Player B identifier

Delivered Whether Player B's message was delivered; Y = yes, N = no

Invest Whether Player A chose to invest $5 with Player B; Y = yes, N = no

Ret\|Del Amount returned by Player B conditional on message delivered

Ret\|Inter Amount returned by Player B conditional on message not being delivered

Message Player B's message, in its entirety

Prom Y = some form of promise, N = no promise

Prom Amt Amount of a specific promise. Promises of "half" are coded as 9.5. Blanks indicate messages without specific promises

because we do not have precise promises. Therefore, we compare mean returns between the promise and no-promise groups rather than correlations. Table 3 shows the mean return within this promise (no promise) group is 7.38 (5.11) ($p = 0.11$ for a t test of difference in means). In summary, there is support for our hypothesis that trustworthiness is generally increasing in the promised return.

Hypothesis 2 predicts that if a promise is delivered to the trustor, trust is increasing in the promise. Of the 41 trustors, 30 decided to invest. If we classify the trustors as either (1) receiving an unambiguous promise with an amount specified or

Table 2 Raw data from Experiment 2

Player #	OT	Invest	RetIOT	RetILate	Message	Prom	Prom Amt
1	Y	N	10	10	Equality is a good thing	Y	9.5
2	N	N	9.5	9.5	Have fun with the money! Hope that's enough money for you ☺	N	
3	Y	Y	6	0	Transfer and lets all benefit from it. I'll return you more than you don't transfer	Y	
4	N	Y	10	10	If you transfer it to me I will return you $10. It's more than $5 and more than half of $19	Y	10
5	Y	Y	10	8	Hello A, Let's make it simple and clear, if you transfer me the $5, I will send $10 back to you. (Since you have to risk of receiving nothing, so I am willing to give you $10–. More risk more return!) I think the most important thing in the business world is trust, so I will definitely keep my word, please trust me! ☺ win–win!	Y	10
6	N	Y	10	10	Hello, I intend to do this civilly to mutually benefit both of us the most. If you sent the $5.00, I will give you $10 and myself $9.00 as a show of good faith. If we are partnered for this whole thing, I implore you to also act civilly for our mutual $ interest	Y	10
7	Y	N	0	0	This is a prisoner's dilemma situation. I will send you $11 keeping $8 for myself. You get $21 total and I get $18. It's fair for both since I am relying on you to transfer the money. let's increase the economic benefit by $14 instead of the $5 *Transfer* *Don't transfer* $21-you $15-you $18-me $10-me	Y	11
8	N	Y	6	8	Your choice determines whether or not you have the opportunity to make more than $15 (your $5 + $10 for showing up). I would advise using this opportunity to your advantage	N	
9	Y	Y	9	9	Player A, I am going to split $19 evenly with you. I PROMISE you that. Please make the transfer. This way we will both be better off. I want you to trust me. No transfer you get $15. I have nothing to lose. Transfer: you get $10 + $9.5 = $19.5	Y	9.5

Table 2 continued

Player #	OT	Invest	RetIOT	RetILate	Message	Prom	Prom Amt
10	N	N	10	10	I will break the $19 so that you get $10 and I get $9 if you transfer. I don't know if there is anything I could say to make you believe me but I think we would like more money and I would feel bad if I took all of the money and lied to you	Y	10
11	Y	Y	9	9	We both get more if you invest. We're both here for the money, and I'll split it $9 and $10 with you	Y	10
12	N	N	7	7	Hello friend! The weather is very nice today. It is nice to have the beautiful sunshine and fresh air before a weekend starts. Hope that the weather is going to be as good as today for the coming week as well!	N	
13	Y	Y	10	10	How would you like to double your investment? I guarantee you $10 if you decide to invest	Y	10
14	N	Y	10	0	Please transfer the $5 to me—this will make your total $19. If you do so, I will transfer half back to you, so you will get to keep $9.50. This is a much better deal than just keeping the $5 and leaving—you'll earn $4.50 more! I promise to give you half ☺	Y	9.5
15	Y	Y	9	9	Hi! If you transfer, I will give you $9. I know it's less than half, but you wouldn't want to walk away with only $5 as well. I hope we can both cooperate! (Please understand that I can't split in half, because I don't know how to deal with decimals)	Y	9
16	N	Y	9	8	Investing is smart. Guaranteed more than $7 back. $7 > $5 so ☺	Y	
17	Y	Y	10	10	I am a very fair individual. If you choose to invest I will transfer $10 of the $19. I am a person that sticks to my word. By having you transfer/invest your $5 you and I can both have positive benefits	Y	10
18	N	Y	9	9	My plan would be to give you $9 of the $19	Y	9
19	Y	Y	10	5	I will give you $9 back so we both have about equal, you get $9 instead of only $5	Y	9
20	N	Y	9	9	If you choose to transfer, you have better chance to get more than $5. I will transfer $9 or $10 to you. We both get more!	Y	9.5

