Promises, Expectations & Causation

Giovanni Di Bartolomeo,[a] Martin Dufwenberg,[b] Stefano Papa[a] & Francesco Passarelli[c]

June 22, 2017

Abstract: Why do people keep their promises? Vanberg (2008) and Ederer & Stremitzer (2016) provide causal evidence in favor of, respectively, an intrinsic preference for keeping one's word and an expectations-based account (suggested by Charness & Dufwenberg 2006) based on guilt aversion. The overall picture is incomplete though, as no study disentangles effects in a design that provides exogenous variation of both (the key features of) promises and beliefs. We present an experimental design that fills the gap.

***JEL* codes:** A13, C91, D03, D64.

Keywords: Promises, expectations, guilt aversion, moral commitment, causation

[a] Sapienza University of Rome; giovanni.dibartolomeo@uniroma1.it, stefano.papa@uniroma1.it

[b] University of Arizona, University of Gothenburg, CESifo; martind@eller.arizona.edu

[c] University of Turin and Bocconi University, Milan; francesco.passarelli@unibocconi.it

1. Introduction

Promises often foster trust and cooperation. A recent literature explores why. Two explanations have been proposed, and experimental tests provided some support and some controversy:

- Charness & Dufwenberg (2006) (C&D) propose an *expectation-based explanation* (EBE): A promise received by *i* changes his expectations, and *j* feels guilty if she hurts *i* relative to his expectations.[1]
- Vanberg (2008) proposes a *commitment-based explanation* (CBE) whereby "people have a preference for promise keeping per se" (p. 1468).[2]

CBE posits that promises have a direct causal effect on promise keeping because people have an intrinsic motivation to keep their word. EBE postulates that the effect of a promise is mediated by its impact on expectations. The driving force is a general motivation to live up to others' expectations. The empirical implications of these two theories are substantially different. CBE implies that people are likely to honor a promise once they have given their word. EBE predicts that they will keep the promise only if it raised the counterparts' expectations. In principle, however, the two explanations are not mutually exclusive. It is plausible that an individual has some inclination both to keep her word and to live up to the higher expectations that her promise may have created.

In C&D's experiment messages are observed and beliefs elicited. This is central to their tests, but the conclusions rely on home-grown variation of messages and beliefs across subjects. Herein lie two potential problems. First, since subjects are randomly assigned neither to their messages nor to their beliefs, C&D's result may reflect correlation rather than causation. Second, as a result of the first problem, C&D cannot independently evaluate EBE and CBE.

Vanberg (who coined the terms EBE and CBE) focuses on the second problem, arguing that C&D's results are confounded in that CBE rather than EBE may be the driving force. He proposes an ingenious "partner-switching" design (described below) which induces exogenous variation in whether a player who sent a promise is actually paired with whoever received that promise. He is thus able to test CBE as a cause of promise-keeping, and reports support. Vanberg did not address the empirical relevance of EBE

[1] The desire to not hurt others relative to what they expect is an instance of Battigalli & Dufwenberg's (2007) model of "guilt aversion." Several papers report experimental evidence supporting guilt aversion in the absence of promises (see, e.g., Dufwenberg & Gneezy 2000, Guerra & Zizzo 2004, Bacharach *et al*. 2007, Reuben *et al*. 2009, Chang *et al*. 2011, Bracht & Regner 2013, Khalmetski *et al*. 2015). Ellingsen *et al*. (2010) do not find support, but Khalmetski *et al*. (2015) offer a possible reconciliation.

[2] He cites Braver (1995), Ostrom *et al*. (1992), and Ellingsen & Johannesson (2004) for exploring related notions. C&D too discuss the idea (Section 5.2), but argue against it.

though. He does not report tests analogous to C&D, regarding within-treatment correlations of choices and beliefs. Even if he did, such data would suffer from the same problem as in C&D: variation in expectations would not be exogenously induced.

Ederer & Stremitzer (2016) (E&S) propose a novel design, including an "unreliable random device" (see footnote 4), which creates exogenous variation in players' expectations. They are able to test EBE as a cause of promise keeping, and report support. E&S do not address the empirical relevance of CBE though. Their design includes no analog to Vanberg's partner-switching feature.

