Promises and Partner-Switch: Vanberg Revisited

Giovanni Di Bartolomeo,[a] Martin Dufwenberg,[b] and Stefano Papa[c]

June 2, 2020

Abstract: Building on Christoph Vanberg's classic partner-switching mechanism, we experimentally test two theories that posit different reasons why promises breed trust and cooperation. The *expectation-based explanation* (EBE) operates via belief-dependent guilt aversion, while the *commitment-based explanation* (CBE) suggests that promises offer commitment power via a (belief-independent) preference to keep one's word. Previous research performed a similar test, which we however argue should be interpreted as concerning informal agreements rather than (unilateral) promises.

***JEL* classification:** A13; C91; D01; D64.

Keywords: Promises, partner-switching, expectations, commitment, guilt, informal agreements.

[a] Department of Economics and Law, Sapienza University of Rome; CIMEO; European School of Political Economy, Luiss, Rome; Council of Experts, Italian Ministry of Economy and Finance; giovanni.dibartolomeo@uniroma1.it.

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

[c] Jean Monnet Department of Political Science, University of Campania "Luigi Vanvitelli"; CIMEO; stefano.papa@uniroma1.it.

1. Introduction

Promises may foster trust and cooperation. A recent literature explores why. Charness & Dufwenberg (2006) (C&D) propose an *expectation-based explanation* (EBE): A promise feeds a self-fulfilling circle of beliefs about beliefs. Promises are honored because if a person broke his promise then he would experience guilt from letting down the co-player's expectation.[1] Therefore, the co-player trusts the promisor. Vanberg (2008) proposes an alternative *commitment-based explanation* (CBE): A promisor delivers because he likes to keep his word.[2] To experimentally test CBE vs. EBE it is crucial to develop a design that *exogenously* varies whether or not a player sent a promise to another. Vanberg ran an experiment which achieved that, by relying on an ingenious "partner-switching" feature. His results support CBE over EBE.

Or do they? On closer scrutiny one realizes that Vanberg did not hew all that closely to C&D's design, and that this affects the interpretation of his result. There are two differences. First, C&D and Vanberg explore different games. C&D focus on a binary trust game, where two players move in sequence. Vanberg instead explores a symmetrized dictator game, where only one player is active along any path of play, and where players initially do not know their role (dictator or recipient). Second, and most importantly, C&D and Vanberg explore different communication protocols. C&D study the effect of a single pre-play message that cannot be responded to. Vanberg instead allows subjects to send messages back-and-forth. The set-up resembles a conversation. Vanberg's players can reciprocate each other's promises, and if they do so their exchange has the flavor of an informal agreement.

These observations invite two reflections. First, Vanberg's result may be reinterpreted as evidence of a preference to honor informal agreements, rather than a broader desire to keep one's word. Second, in order to test CBE vs. EBE without regard to the informal agreement interpretation just given, a new experiment is called for. One should incorporate Vanberg's partner-switching feature in an environment that hews closer to C&D's game and communication-protocol.

We have run such an experiment, and now report our results.

[1] EBE's reference to guilt is grounded in the theory of guilt aversion. See Battigalli & Dufwenberg (2007) for a general approach, based on the framework of so-called called psychological game theory (Geanakoplos *et al*., 1989; Battigalli & Dufwenberg 2009, 2020).

[2] Ostrom *et al*. (1992), Ellingsen & Johannesson (2004), C&D (Section 5.2), and Di Bartolomeo *et al*. (2019a) discuss similar ideas.

Section 2 provides in-depth scientific background: hypotheses, designs (C&D's, Vanberg's, ours), related literature. Section 3 describes our procedures. Section 4 explains what we found, while Section 5, finally, offers an interpretation, emphasizing the distinction between one-sided promises and informal agreements, as well as how the relevance of these notions may depend on the nature of a strategic situation.

2. Scientific background

We first recall what C&D and Vanberg did, then explain what we add. Figure 1 depicts C&D's game (form) to the left and Vanberg's to the right. Note how they differ, as indicated above.