Table 2 continued

Player #	OT	Invest	Ret|OT	Ret|Late	Message	Prom	Prom Amt
21	Y	Y	9	0	Gimme da cash	N	
22	N	Y	9	9	Hopefully you choose to transfer. I would like to split it evenly and give you $9 and I keep $10. Transfer: $9	Y	9
23	Y	Y	10	10	If you transfer $5, I'll return to you $10. I'll get $9 but that's way better than just $10 (show-up). You can make $5 more than if you didn't transfer	Y	10
24	N	Y	8	6	Hi, Player A! If you choose to transfer your $5, I will send you at least $6 back! I think a fair exchange would be to send $9 back so you walk home with $19 today. Either way, if you transfer $5, you have my word (for what it is worth) that you'll get more than $5 back ☺	Y	
25	Y	Y	0	0	Invest. I'll get 10, give you 9	Y	9
26	N	Y	4	5	$6	N	
27	Y	N	0	0	<Blank Message>	N	
28	N	Y	9	9	<Blank Message>	N	
29	Y	Y	8	8	Transfer and you will get at least $8.—Go Bucks	Y	
30	N	Y	7	5	Transfer and we'll split the money	Y	9.5
31	Y	Y	10	9	If you transfer I'll give you $9 so we both make more (even more) money! Hopefully I'm allowed to say that…	Y	9
32	N	Y	0	0	If the right choice is made, we both can go home happy	N	

Player # Player B identifier

OT Whether Player B's message was delivered on-time; Y = on-time, N = late

Invest Whether Player A chose to invest $5 with Player B; Y = yes, N = no

Ret\|OT Amount returned by Player B conditional on message being delivered on-time

Ret\|Late Amount returned by Player B conditional on message being delivered late

Message Player B's message, in its entirety

Prom Y = some form of promise, N = no promise

Prom Amt Amount of a specific promise. Promises of "half" are coded as 9.5. Blanks indicate messages without specific promises

(2) receiving no message or an ambiguous message, 11 trustors fall in the former category and 30 in the latter category. Of those receiving an unambiguous promise, 8 of 11 (73%) invested, and of those not receiving an unambiguous promise, 22 of 30 (73%) invested, with no significant difference in proportions. Further, the mean amount promised to the 8 of 11 who chose to invest is 9.63, which is not significantly greater than the 9.17 mean amount promised to the 3 of 11 who chose not to invest. The data provide almost no support for Hypothesis 2.

Table 3 Summary statistics from Experiment 1

	All B-players (n = 41)		Promise (n = 32)			No promise (n = 9)	
	Ret|Del	Ret|Inter	Ret|Del	Ret|Inter	Prom Amt (n = 20)	Ret|Del	Ret|Inter
Mean	6.88	4.83	7.38	4.59	9.48	5.11	5.67
Median	9.00	6.00	9.00	6.00	9.75	6.00	7.00
SD	3.74	4.17	3.47	4.32	1.22	4.31	3.67
# ≥ 5	32	24	27	18	20	5	6

Ret|Del Amount returned by Player B conditional on message delivered

Ret|Inter Amount returned by Player B conditional on message not being delivered

Prom Amt Amount of a specific promise. Promises of "half" are coded as 9.5. For those 20 B-Players who promised a specific amount, mean (SD) Ret|Del is 6.85 (4.19) and mean (SD) Ret|Inter is 3.60 (4.32)

 # ≥ 5 The number of B-Players for whom Ret|Del, Ret|Inter, or Prom Amt equals 5 or more

Hypothesis 3 analyzes trustworthiness to assess whether the internal consistency rationale or communication rationale (or both) is explanatory. We measure trustworthiness in two ways. The first measure is binary: whether Player B returns an amount greater than or equal to 5 or less than 5. This measure may be justified because if Player B returns 5 or greater, Player A would be weakly better off for having trusted Player B by investing. The second measure is the amount returned by Player B: the greater the amount Player B returns, the greater is the level of trustworthiness.