To sum things up, Vanberg and E&S provide evidence supporting, respectively, CBE and EBE. The overall picture is incomplete, however, since neither study disentangles CBE and EBE in a design that provides exogenous variation of *both* promises and beliefs. Our main goal is to fill this gap. We report experimental evidence from a design with exogenous variation of promises as well as of beliefs, allowing a test of the relative importance of EBE vs. CBE.

Section 2 describes design and procedures, section 3 presents result, and section 4 sums up.

2. Experimental design and procedures

2.1 Experimental design

We use Vanberg's binary-choice random-dictator game with a partner-switching mechanism. To explain the rules precisely, it is helpful to first consider the game depicted in Figure 1: a binary-choice random-dictator game *without* partner-switching.

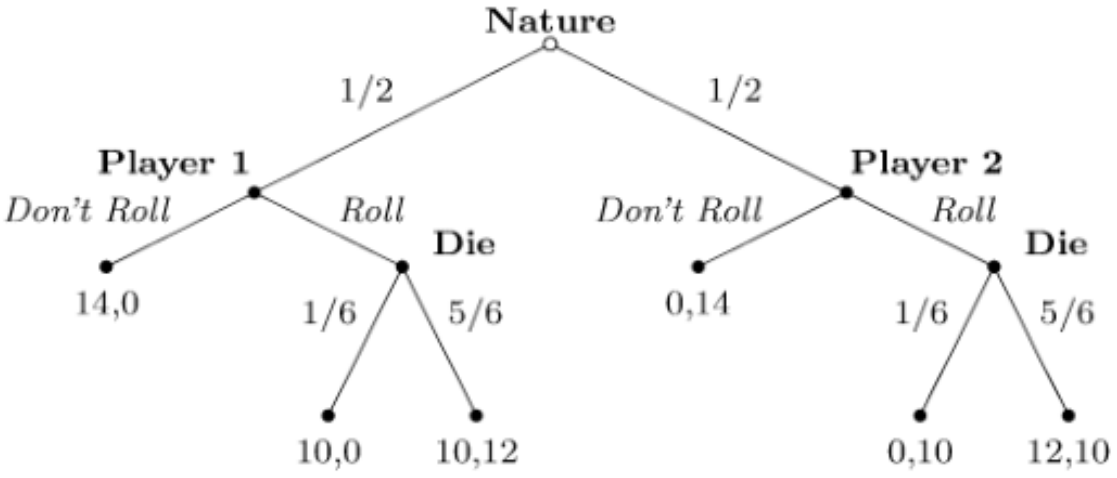

Figure 1 – A Binary-choice random-dictator Game

It is randomly determined whether player 1 or 2 will be the "dictator" who chooses between *Don't Roll* and *Roll*. The other player will be the "recipient." Payoffs are as indicated.[3]

Vanberg's design makes two critical changes relative to Figure 1. First, before subjects are told their roles, they are given the opportunity to communicate and (if they so choose) exchange promises. Then, they are randomly assigned their roles, dictator or recipient. This is when the second change appears. To generate exogenous variation in promises, *some of the recipients are randomly re-matched with a new dictator*, according to a given switching probability (see below). Each dictator whose co-player was switched then has the opportunity to read the messages that occurred between the new recipient and the dictator with whom that new recipient was initially matched. Finally, each dictator (whether or not involved in a switch) chooses between *Roll* and *Don't Roll*, like in Figure 1.

The design achieves exogenous variation in whether a player who sent a promise is actually paired with someone who received that promise. It involves asymmetric information between dictators and recipients; all subjects know that they will be re-matched with a certain probability, but after re-matching only dictators are told whether their recipient was switched. Therefore, recipients' first-order beliefs (and in turn dictators' second-order beliefs) can depend on whether they have received a promise or not, but not on whether there was a switch.

In Vanberg's study, the switching probability is always 50%. In order to achieve exogenous variation not only regarding promises but also regarding expectations, we crucially introduce a random change in the switching probability across treatments:[4] either *high* (75%) or *low* (25%). Recipients' first-order beliefs, and dictators' second-order beliefs, at the time the dictator makes her choice, may now plausibly and directionally depend on the value of the switching probability. Namely, in light of the relevance of CBE (as documented by Vanberg), it is plausible that people expect dictators to be more inclined to keep their own promise than a promise made by someone else. Hence recipients who received a promise should expect it to be kept with higher probability if the switching probability is low (i.e. 25%) rather than high (i.e. 75%). And, if dictators understand that, their second-order beliefs should vary by switching probability in the same direction as recipients' first-order beliefs.