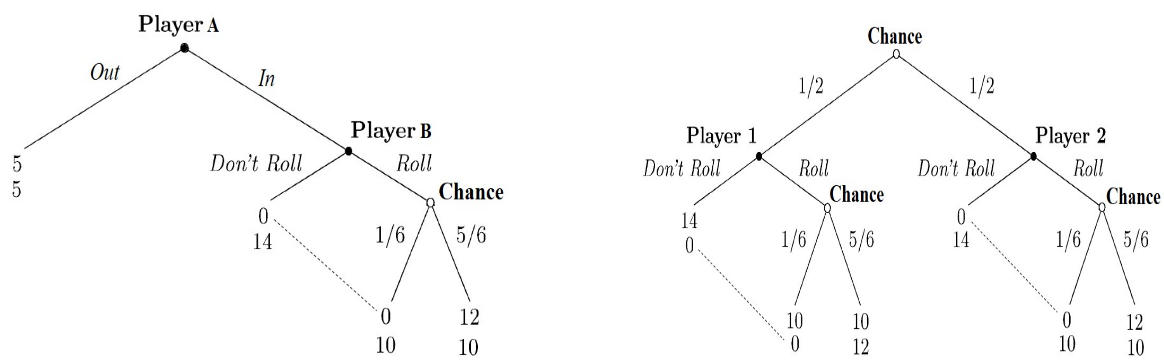

Figure 1 – The game trees: C&D's (to the left) and Vanberg's (to the right)

EBE. C&D explored experimental treatments with and without pre-play communication. In particular, in one treatment B could send a single pre-play message to A. Suppose that B experiences guilt if he chooses *Don't*, and that the amount of guilt increases the more strongly B believes that A believes that B will choose *Roll*.[3] A promise from B to A may then feed a self-fulfilling circle of beliefs about beliefs that B will *Roll*, and therefore A will choose *In*. C&D articulated this idea – aka EBE, – tested it, and found support.

[3] Note again the reference to guilt aversion (compare with footnote 1). Several other experiments, starting with Dufwenberg & Gneezy (2000), tested hypotheses related to guilt aversion, often without communication in the picture. See Cartwright (2019) for a survey.

CBE. Vanberg points out that C&D's story is confounded. Suppose B has an ingrained preference for keeping his word: If B promises to *Roll* then he will prefer to not renege. If A anticipates this, he will choose *In*. This idea – aka CBE – generates the same prediction as EBE.

Partner-switching. Vanberg came up with a clever experimental device to test CBE vs. EBE, enabling him to draw robust causal inference regarding the impact of a promise. Namely, he proposed that if subject *i* makes a promise to *j*, then, with 50% probability, *j* would be "switched" and replaced by another subject *k* who previously received a message from yet another subject *l*. Moreover, if there was a switch, *j*, but not *k*, would be informed of this. For cases where *l* sent a similar message to *k* as *i* sent to *j* (note: *i* could read *l*'s message to *k*) EBE suggests that *i* would behave the same way with or without a switch. CBE, by contrast, implies that *i* will fulfill his promise if and only if there was no switch.

Vanberg's design. Relying on partner-switching, it would seem natural to test the CBE vs. EBE prediction using C&D's game. However, that is not what Vanberg did. Instead, he used the game to the right in Figure 1. That is, if a subject *i* in player 1's position was selected to be the dictator (by the initial chance move), then the subject *j* (in player 2's position) with whom *i* had initially communicated would be switched to another subject *k* who had initially communicated with a fourth subject *l*. Moreover, instead of using C&D's single-shot messages from one player to the other, Vanberg allowed the two players four rounds of back-and-forth messaging. Based on this design, Vanberg reported support for CBE over EBE.[4] While this is not outrageous – C&D's hypotheses of EBE could, in principle, be reformulated for Vanberg's game – it is nevertheless a questionable choice. First, since C&D had the original result, while not stick to and bridge with their design? Second, and most importantly, the back-and-forth nature of Vanberg's communication protocol would seem to generate experiences that look more like informal agreements generated by conversations than one-sided promises. While exploring the impact of informal agreements may be useful, it does not address the theme of one-sided messages that C&D introduced.[5]

---

[4] See also Di Bartolomeo *et al*. (2019b) who report similar results from a related design.