Table 3 discloses the frequencies of B-Players who returned 5 or greater for each of the treatment combinations needed to test Hypothesis 3. For delivered promises, 27 of 32 B-Players (84%) were trustworthy; for nondelivered promises, 18 of 32 (56%). The difference is statistically significant, two-tailed $p < 0.004$ using the binomial test for correlated proportions described in Hays 1963.[9] Thus, whether the message is communicated affects trustworthiness, supporting Hypothesis 3A (communication hypothesis).

Further, for nondelivered messages that contain no promise, 6 of 9 B-Players (67%) were trustworthy. This proportion (6/9) does not significantly differ from the proportion of trustworthy B-Players non-delivered promises (18/32) ($p > 0.50$ using the Fisher exact probability test). Again, this finding is supportive Hypothesis 3A but not Hypothesis 3B (internal consistency).

Considering the other measure of trustworthiness, the amount returned by the B-Players, results again support Hypothesis 3A, but not Hypothesis 3B. Table 3 shows the mean (median) amount returned for delivered promises is 7.38 (9), with SD 3.47, while for nondelivered promises, the mean (median) returned is 4.59 (6), with SD 4.32. A paired t test shows the difference in means to be significant (two-tailed $p < 0.001$) and a Wilcoxon signed ranks test indicates the difference in medians is significant (two-tailed $p < 0.001$). The delivery of the promise has an effect on trustworthiness as measured by the amount returned (Table 4).

[9] To perform the test, the data were cast in a 2 × 2 format as presented in Table 5.

Table 4 Summary statistics from Experiment 2

	All B-players (n = 32)		Promise (n = 24)			No promise (n = 8)	
	Ret|OT	Ret|Late	Ret|OT	Ret|Late	Prom Amt (n = 20)	Ret|OT	Ret|Late
Mean	7.70	6.61	8.42	7.21	9.63	5.56	4.81
Median	9.00	8.50	9.00	9.00	9.50	6.50	6.00
SD	3.28	3.84	2.80	3.60	0.53	3.89	4.21
# $\geq$ 5	27	25	22	20	20	5	5

Ret|OT Amount returned by Player B conditional on message being delivered on-time

Ret|Late Amount returned by Player B conditional on message being delivered late

Prom Amt Amount of a specific promise. Promises of "half" are coded as 9.5. For those 20 B-Players who promised a specific amount, mean (SD) Ret|OT is 8.55 (3.02) and mean (SD) Ret|Late is 7.55 (3.56)

 # $\geq$ 5 The number of B-Players for whom Ret|OT, Ret|Late, or Prom Amt equals 5 or more

For nondelivered messages with no promise, Table 3 shows a mean (median) amount returned of 5.67 (7), with SD 3.67. These values do not significantly differ from 4.59 (6), the mean (median) for nondelivered promises (two-tailed $p > 0.50$ for the means using a two-sample t test and two-tailed $p > 0.50$ for the medians using the robust rank-order test, Siegel and Castellan 1988). This finding does not support Hypothesis 3B, the internal consistency hypothesis.

Although not necessary to test Hypothesis 3, results indicate that delivery has different effects on trustworthiness for messages containing promises as compared to messages that do not contain promises. For promises, delivery is associated with an increase in mean and median amount returned of 2.79 (7.38 − 4.59) and 3 (9 − 6), respectively. For messages that do not contain promises, the mean and median differences between delivered and nondelivered messages are − 0.56 (5.11 − 5.67) and − 1 (6 − 7), respectively. The means differ at two-tailed $p < 0.01$ (two-sample t test) and the medians differ at two-tailed $p < 0.001$ (robust rank-order test). These results do not support I&P's finding that delivery status did not affect trustworthiness differently for promises and non-promises.