This explains how our treatment variable, the switching probability, achieves exogenous variations in first- and second-order beliefs. We use that feature to test the effect of second-order beliefs in promise

[3] Payoffs and names of choices in subgames match those found in C&D's related trust game.

[4] The idea to use the partner-switch probability as a treatment variable is inspired by E&S. Their design is in other regards very different than ours, however. Instead of using Vanberg's game, they introduce a modified version of the C&D's trust game. Instead of a switching probability, they have a move by Nature determining whether the trustee will have a choice. The associated probability can be high or low, and only the trustee observes Nature's choice; this too yields exogenous belief-variation.

keeping. A higher *Roll* rate in the presence of higher second-order expectations would imply that the dictators' behavior is causally affected by their expectations, as predicted by EBE.

Following Vanberg, we furthermore use exogenous variation in promises to test CBE. A higher *Roll* rate in the cohort of non-switched dictators would imply that people have a stronger taste to keep their own promises than promises made by others. Then, as predicted by CBE, making a promise *per se* would determine promise-keeping behavior.

2.2 Procedures

The experiment was conducted at the *CIMEO Experimental Economics Lab* of Sapienza University of Rome (December 2015). The design involved 192 undergraduate student subjects (6 sessions, 32 subjects each), recruited using an online system. Upon arrival, subjects were randomly assigned to 32 isolated computer terminals.[5] Three assistants handed out instructions (cf. Supplementary Material online) and checked that participants correctly followed the procedures. Before playing any game, subjects filled out a short questionnaire testing their comprehension.

Each session consisted of eight rounds, with perfect stranger matching. Payoffs, as shown in Figure 1, were computed in tokens (where 1 token = 0.5 euro). At the end of each session, one of the rounds was randomly chosen for payment. First- and second-order beliefs were elicited by asking subjects to guess their counterparts' actions and guesses, respectively. Incentives were provided for all rounds except the one chosen for payment, implying that subjects had no incentive to hedge against bad outcomes and thus to misreport their beliefs.[6] All subjects received a fixed show-up fee of 2.50 tokens.

Each round implemented the following sequence of five stages:

1. **Communication.** Subjects were randomly matched to form 16 chatting pairs, with random determination of who would start the chat. As in Vanberg's design, each chat consisted of four one-way messages in sequence. Each message could be of at most 90 characters, and were catalogued as involving promise or not (see below).

2. **Role assignment and revelation of the switching probability.** After communication, roles were randomly assigned in each pair and subjects informed of that. Depending on treatment, the switching probability was announced (either 25% or 75%).

3. **Belief elicitation.** This stage has two parts:

[5] The experiment was programmed and conducted with the software z-Tree (Fischbacher, 2007).

[6] Our elicitation procedure is described in details in Appendix A.

a. First-order beliefs: each recipient was asked to guess whether his or her dictator would choose *Roll* or *Don't Roll*.

b. Second-order beliefs: dictators were asked to guess the guess of the person with whom they had formed a chatting pair.

4. **Switching.** Some recipients were switched: 25% or 75%, depending on treatment. Only dictators were informed whether or not a switch occurred. Dictators with switched recipients were then allowed to read the prior conversation of their new recipients.

5. **Dictators' action.** All dictators made their choice: *Roll* or *Don't Roll*. All subjects were informed of their payoff for the round. Recipients were not informed whether they had been switched or not, nor could they infer the dictator's choice when their payoffs were zero.[7]

Since we elicited second-order beliefs before switching occurred, we plausibly induce exogenous variation in expectations prior to switching. We assume (and verify in a test[8]) that whatever exogenous variation we have in second-order beliefs is unchanged after the switch/no-switch condition is revealed. This is a reasonable since dictators know recipients do not observe whether there is any switch, so their first-order beliefs should be independent of that.

Our sample consists of 768 dictator decisions, equally split between switch and no-switch subsets. Messages were classified according to Vanberg's protocol, each pair of messages sent by a subject in a round was treated as a unit. Hence, we had 1,536 messages. We asked a research assistant to code these according to whether they conveyed a promise, obtaining 575 promises out of 768 messages.