[5] In Vanberg's design, even if *i*'s promise is not reciprocated by *j*, *i*'s promise will not be comparable to a promise in C&D's set-up. A promise which could have been reciprocated, but was not, is not the same as a promise that cannot be reciprocated as a matter of design.

Our design. We apply Vanberg's partner-switching feature to C&D's original game, thus providing a new independent test of CBE vs. EBE. A subject in B's position is allowed to send a single written free-form message to a subject in A's position. Subsequently, there was a 50% probability that the A-subject would be switched and replaced by another subject (also in the position of A) who previously received a message from yet another subject *l* (in the position of player B). If there was a switch, only the B-subject was informed.

Related literature. Several less closely related studies explored various other aspects of how different types of communication (e.g., face-to-face vs. anonymous, bilateral vs. unilateral, free-form vs. prefab, passage of time) affect play in various games. See, e.g., Ellingsen & Johannesson (2004), Charness & Dufwenberg (2010, 2011), Ismayilov & Potters (2016, 2017), Ederer & Stremitzer (2017), Krupka *et al*. (2017), Dufwenberg *et al.* (2017), Di Bartolomeo *et al*. (2019a, 2019b), and Ederer & Schneider (2020).

3. Procedures

Our experiment was conducted at CIMEO Experimental Economics Lab of Sapienza University of Rome in May 2019. On aggregate, it involved 226 undergraduate students (8 sessions), recruited using an online recruitment system. Upon arrival at the lab, subjects were randomly assigned to isolated computer terminals.[6] Three assistants handed out instructions 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 10 rounds, with perfect stranger matching. At the end of each session, one of the rounds was randomly chosen for payment. All subjects received a fixed show-up fee of 2.50 tokens, where 1 token = 0.5 euro.

Each round implemented the following sequence of six stages.

1. **Role assignment.** Player positions B and A are randomly assigned, and pairs formed.
2. **Communication.** B can send a free form message to A (≤90 characters).
3. **A's action.** A reads B's message, and then A has to choose *In* or *Out*.

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

4. **Switching.** Some As were switched with the 50% probability. Only Bs were informed whether or not a switch occurred. Bs with switched As were allowed to read the message that had previously been received by the new A's pre-switch B.[7]
5. **Belief elicitation.** This stage has two parts: a) First-order beliefs: each A was asked to guess if his/her unknown B would choose to *Roll* or *Don't*; b) Second-order beliefs: each B was asked to guess the guess of the A with whom they would play after the switching stage occurred.[8]
6. **B's action.** B chooses between *Roll* or *Don't*. Then all subjects are informed about their payoff in that round. As are neither informed whether they had been switched nor about B's choice; only payoffs are revealed.[9]

Eliciting first- and second-order beliefs is common in the literature on guilt aversion (footnote 3). Doing so here allows us to compare our findings regarding beliefs to C&D's. Incentives for beliefs elicitation 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.[10]

4. Results

We report our findings in three steps. First, we discuss the relation between promises and the proportion of As who choose *In*. Second, we verify whether our design delivers an exogenous variation in promises. Third, we present our main results, concerning CBE and EBE.

3.1 Promises and opt-out decisions

Our sample consists of 1130 pairs of subjects, and 1130 messages from B to A. We asked a research assistant to code these messages according to whether or not they conveyed a promise to *Roll* (or similar-in-spirit clear statement of intent to *Roll*). This way we obtained 527 promises out of 1130 messages (47%).

[7] It is worth noting that only pairs, whose A has decided to continue the game by choosing *In*, could be switched with 50% probability. The other pairs, whose A chose *Out*, finish the game in that round.