Thus far, the data strongly support the communication rationale for promise keeping. However, we have yet to address directly whether delivered promises are "kept" more often than undelivered promises. To do this, we investigate the 32 individual trustees who made promises, assessing the effect of delivery on their tendency to keep their promise. We operationalize "keeping" a promise in two ways: (1) returning an amount greater than or equal to 5 and (2) returning an amount greater than or equal to the amount specified or implied in the message. The first assumes that the basic promise is that the trustor will be made no worse off by investing; thus if the trustee returns at least 5, even if it is less than the specific promise, the promise is deemed to have been kept. This first definition provides for an analysis similar to the binary trustworthiness analysis presented earlier. The second operationalization of keeping the promise is strict and literal: the promise is deemed to have been kept if and only if the trustee returns an amount equal to or greater than the amount specified or implied in the message.

Table 5 Hypothesis 3 tests of kept promises. (A) Define "kept" promise as amount returned is greater than or equal to 5, (B) define "kept" promise as amount returned is greater than or equal to amount specified or implied in message

Message not delivered	Message delivered	
	Promise kept	Promise broke
(A)		
Promise kept	18	0
Promise broke	9	5
(B)		
Promise kept	11	0
Promise broke	11	9

This table contains observations only where a promise was made. Cells include counts of the trustees who kept or broke promises, with each promise-making trustee included in only one cell. For example, the 18 trustees in the promise kept/promise kept cell in this table (A) kept their promises both when the message was and was not delivered

In this table (B), trustee #17 (see Table 1) was not included because it is indeterminable whether the promise was kept if the message was not delivered

We counted the number of kept promises for delivered and undelivered messages under both definitions. Table 5 summarizes the results. Under the first definition (Panel A), 27 (18) of 32 promises were kept for delivered (undelivered) messages. The difference in proportions between delivered and undelivered messages is significant at $p < 0.004$, as described earlier. Under the second definition (Panel B), the difference in proportions of kept promises for delivered messages (22/31) and undelivered messages (11/31) is significant at two-tailed $p < 0.001$.[10] These findings further support the communication hypothesis, in that delivered promises are more likely to be kept than are undelivered promises.

4.2 Second experiment

The results of our first experiment indicate that promises must be heard by another to be relevant. To further understand the role of the promisee knowing of the promise, we administered a second experiment wherein all subjects were informed that 50% of the messages would be delivered *before* the trustor chose whether to invest and 50% of the messages would be delivered *after* the trustor chose whether to invest. That is, the message would always be delivered, but sometimes not until after Player A made the decision to invest.

If the legitimizing aspect of promises is *solely* that the promise influences the behavior of the promisee, then (1) late-delivered promises should lead to less trustworthiness than on time-delivered promises in experiment 2 and (2) late-

[10] Table 5 indicates that trustee #17 was omitted from the second-definition tests. This trustee's promise was "will be fair, I promise." Because there are two reasonable ways to define fairness (return at least 5 or return approximately half, i.e., 9 or 10), we were not comfortable in classifying this trustee's promise contingent on non-delivery because the Ret|Inter amount was 7. In the Table 5 analysis we required observations under both delivery conditions and so had to delete that observation.

delivered promises in experiment 2 should lead to the same level of trustworthy behavior as non-delivered promises in experiment 1. On the other hand, if the legitimizing aspect of promises is *solely* that the promisee becomes aware of the promise, then (1) late-delivered promises should lead to the same level of trustworthiness as on time-delivered promises in experiment 2, and (2) late-delivered promises in experiment 2 should lead to greater levels of trustworthiness than non-delivered promises in experiment 1. We refer to the behavior-influencing explanation as the *contracting explanation* and the awareness explanation as the *communication sufficient explanation*.

Hypothesis 4A (*Contracting explanation*) Late delivery of promises leads to less trustworthiness than on-time delivery of promises, and late delivery of promises leads to the same level of trustworthy behavior as undelivered promises.

Hypothesis 4B (*Communication sufficient explanation*) Late delivery of promises leads to the same level of trustworthiness as on-time delivery of promises, and late delivery of promises leads to greater levels of trustworthiness than undelivered promises.

Three sessions of experiment 2 were run, with 22, 14 and 28 student-participants. The procedure and parameters were the same as for experiment 1 except for the different treatments noted above.