3. Results

3.1 Expectations and switching probabilities

Before we explore the causal effect of promises and second-order beliefs, we must establish that our design provides an adequate test bed in the sense of actually inducing exogenous variation in first- and second-order beliefs, as described in subsection 2.1. Table 1 presents those beliefs, in the form of fractions of dictators who chose to roll.[9] The table reports also standard deviations (s.d.) and observations (obs.). All reported statistics adopt the Wilcoxon rank-sum test (Mann-Whitney). We check robustness

[7] Recipients could obtain a zero payoff in two cases: (i) their dictator had chosen *Don't Roll*; (ii) their dictator had chosen *Roll* and the outcome of the die-roll was "1".

[8] See Table B2 in Appendix B.

[9] Appendix A describes how the beliefs in Table 1 (and in Tables B1 and B2 of Appendix B) were elicited and computed.

using the Fligner-Policello test, a robust rank check for unmatched data which does not require two compared populations to have the same variance (see Hollander *et al*. 2014). These tests always confirms the Wilcoxon rank-sum test, so we omit to report them.[10]

Consider recipients who received and dictators who made a promise (columns (1) and (2), respectively). Their average first- and second-order beliefs are significantly higher when the chance of being re-matched is low rather than high: 0.74 is significantly higher than 0.55 (Z=6.71, p=0.000) and 0.80 is significantly higher than 0.67 (Z=4.82, p=0.000). Thus, a positive correlation between switching probability and beliefs exists when a promise was made.

This is not the case when no promise was made. Columns (3) and (4) show that first- and second-order beliefs are not affected by the switching probability.[12] Only if a promise were made did people take the value of the switching probability into account while forming expectations.

Table 1 also unveils a positive correlation between promises and expectations. On average, recipients who received a *Roll*-promise, and dictators who promised to choose *Roll*, have higher first- and second-order beliefs compared to those who did not receive/make such a promise. This positive correlation between promises and expectations confirms the findings of C&D and Vanberg. It exists both when the switching probability is high and when it is low.[13]

Table 1 – First- and Second-order beliefs (768 obs.)

	PROMISE (575 OBS.)		NO PROMISE (193 OBS.)	
	1ST-ORDER BELIEFS	2ND-ORDER BELIEFS*	1ST-ORDER BELIEFS	2ND-ORDER BELIEFS*
	(1)	(2)	(3)	(4)
SWITCHING PROBABILITY				
LOW (25%)	0.74 (s.d. 0.30, obs. 282)	0.80 (s.d. 0.27, obs. 282)	0.42 (s.d. 0.38, obs. 102)	0.38 (s.d. 0.38, obs. 102)
HIGH (75%)	0.55 (s.d. 0.34, obs. 293)	0.67 (s.d. 0.34, obs. 293)	0.41 (s.d. 0.32, obs. 91)	0.40 (s.d. 0.35, obs. 91)

* Note that second-order beliefs were elicited before dictators knew whether they had been re-matched or not.

[10] As we do here, E&S use the Wilcoxon rank-sum and Fligner-Policello tests. Vanberg reports Wilcoxon signed-rank tests paring data by sessions, where session comparisons are based on six matched pairs. For less than eight pairs, however, the outcome of the test is unreliable (see Welkowitz *at al*. 2006, pp. 444-448). For this reason, we chose not to use the Wilcoxon signed-rank test. Note that, as a robustness check, Vanberg also reports the Wilcoxon rank-sum tests (for the first-round observations).

[12] In column (3), 0.42 is not significantly different from 0.41 (Z=0.16, p=0.436) and in column (4) 0.38 is not significantly different from 0.40 (Z=0.53, p=0.297).

[13] Regarding first-order beliefs, 0.74 is significantly higher than 0.42 (Z=7.21, p=0.000) and 0.55 is significantly higher than 0.41 (Z=3.47, p=0.000). Regarding the second-order beliefs, 0.80 is significantly higher than 0.38 (Z=9.11, p=0.000) and 0.67 is significantly higher than 0.40 (Z=6.14, p=0.000).