[8] Specifically, if B's partner is switched, then B must guess the guess of the new partner after she read the message that the new partner has received by his old partner during communication. Conversely, if the partner of B is not switched, B must guess the guess of the partner with whom she communicated before.

[9] As could obtain a zero payoff either because B chose to *Roll* or because the outcome of the die-roll was #1 when B chose to *Roll*.

[10] Our elicitation procedure is described in detail in Appendix A.

The proportion of As who choose *In* is 61% (691 out of 1130). The proportion of As who choose *In* after receiving a promise is 76% (398 out of 527). The proportion of As who choose *In* when Bs did not promise is 49% (293 out of 603).

As expected, the proportion of As who received a promise and choose *In* is significantly higher than the proportion of As who did not receive a promise and choose *In* (76% vs. 49%: Z=2.52, p=0.012).

3.2 Second-order beliefs and the exogenous variation in promises

Let us now focus on second-order beliefs.[11] As said before, we observe that As choose *In* in 691 cases. Therefore, we have an equal number of Bs who choose between *Roll* and *Don't*. However, due to the partner-switching feature, these Bs are not necessarily those who sent a message to the As choosing *In*. Table 1 reports their second-order beliefs. It is organized as follows. The first two columns (a and b) refer to the non-switched cases, while the second and third columns (c and d) refer to Bs whose partner was switched. Odd columns refer to the case where Bs read a promise, while even ones refer to the case where Bs did not read a promise. Rows indicate whether Bs made a promise (1) or not (2). Note that some cells are empty by design, in the case where Bs partners were not switched (because in this case Bs only read their own messages).

Table 1 – Second-order beliefs of B's

	NO SWITCH		SWITCH	
	PROMISE READ	NO PROMISE READ	PROMISE READ	NO PROMISE READ
	(a)	(b)	(c)	(d)
(1) B makes a PROMISE	**70%** (0.29/204)		**67%** (0.32/104)	**54%** (0.31/90)
(2) B does not make a PROMISE		**55%** (0.31/120)	**59%** (0.31/90)	**57%** (0.30/83)

Looking at Bs with non-switched partners, as in C&D, we find that the second-order beliefs of Bs who made a promise are significantly different from those of Bs who did not sent a promise (70% vs. 55%: Z=2.38, p=0.017). The statistics reported are obtained from the Wilcoxon signed rank test, which compares averages at the session level. Our data are independent at session level, but not at the individual level. The Wilcoxon signed rank tests accounts for such structure in the data.

[11] First-order beliefs display similar patterns as the second-order beliefs and are reported in Appendix B.

Among the subsample of Bs who made a promise, as expected, the average second-order belief of those who read a promise is independent of the switch (70% vs. 67%: $Z=0.00$ and $p=1.000$), i.e., second-order beliefs of Bs with non-switched partners who made a promise are not significantly different from those of other Bs who made a promise and were re-matched with As who received a promise by someone else. Therefore, like Vanberg, we obtain exogenous variation whether a promise was transmitted to the eventual partner at play, among the subjects who made a promise.

The table shows that the correlation found by C&D between promises and second-order beliefs holds independently of the switch. The second-order beliefs of switched promisors who are re-matched with an A who received a promise are higher than those of switched promisors who are re-matched with an A who did not receive a promise by someone else (67% vs. 54%: $Z=2.10$, $p=0.036$). Similarly, the second-order beliefs of Bs with non-switched partners who made a promise are higher than those of Bs who made a promise and are re-matched with an A who did not receive a promise by someone else (70% vs. 54%: $Z=2.52$, $p=0.012$).[12]

3.3 Main results: CBE & EBE

Table 2 reports the average *Roll* rates of Bs. The structure is like that of Table 1. Therefore, we distinguish the average *Roll* rates of Bs who promise and Bs who do not promise by rows. Columns refer to the message they read and indicate if a switch occurred or not.