The raw data from experiment 2 are shown in Table 2 and summary statistics are in Table 4. The mean (SD) returned amount conditional on on-time delivery is 7.70 (3.28) and the mean (SD) returned amount for late delivery is 6.61 (3.84). The difference in returned amounts between the delivery protocols ($7.70 - 6.61 = 1.09$) is smaller than in experiment 1 ($6.88 - 4.83 = 2.05$). The returned amount for on-time delivery is higher in experiment 2 (7.70) than in experiment 1 (6.88). However, much of this difference is due to a higher proportion of 0 returns in experiment 1 than experiment 2 (8/42 vs. 4/32). In any case, the difference in returned amounts between the experiments is not significant.

Table 6 shows how the B-Players conditioned on delivery status across the experiments, specifically whether they conditioned their return differently for the two delivery methods. Considering all the B-Players (Table 6A), 24 of 41 (59%) in experiment 1 changed the amount they returned depending on the delivery status, whereas 11 of 32 (34%) did so in experiment 2. The difference is significant at $p < 0.05$ via a Chi squared contingency-table test. This suggests that B-Players perceived a more important difference between on-time delivery and non-delivery than between on-time delivery and late delivery. For only those B-Players making promises, the same relationship holds, as shown in Table 6B ($p < 0.017$).

Tests of Hypotheses 4 utilize only messages that contain promises, of which there are 24 (see Table 2). The first set of tests compares trustworthiness for on-time and late delivery and we employ the same two measures of trustworthiness developed earlier. For the first measure, Table 4 shows that 22 of the 24 B-Players making a promise returned 5 or more for on-time delivery and 20 of the 24 returned 5 or more for late delivery. The difference is not significant.

Table 6 Conditioning on delivery status by B-players. (A) All B-players, (B) B-players making a promise

	Experiment	
	1 (n = 41)	2 (n = 32)
(A)		
Different Amt returned	24	11
Same Amt returned	17	21
	Experiment	
	1 (n = 32)	2 (n = 24)
(B)		
Different Amt returned	21	8
Same Amt returned	11	16

This table shows whether B-Players changed the amount they agreed to return based on the delivery status. In experiment 1, the status is either delivered or not delivered. In experiment 2, the status is either delivered on-time or delivered late. Cells contain the frequency of B-Players who agreed to return the same amount, irrespective of delivery status, or changed the amount depending on the delivery status

For the second measure of trustworthiness, Table 4 shows that the mean (median) amount returned for on-time delivery is 8.42 (9), with SD 2.80. For late delivery, the mean (median) returned is 7.21 (9), with SD 3.60. The difference in means is significant (two-tailed $p < 0.05$ via a paired t test) as is the median test (two-tailed $p < 0.05$ using the Wilcoxon signed ranks test[11]).

On balance, these tests indicate that late delivery is associated with less trustworthiness than on-time delivery, which supports Hypothesis 4A (contracting).

The second set of tests compares trustworthiness of late delivery (from experiment 2) and non-delivery (from experiment 1). As noted earlier, the level of trustworthiness across the experiments differed (albeit insignificantly), as reflected, for example, in the mean return for (on-time) delivery of promises of 7.38 in experiment 1 compared to 8.42 in experiment 2; thus, our tests of late delivery versus non-delivery control for that difference across experiments.

The effect on trustworthiness of late delivery (experiment 2) is measured by taking the difference between Ret|OT and Ret|Late, 8.42 − 7.21 = 1.21 (Table 4). The SD of the differences is 2.47; the median difference is 0. The effect on trustworthiness of non-delivery (experiment 1) is the difference between Ret|Del and Ret|Inter (Table 3), or 7.38 − 4.59 = 2.79. The SD is 3.58 and the median difference is 1. A two-sample t test of means shows a significant difference (two-

[11] It is odd that the medians for the two conditions are both equal to 9, but the Wilcoxon signed ranks test yields a significant difference. This occurs because the test, strictly speaking, tests differences in signed ranks rather the median and therefore can detect other distributional differences. It is possible that the ranks of one group (except for those observations at the median) are substantially higher or lower than the ranks of the other group (again, except for those at the median). Also, note in Table 2 that there are no instances, considering promises only, in which Ret|Late is greater than Ret|OT, but several in the other direction, with some of them large in magnitude.

tailed $p < 0.06$) and a robust rank-order test shows a significant difference in medians (two-tailed $p < 0.07$). Thus, late delivery is associated with more trustworthiness than is non-delivery. It appears that communication of a promise, even if it cannot influence the promisee's behavior, has a role in legitimizing a promise. Overall, the results support both Hypotheses 4A and 4B: both contracting and communication sufficient explanations play a role.