3.2 Testing the preference to live up to expectations

We now investigate dictators' promise-keeping. We consider dictators who at the time they made their choice were matched with a co-player who received a promise (from the dictator, if there was no switch); from someone else, if there was a switch) during the chatting stage.[14] Table 2 reports average *Roll* rates, standard deviations, and observations. There are four cells/categories structured across two dimensions: (i) dictators who made a promise and were not re-matched (column (a)) vs. dictators who were re-matched and paired with a co-player who received a promise by someone else (column (b)); (ii) treatments with high second-order beliefs (i.e., low switching probability – row (1)) vs. treatments with low average second-order beliefs (i.e., high switching probability – row (2)).

Table 2 – Dictators' behavior (*Roll* rates)
Partners received a promise (575 obs.)

	NO SWITCH	SWITCH
SECOND-ORDER BELIEFS	(a)	(b)
(1) High SOBs	51% (s.d 0.50, obs. 214)	41% (s.d. 0.50, obs. 68)
(2) Low SOBs	58% (s.d. 0.50, obs. 72)	28% (s.d. 0.45, obs. 221)

According to EBE, the causal factor is a change in the expectations associated with a promise made. Tests for such effects appear in the *columns* of Table 2. Consider first the dictators who made a promise and were not re-matched (column (a)). We do not find support for the hypothesis that these *Roll* rates are higher when SOBs are high: the 51% *Roll* rate of those with high second-order beliefs is obviously not higher than the 58% *Roll* rate of dictators with low second-order beliefs.

By contrast, expectations have a significant influence on dictators' behavior when dictators are re-matched with a new partner who has been made a promise by another person. As seen in column (b), the *Roll* rate of re-matched dictators is 41% when their average second-order beliefs are high, which is significantly higher than the 28% when their second-order beliefs are low (Z=2.12, p=0.017).

3.3 Testing the preference to keep one's word

According to CBE, the casual factor leading to promise keeping is the act of making a promise; people have a preference for keeping their word. Tests for such effects are conducted using the *rows* of Table 2, comparing dictators who chose whether to keep their own promises with dictators who chose whether to

[14] Data on what happens when no promise was made during communication is not immediately relevant to our main focus on CBE and EBE. We report on that data in Appendix C.

keep somebody else's promise. Vanberg conducts test of that sort, but does not control for different exogenously induced second-order beliefs. He finds that the former dictators were more likely to choose *Roll*, and concludes that "…the effects of promises cannot be accounted for by changes in payoff expectations. This suggests that people have a preference for promise keeping per se (p. 1478)." However, he did not check whether this is true for different values of second-order beliefs. We do that here, reaching a qualified conclusion: people have a stronger taste for keeping their own promises than the promises made by someone else when their second-order beliefs are low. When they are high, people are equally likely to keep their own promises and the promises made by others.

Consider row (2), with low second-order beliefs. Among dictators who were not re-matched the fraction who chose to keep the promise is 58%, significantly larger than the 28% who chose *Roll* in the "Re-Match" group (Z=4.73, p=0.000). The difference occurs despite dictators' average second-order beliefs not being significantly different between the two groups.[15] The picture is different when second-order beliefs are high (row (1)). The *Roll* rate of dictators that were not re-matched is 51%, not significantly greater than the 41% of dictators who were re-matched (Z=1.40, p=0.160). The intrinsic motivation for keeping one's word leads to higher *Roll* rates, but only when expectations are low.

4. Conclusion

Why do people keep their promises? The question tends on the brazen, obscuring how in many cases people actually do not keep their promises. This is true in our data; in Table 2, the cell with the most frequent promise-keeping gets 58%, so 42% reneged. In other situations, reneging rates may even higher.[16] Yet, it is undeniably the case that in many contexts, including the one we studied, promises can have a dramatic effect on trust and cooperation, even if not foolproof. It is important to understand why.

The expectations-based explanation (EBE) of promise-keeping says that people have a general tendency to fulfill others' expectations. Thus, they will keep their promise if it substantially increased their counterpart's expectations. Commitment-based explanation (CBE) says that people attach a value to the act of keeping their word, per se. Promises have an independent effect on behavior. Thus people will exhibit a higher inclination to keep their promises than the promises made by others. We argued that in

[15] See Table B2 in Appendix.