Table 2 – B's Roll rates

	NO SWITCH		SWITCH	
	PROMISE READ	NO PROMISE READ	PROMISE READ	NO PROMISE READ
	(a)	(b)	(c)	(d)
(1) B makes a PROMISE	**74%** (0.44/204)		**70%** (0.46/104)	**59%** (0.49/90)
(2) B does not make a PROMISE		**29%** (0.46/120)	**31%** (0.47/90)	**39%** (0.49/83)

Focusing on columns (a) and (b) of Tables 1 and 2, we observe a correlation between promise keeping and second-order beliefs, just as in C&D. This is consistent with EBE. Second-

[12] The second-order beliefs of Bs who did not sent a promise (all those in row (2)) are not statistically different. Moreover, those are also not significantly different from the second-order beliefs of Bs with switched partners who sent a promise and were re-matched with As who did not receive a promise (row (1) column (d)). All statistics are reported in Appendix B.

order beliefs of promisors are higher in Table 1 (70% vs. 55%: Z=2.38 and p=0.017) as are average *Roll* rates (74% vs. 29%: Z=2.52, p=0.012) in Table 2. As expected, our results show a correlation between promise keeping and second-order beliefs. People are more likely to keep promises and these are correlated with high second-order beliefs. However, as argued by Vanberg, the correlation does not necessarily imply causation. We need to check it by using our exogenous variation in promises.

Among the subsample of Bs who made a promise, the average *Roll* rate of Bs with non-switched partners is not statistically different from that of Bs who read a promise made by someone else (74% vs. 70%: Z=0.14, p=0.889). Thus, we do not find support for CBE. The behavior of Bs with non-switched partners who are requested to keep their own promises is not different from the behavior of Bs who are requested to keep promise done by another. Our results here is different from that of Vanberg, who found support for CBE when he ran an analogous test.

The result is, however, consistent with the idea that people feel obliged to keep promises made by others since those are associated with higher second-order beliefs, as predicted by EBE. A direct test for EBE is obtained by comparing dictators who read a promise with those who did not. Do they have different second-order beliefs? See Table 1. The average *Roll* rate of Bs with non-switched partners and that of other Bs who read a promise made by someone else are both are higher than that of Bs with switched partners who did not read a promise made by someone else (74% vs. 59%: Z=2.52, p=0.012; 70% vs. 59%: Z=2.10, p=0.036).[13] This finding support EBE. Our results here is different from that of Vanberg, who ran an analogous test but found no statistically significant difference and hence no support for CBE.

5. Interpretation

C&D proposed a theory – EBE – to explain why promises foster trust and cooperation. Vanberg pointed out that C&D's results are confounded: an alternative explanation – CBE – is conceivable. Vanberg introduced an imaginative partner-switching technique, which admits robust causal inference. He developed a design based on which he reported support of CBE over

[13] Although, it is not the scope of the present paper, it is interesting to note that Bs who did not make promises seem to systematically choose *Roll* less than Bs who did. They do not respond to promise written by others, which on turn do not raise their second-order beliefs. These results seem to be related to the study of Ismayilov & Potters (2015). All statistics are reported in Appendix B.

EBE. However, as a follow-up study to C&D's, Vanberg's test is not fully to the point. Apart from the partner-switching feature, he deviated from C&D's design as regards both game and communication protocol. Vanberg thereby (i) left the research question he set out to address insufficiently precisely explored, and (ii) he conducted a study which is valuable for a different reason than he intended. Let us elaborate on (i) and (ii), and then wrap up:

(i) If the goal is to use Vanberg's partner-switching feature to test CBE vs. EBE in C&D's asymmetric setting, then the appropriate design is ours rather than Vanberg's. Our results support EBE over CBE.