In the post-experiment questionnaires, we often observe the difference in opinion on what makes a delivered promise valid (awareness or promisee reliance) that we intended to investigate. From observation 9, a promisor expressed an opinion that a delivered promise must be kept regardless of when it was delivered: "Player A is going to read my message anyways, whether before or after his/her decision is made…. I really am a person who keeps his promises. It's a morality issue, and I will not be comfortable with money earned by cheating." Observation 4 had similar comments. On the other hand, from observation 14, only promises that affected promisee behavior are legitimate: "… I kept $9 and gave them $10 (as promised) if they read it before deciding. If they didn't read my message (before), they wouldn't be relying on my promise, so I could keep all $19 for myself." One promisor, observation 29, gave a clear articulation of the internal consistency rationale stating, "even if the other person doesn't see it, I still do."

4.3 Characterizations across experiments

Combining the data from experiments 1 and 2 allows for several characterizations that might further help explain why individuals keep their promises. The first three characterizations refer to promises that were delivered on time, while the last ties together the three delivery methods: on-time, late and no delivery. We use these characterizations along with the other reported results to form conclusions.

Characterization 1: Promise keeping tends to be all-or-nothing Across the two experiments there are 40 specific promises. Of those, seven are promises to return at least nine but where zero was actually returned. Of the remaining 33 specific promises, in 25 the promisors returned the exact amount promised (counting a return of nine or ten as exact if a promise of "half" was given). In six of the remaining eight specific promises, promisors returned an amount that differed by one from their promise.

Characterization 2: Non-specific promises are non-specific for a reason Excluding the promisors who returned nothing, specific promises yielded an average return of 9.33 (on an average promise of 9.47). Non-specific promises yielded an average return of 7.85. The reason the return is lower for non-specific promises is that, of the 10 promisors who essentially promised to return more than what was invested (5), four returned 6. In the case of the 33 specific promises, none returned less than 7, and only two returned 7. When individuals promise to return more than what is given, they often mean just barely.

Characterization 3: Specific rather than non-specific promises are used by deceivers All of the promises that are associated with a zero return are specific promises; none of the non-specific promises result in a zero return. Conditional on receiving something, a specific promise is better than a non-specific promise, but

unconditionally not so. Perhaps not surprisingly, deceivers hide behind specific promises.

Characterization 4: The legitimacy of late promises is all-or-nothing on an individual basis As noted, the *average* difference between the amount returned for on-time versus late delivery is less than for on-time versus no delivery. But those who do condition, do so to approximately the same degree in each treatment. Specifically, conditional on a promise being made and a non-zero difference in returns based on delivery status, the mean difference in experiment 1 is 4.29 and the mean difference in experiment 2 is 3.63; the difference in differences is not significant. However, in experiment 1, 21 of 32 promisors condition on delivery status (19 in the predicted direction). In experiment 2, only 8 out of 24 promisors condition on delivery status (see Table 6B).

5 Conclusions

The purpose of these experiments was twofold. The first purpose was investigate the explanatory power of internal consistency and communication rationales for promise keeping. There appears to be no support for the internal consistency rationale. Undelivered promises appear to produce the same behavior as non-promises and both appear to be consistent with similar experiments that did not include promises. The second purpose was to investigate the legitimizing element in delivered promises; promisee awareness of implicit contracting. Our results suggest that on an individual basis some promisors view promisee delivery while others implicit contracting as legitimizing the promise. Overall, promises are associated with four types of behavior including behaviors where the promise appears to have no relevance to promisor behavior:

1. *Promises as a form of deception* Promises are not kept regardless of the delivery status. Promisors return nothing to promisees and so the promises are a ruse.
2. *Promises as a statement of intent* Promise making does not affect the behavior of the promisor; the promisor is trustworthy, but would be trustworthy absent a promise.
3. *Promises are kept if communicated* Promises are kept only if the promisee learns of the promise at *any* point.
4. *Promises are kept if communicated prior to the promisee's action* This behavior most resembles contracting.