[16] Charness & Dufwenberg (2010) suggest that, outside the lab, this includes "used car sales, promises made by politicians, tax returns sent to the IRS, and testimony in traffic courts." And in-the-lab they show that this is the case when promises are "bare" (viz., circling a pre-fab message) as opposed to "rich" (free-form, as in C&D). Realistically, intrinsic motivations to keep one's word or to live up to others' expectations are not unconditional. Not only do motivations incorporate absolute principles of justice, they also depend on contextual factors which may strengthen or weaken them (e.g., habits, norms, culture, merit, need, …). How these factors influence the pro-social choice of keeping a promise is an important issue yet to be unveiled by future research.

principle the two explanations are not mutually exclusive, since the inclination to keep own promises may still positively depend on expectations.

We consider the prominent lab-context introduced by Vanberg (2008), with rich free-form communication followed by a potential "partner-switch." We make the probability of a partner-switch a treatment variable, thereby inducing high vs. low recipient expectations (for low vs. high switch-probability). Differently from previous work, our design allows us to test the empirical relevance of both EBE and CBE. We find that if beliefs are high (as induced by a low switch-probability) promise-keeping rates are high, and if beliefs are low (as induced by a high switch-probability) promise-keeping rates are high anyway. This is suggestive of the fact that the value a person attaches to the act of keeping her promise is *independent* of her beliefs. People's intrinsic motivation to keep their word appears to be the only force leading to promise keeping.

Does this imply that beliefs never matter in pro-social behavior? The answer is "No". Our results show that when people have to choose whether to keep a promise made by another person, their choice positively depends on their beliefs. People have no intrinsic motivation to keep their word since they did not make that promise, but nevertheless display a preference to live up to their counterparts' expectations, as postulated by the theory of guilt aversion. Specifically, we find that, when second-order beliefs are high, the preference to fulfill expectations induced by others' promises is as strong as the preference to keep one's promises. Own promises do not have an independent effect in this case.

Our result that people's preference for keeping somebody else's promise depends on their second-order beliefs seems plausible and is descriptive of a wide range of real-word situations. Think for instance of a situation in which a newly appointed boss finds herself in the position of choosing whether to keep a promise made to an employee by her predecessor. Our results suggest that she is less likely to renege the promise when the employee's expectations are higher.

Our evidence thus supports the notion that the prevalence of pro-social choice is higher if *either* of *two* "reasons" apply:

- A person has made a promise to the recipient he interacts with.
- A person interacts with a recipient who expects much.

The first reason is indicative of a preference for keeping one's word. The second is indicative of guilt aversion. We find it comforting that pro-social choice may happen for many reasons.

Appendix A – Elicitation of beliefs

Elicitation of first-order beliefs: After communication, recipients were asked to guess what their (unknown) dictators would choose to do. They had been told the value of the switching probability in their treatment. Thus, they were aware that their paired subject could be switched according to that probability. Recipients could make their guess by ticking one of the five-point scale in Table A. This scale is the same as in Vanberg. Beliefs are then re-scaled to 1, 0.75, 0.5, 0.25, and 0. Thus the numbers shown in Table 1 (and in Tables B1 and B2 below) represent the averages of recipients' re-scaled responses. The payoffs correspond to a quadratic scoring rule for probability values 85%, 68%, 50%, 32%, and 15%, because due to the risk neutrality assumption, quadratic scoring yields flat payoffs as probabilities approach one (see Vanberg, p. 1472).

Table A – Incentives for first-order belief elicitation

The dictator will	choose *Roll*			choose *Don't Roll*	
	Certainly	Probably	Unsure	Probably	Certainly
Please tick your guess	◯	◯	◯	◯	◯
Your earnings if the dictator					
chooses *Roll*	0.65 tokens	0.60 tokens	0.50 tokens	0.35 tokens	0.15 tokens
chooses *Don't Roll*	0.15 tokens	0.35 tokens	0.50 tokens	0.60 tokens	0.65 tokens

Elicitation of second-order beliefs: Before dictators were told whether their paired subject had been switched or not, they were asked to guess his guess. Specifically, they had to guess which of the five points of Table A had been ticked by their counterpart. Correct guesses were paid 0.50 tokens.