(ii) C&D's game is *asymmetric*. At the root, both players know that player A has to trust player B to *Roll*, not the other way around. By contrast, Vanberg's games is *symmetric*. At the root, both players know that either may have to trust the other to *Roll*. Moreover, the communication protocols differ, with a one-sided message from B to A in C&D's case and a conversation-like exchange in Vanberg's case. The symmetry of Vanberg's game, and the back-and-forth nature of his communication protocol, invites the reflection that players may be inclined and able to strike a deal of conditional cooperation: "I'll promise to *Roll* if you promise to *Roll*." And if both players do promise to *Roll*, their exchange has the semblance of an informal agreement. Vanberg's results are consistent with and supportive of the idea that players have a belief-independent preference to honor such agreements. He should be viewed as a pioneer at exploring the impact of informal agreements. One may take inspiration from his findings, and conduct further research theorizing and experimenting about informal agreements.[14]

As regards CBE, there is no tension between our results and Vanberg's. His study, interpreted as we have suggested, explores a preference for honoring informal agreements. This aspect has no counterpart in our (or C&D's) design. A preference for keeping a unilateral promise may be a rather different animal than a preference for honoring a gentleman's agreement. Thus, different forms of CBE are considered by us and by Vanberg.

As regards EBE, it may seem puzzling that this theory is supported in C&D's setting, but not in Vanberg's. Data is what it is though, and we have the following reflection: Different games may trigger different thinking. Humans are motivated in many ways (e.g., by reciprocity,

[14] Very little such research seems to have been done, but see Miettinen (2012) and Dufwenberg *et al.* (2017) for some theory and Kessler & Leider (2011) and Dufwenberg *et al.* (2017) for experiments.

emotions, image concerns, on top of more classical items like income or concern for fair distributions; see Battigalli & Dufwenberg 2020 for a systematic discussion). However, perhaps humans cannot consider more than a few such motivations at a time, and perhaps Vanberg's setting, relative to ours (and C&D's) triggers other motivations that may crowd out the belief-dependent feelings that are built into EBE. To mention two such potential motivations, first consider the preference for honoring an informal agreement, described above. Second, consider reciprocity, such that a player would wish to choose *Don't* [*Roll*] if and only if he or she believed that the other player would have done likewise had he or she been the dictator. In Vanberg's game, this motivation would be potent (since there is a node where the other player may choose *Roll*), whereas in C&D's game it would be muted (since the other player has no *Roll* choice).

Appendix A – Elicitation of beliefs

Elicitation of first-order beliefs. After communication, As were asked to guess whether their (unknown) Bs would choose *Roll* or *Don't Roll*. As could make their guess by ticking one of the five-point scale in Table A. This scale is the same as Vanberg's. Beliefs are then re-scaled to 1, 0.75, 0.5, 0.25, and 0. Thus the numbers shown in Table 2 represent the averages of As' 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).[15]

Table A – Incentives for the elicitation of first-order beliefs

B will	choose *Roll*			choose *Don't Roll*	
	Certainly	Probably	Unsure	Probably	Certainly
Please tick your guess	◯	◯	◯	◯	◯
Your earnings if B					
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. Soon after Bs were told whether their paired subject had been switched or not, they were asked to guess the partner's guess. Specifically, they had to

[15] We also verify the robustness or our results to the quadratic scoring rule. Results are available upon request.

guess which of the five points of Table A had been ticked by their counterpart. Correct guesses were paid 0.50 tokens.

Appendix B – First-order beliefs

The table below reports first-order beliefs.

Table B – First-order beliefs of A's

	NO SWITCH	SWITCH
	(a)	(b)
(1) A receives a PROMISE	**63%** (0.31/204)	**63%** (0.31/194)
(2) A does not receive a PROMISE	**55%** (0.28/120)	**54%** (0.31/173)

As expected from the information structure implied by the experimental design, the first-order beliefs of As who receive a promise are the same between those non-switched and switched (63% vs. 63%: Z=0.00, p=1.000). Similarly, they are the same when those who did not receive a promise (55% vs. 54%: Z=0.28, p=0.779). Comparing now the first-order beliefs of As who receive a promise to those of As who did not in no-switched condition, we find a positive difference in favor of the former (63% vs. 55%: z=2.10 p=0.036). First-order beliefs of As who receive a promise to those of As who did not in switched condition, we find a positive difference in favor of the former (63% vs. 54%: z=2.52 p=0.012).