There are three main conclusions from the experiments. First, as explained above, communication of the promise is necessary to make it legitimate. Unheard promises are the equivalent of unmade promises. Second, promisors seem to either view late promises as legitimate as on-time promises or as the equivalent of non-delivered promises. There does not seem to be an intermediate level of promise keeping for late arriving promises. We infer this conclusion from a between-subject analysis (as discussed in characterization 4)—further research might use a within-subject approach to determine if this conclusion holds. Third, promise-keeping is

basically an all-or-nothing affair. Conditional on on-time delivery, promisors either closely adhere to the promise or return nothing. These results are difficult to reconcile with the expectations rationale of CD for promise-keeping. Comparing the second-order expectations of those promisors who return nothing regardless of delivery status to those who adhere to their promise could shed further light on this issue.[12]

Two other issues are worthy of future investigation as well. Some trustees appear to disregard the welfare of trustors in that they return nothing under all circumstances, yet they refrain from making a promise. Are they selfish with respect to financial gain, but still receive disutility from lying? Second, there are quite a few trustees who return more than the trustor's original investment yet make no promise. This is particularly surprising, given the trustor is not faced with a situation where it is infeasible to fulfill a promise.

Finally, we should note there is the potential for an experimenter-induced demand effect given our use of the strategy method. The potential of demand effects related to the use of the strategy method is that the strategy method cues participants that conditional responses are expected (Zizzo 2010). Therefore, caution is necessary in interpreting the results, so further experiments without the use of the strategy method might be worthwhile.

References

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity and social history. *Games and Economic Behavior, 10*(1), 122–142.

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

Charness, G., & Dufwenberg, M. (2006). Promises and partnership. *Econometrica, 74*(6), 1579–1601.

Charness, G., & Dufwenberg, M. (2010). Broken promises: An experiment. *Economic Letters, 107*(2), 281–283.

Douthit, J., Kearney, L., & Stevens, D. (2012). Can agent cheap talk mitigate agency problems in the presence of a noisy performance measure? An experimental test in a single- and multi-period setting. *Journal of Management Accounting Research, 24,* 135–158.

Dufwenberg, M., & Gneezy, U. (2000). Measuring beliefs in an experimental lost wallet game. *Games and Economic Behavior, 30*(2), 163–182.

Ederer, F. & Stremitzer, A. (2016). *Promises and expectations*. Yale University working paper.

Ellingsen, T., & Johannesson, M. (2004a). Promises, threats and fairness. *The Economic Journal, 114*(April), 397–420.

Ellingsen, T., & Johannesson, M. (2004b). Is there a hold-up problem? *The Scandinavian Journal of Economics, 106*(3), 475–494.

Ellingsen, T., Johannesson, M., Tjøtta, S., & Torsvik, G. (2010). Testing guilt aversion. *Games and Economic Behavior, 68*(1), 95–107.

Hays, W. (1963). *Statistics*. New York: Holt, Rinehart and Winston.

Ismayilov, H., & Potters, J. (2016). Why do promises affect trustworthiness, or do they? *Experimental Economics, 19*(2), 382–393.

Johnson, N., & Mislin, A. (2011). Trust games: A meta-analysis. *Journal of Economic Psychology, 32*(5), 865–889.

[12] As mentioned, we did not collect expectations data.

Kawagoe, T., & Narita, Y. (2014). Guilt aversion revisited: An experimental test of a new model. *Journal of Economic Behavior & Organization, 102,* 1–9.

Khalmetski, K., Ockenfels, A., & Werner, P. (2015). Surprising gifts: Theory and laboratory evidence. *Journal of Economic Theory, 159,* 163–208.

Klein, D., & O'Flaherty, B. (1993). A game-theoretic rendering of promises and threats. *Journal of Economic Behavior and Organizations, 21*(3), 295–314.

Lundquist, T., Ellingsen, T., Gribbe, E., & Johannesson, M. (2009). The aversion to lying. *Journal of Economic Behavior and Organizations, 70*(1–2), 81–92.

Siegel, S., & Castellan, N. J. (1988). *Nonparametric statistics for the behavioral sciences* (2nd ed.). New York: McGraw-Hill.

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

Zizzo, D. (2010). Experiment demand effects in economic experiments. *Experimental Economics, 13,* 75–98.