Appendix B – Beliefs

Table B1 – First-order beliefs (obs. 768)

	RECEIVED A PROMISE		DID NOT RECEIVE A PROMISE	
	NO SWITCH	SWITCH	NO SWITCH	SWITCH
SWITCHING PROBABILITY	(1)	(2)	(3)	(4)
LOW (25%)	0.74	0.75	0.41	0.44
	(s.d. 0.30, obs.214)	(s.d. 0.31, obs.68)	(s.d. 0.37, obs.74)	(s.d. 0.42, obs.28)
HIGH (75%)	0.51	0.57	0.41	0.41
	(s.d. 0.36, obs. 72)	(s.d. 0.34, obs.221)	(s.d. 0.34, obs.24)	(s.d. 0.31, obs. 67)

As shown in Table B1, recipients' first-order beliefs are independent of whether pairs were re-matched or not. This result is of course consistent with the fact that recipients knew that only dictators would have been informed of a possible switch, in the case it occurred. Consider the first row. Beliefs of recipients who received a promise are 0.74 in the no-switch condition and 0.75 in the switch condition (Z=0.39, p=0.697). Beliefs of recipients who received no promise are 0.41 in the no-switch condition and

0.44 in the switch one (Z=0.12, p=0.905). Similar results hold in the second row. Beliefs of recipients who received a promise are 0.51 in the no-switch condition and 0.57 in the switch one (Z=1.24, p=0.217). Beliefs of those who received no promise are 0.41 in the no–switch condition and 0.41 in the switch one (Z=0.12, p=0.901).

The data reported in the columns of Table B1 is also consistent with the exogenous variation in expectations described in Table 1:

1. First-order beliefs of recipients who received no promise appear to be independent of the switch condition. In the no-switch condition they are 0.41 when the switching probability is 25% and 0.41 when it is 75% (Z=0.13, p=0.899). First-order beliefs in the switch condition are 0.44 when the switching probability is 25% and 0.41 when it is 75% (Z=0.12, p=0.903).

2. By contrast, average first-order beliefs of recipients who received a promise depend on the switching probability. In the no–switch condition they are 0.74 when the switching probability is 25% and 0.51 when it is 75% (Z=4.74, p=0.000). In the switch condition they are 0.75 when the switching probability is 25% and 0.57 when it is 75% (Z=4.00, p=0.000). As expected, these beliefs are independent of whether pairs were re-matched. Moreover, all these values are significantly higher compared to those of recipients who did not receive a promise.

Table B2 presents dictators' average second-order beliefs.

Table B2 – Second-order beliefs

	NO SWITCH (384 OBS.)*		SWITCH (384 OBS.)*	
	PROMISED TO ROLL (1)	DID NOT PROMISE (2)	PROMISED TO ROLL (3)	DID NOT PROMISE (4)
SWITCHING PROBABILITY				
LOW (25%)	0.80 (s.d. 0.27, obs. 214)	0.38 (s.d. 0.38, obs. 74)	0.79 (s.d. 0.27, obs. 68)	0.38 (s.d. 0.27, obs. 28)
HIGH (75%)	0.65 (s.d. 0.35, obs. 72)	0.47 (s.d. 0.35, obs. 24)	0.67 (s.d. 0.27, obs. 221)	0.37 (s.d. 0.27, obs. 67)

* Note that second-order beliefs were elicited before dictators knew whether they had been switched or not. The values in the table have been split in *no-switch* and *switch* by the experimenter.

When the chance of being switched is low, dictators who promised to *Roll* (columns (1) and (3)) display higher average second-order beliefs: 0.80 is significantly higher than 0.65 (Z=3.28, p=0.001) and 0.79 is significantly higher than 0.67 (Z=2.60, p=0.010).

The value of the switching probability does not affect second-order beliefs of dictators who are matched with recipients that received no promise (cf. columns (2) and (4)): 0.38 is not significantly

different from 0.47 (Z=1.32, p=0.187) and 0.38 is not significantly different from 0.37 (Z=0.03, p=0.973). This is consistent with the fact that recipients who received no promise do not care about who they will be paired with. Since no promise has been made, they expect dictators are equally likely to choose *Don't Roll*, independently of whether they have been re-matched or not, and independently of the switching probability.