Appendix C – Test statistics

Second-order beliefs in Table 1. Among Bs who did not make a promise, second-order beliefs of no-switched Bs are not significantly different from those of switched Bs who are re-matched with other As who did not receive a promise by someone else (59% vs. 55%: Z=0.912, p=0.362. Second-order beliefs of Bs who did not make a promise and are re-matched with other A who did not receive a promise by someone else are not significantly different from those of Bs who did not make a promise and whose partners were re-matched with other As who received a promise by someone else (57% vs. 59%: Z=0.42, p=0.674. Second-order beliefs of Bs who did not make a promise and were re-matched with other As who did not receive a promise by someone else are not significantly different from those of trustees who made a promise and are re-matched with other As who did not receive a promise by someone else (57% vs. 55%: *Z*=0.98, *p*=0.327).

Roll rates in Table 2. Rows (1) shows that Bs who did not made a promise do not care about the others when they received a promise, in fact Bs' behavior when their partner is switched to someone who received a promise by another person (column (c)) is 31% which is not significantly different from the 29% when Bs whose partner was not switched did not made a promise (i.e., 29% vs. 31%: *Z*=0.912, *p*=0.362). Moreover, re-matched Bs who did not make a promise and did not read a promise sent by others is 39% (column (b)) which is not significantly different from the 29% when Bs whose partner was not switched and who did not make a promise (i.e., 29% vs. 39%: *Z*=1.614, *p*=0.106). The *Roll* rate of the re-matched B who did not make a promise and read a promise is 31% which is not significantly different from the 39% observed from re-matched Bs who do not make a promise and did not read any promise (i.e., 31% vs. 39%: *Z*=1.614, *p*=0.106).

References

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

Battigalli, P. and M. Dufwenberg (2009), "Dynamic psychological games," *Journal of Economic Theory*, 144: 1-35.

Battigalli, P. and M. Dufwenberg (2020), "Belief-dependent motivations and psychological game theory," Bocconi University IGIER w.p. 646.

Cartwright, E. (2019), "A survey of belief-based guilt aversion in trust and dictator games," *Journal of Economic Behavior and Organization*, 167: 430-444.

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

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

Charness, G. and M. Dufwenberg (2011), "Participation," *American Economic Review*, 101: 1213-1239.

Di Bartolomeo, G., M. Dufwenberg, and S. Papa (2019a), "The sound of silence: A license to be selfish," *Economics Letters*, 182: 68-70.

Di Bartolomeo, G., M. Dufwenberg, S. Papa, and F. Passarelli (2019b), "Promises, expectations & causation," *Games & Economic Behavior*, 113: 137-146.

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

Dufwenberg, M., M. Servátka, and R. Vadovič (2017), "Honesty and informal agreements," *Games & Economic Behavior*, 102: 269-285.

Ederer, F. and F. Schneider (2020), "Trust and promises over time," *The American Economic Journal: Microeconomics*, forthcoming.

Ederer, F. and A. Stremitzer (2017), "Promises and expectations," *Games & Economic Behavior*, 106: 161-178.

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

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

Geanakoplos, J., Pearce D., and E. Stacchetti (1989), "Psychological games and sequential rationality," *Games & Economic Behavior*, 1: 60-79.

Ismayilov, H. and J. Potters (2016), "Why do promises affect trustworthiness, or do they?" *Experimental Economics*, 19: 382-393.

Ismayilov, H. and J. Potters (2017), "Elicited vs. voluntary promises," *Journal of Economic Psychology*, 62: 295-312.

Kessler, J., and Leider, S. (2012), "Norms & contracting," *Management Science*, 58, 62–77.

Khalmetski, K. (2016), "Testing guilt aversion with exogenous shift in beliefs," *Games and Economic Behavior*, 97: 110-119.

Krupka, E.L., S. Leider, and M. Jiang (2017), "A meeting of the minds: Contracts and social norms," *Management Science*, 63: 1708-1729.

Miettinen, T. (2013), "Promises and conventions – an approach to pre-play agreements," *Games and Economic Behavior,* 80: 68-84.

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.

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