Reading Table B2 by "rows" indicates that promises are positively correlated with expectations. Such correlation exists both when the switching probability is high and when it is low.[18]

Appendix C – No promise

What happens when no promise was made during communication? As expected, there is strong correlation between the existence of a promise and *Roll* rates, a result in line with C&D and Vanberg. Table C summarizes average *Roll* rates when no promise was made. It distinguishes between switch and no-switch conditions. It also reports standard deviations (s.d.) and observations (obs.). *Roll* rates drop substantially compared to when a promise was made (cf. Table 2). There is no significant difference between the average behavior of dictators who made no promise and were not switched (0.27) and dictators who were re-matched with a new recipient who had received no promise (0.23; Z=0.54, p=0.294). As for second-order beliefs, they are also quite low and not statistically different between the two groups (cf. Table B1). For this reason, we have only one single row in Table C. Finally, as the second order-beliefs of dictators who did not make a promise are lower than those of dictators who made a promise, a positive correlation between promise keeping and beliefs is observed.

Table C – Dictators' behavior (*Roll* rates)
Partners received no promise (193 obs.)

	NO SWITCH	SWITCH
SECOND-ORDER BELIEFS	(a)	(b)
Lowest SOBs *	0.27 (s.d. 0.44, obs. 98)	0.23 (s.d. 0.42, obs. 95)

* When no promise was made we obtain the lowest SOBs (see Appendix B, Table B2 –columns (2) and (4)). Appendix B also shows that the change in the switching probability does not yield any significant variation in expectations.

[18] Among non-switched dictators: 0.80 vs. 0.38 (Z=7.79, p=0.000) and 0.65 vs. 0.47 (Z=2.17, p=0.030), in the low and the high switching probability case, respectively. Among switched dictators: 0.79 vs. 0.38 (Z=4.65 p=0.000) and 0.67 vs. 0.37 (Z=5.83 p=0.000), in the low and the high switching probability case, respectively.

References

Bacharach, M., G. Guerra, and D. J. Zizzo (2007), "The self-fulfilling property of trust: An experimental study," *Theory and Decision*, 63: 349-388.

Battigalli, P. and M. Dufwenberg (2007), "Guilt in games," *American Economic Review*, 7: 170-176.

Bracht, J., and T. Regner (2013), "Moral emotions and partnership," *Journal of Economic Psychology*, 39: 313-326.

Braver, S. (1995), "Social contracts and the provision of public goods," in *Social dilemmas: Perspectives on individuals and groups*, ed. by D. Schroeder. New York: Praeger.

Chang, L. J., A. Smith, M. Dufwenberg, and A. Sanfey (2011), "Triangulating the neural, psychological and economic bases of guilt aversion," *Neuron*, 70: 560-572.

Charness, G. and M. Dufwenberg (2006), "Promises and partnership," *Econometrica*, 74: 1579-1601.

Charness, G. and M. Dufwenberg (2011), "Bare promises: an experiment," *Economics Letters*, 107: 281-283.

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

Ederer, F. and A. Stremitzer (2016), "Promises and expectations," Cowles Foundation Discussion Paper No. 1931.

Ellingsen, T. and M. Johannesson (2004), "Promises, threats and fairness," *The Economic Journal*, 114: 397-420.

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

Fischbacher, U. (2007), "z–Tree: Zurich toolbox for ready-made economic experiments," *Experimental Economics*, 10: 171-178.

Guerra, G., and D. J. Zizzo (2004), "Trust responsiveness and beliefs," *Journal of Economic Behavior and Organization*, 55: 25-30.

Hollander, M., D. Wolfe, and E. Chicken (2013), *Nonparametric Statistical Methods*, Wiley Series in Probability and Statistics, Wiley, New York.

Huang, P.H. and H. M. Wu (1994), "More order without more law: A theory of social norms and organizational cultures," *Journal of Law, Economics, and Organization*, 10: 390-406.

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

Ostrom, E., J. Walker, and R. Gardner (1992), "Covenants with and without a sword: Self–governance is possible," *The American Political Science Review*, 86: 404-417.

Reuben, E., P. Sapienza, and L. Zingales (2009), "Is mistrust self-fulfilling?," *Economic Letters*, 104: 89-91.

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

Welkowitz, J., B. H. Coen, and R. B. Ewen (2006), *Introductory Statistics for the Behavioral Sciences*, John Wiley & Sons, Hoboken NJ.