DEPARTMENT OF ECONOMICS

Promises and Partner-Switch

Giovanni Di Bartolomeo, Martin Dufwenberg & Stefano Papa

UNIVERSITY OF ANTWERP
Faculty of Business and Economics
City Campus
Prinsstraat 13, B.226
B-2000 Antwerp
Tel. +32 (0)3 265 40 32
www.uantwerpen.be

FACULTY OF BUSINESS AND ECONOMICS

DEPARTMENT OF ECONOMICS

Promises and Partner-Switch

Giovanni Di Bartolomeo, Martin Dufwenberg & Stefano Papa

RESEARCH PAPER 2021-005
DECEMBER 2021

University of Antwerp, City Campus, Prinsstraat 13, B-2000 Antwerp, Belgium
Research Administration – room B.226
phone: (32) 3 265 40 32
e-mail: joeri.nys@uantwerpen.be

The research papers from the Faculty of Business and Economics are also available at www.repec.org (Research Papers in Economics - RePEc)

D/2021/1169/005

Promises and Partner-Switch*

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

December 2021

Abstract: Building on a 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.

---

* We are grateful to Francesco Bloise for comments on a previous version.

[a] Department of Economics and Law, Sapienza University of Rome; Department of Economics, University of Antwerp; CIMEO; giovanni.dibartolomeo@uniroma1.it.

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

[c] Department of Economics and Finance, University of Rome Tor Vergata; CIMEO; stefano.papa@uniroma2.it.

1. Introduction

Promises may foster trust and cooperation. A recent literature explores why. Charness & Dufwenberg (2006) (C&D) propose, and report experimental support for, 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): People like to keep their word.[2] To experimentally test CBE 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. [3]

While the title of Vanberg's paper includes the question "Why do people keep their promises?," his approach to CBE is broader as he also refers to obligations "based on agreements or contracts" (p. 1467). His experiment reflects this too. Let us highlight two differences between C&D's and Vanberg's designs: First, 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, C&D and Vanberg explore different communication protocols. C&D study a single pre-play message that cannot be responded to. Vanberg instead allows subjects to send messages back-and-forth. If they then reciprocate each other's promises, their exchange may have the flavor of a conversation that generates an informal agreement.

A summary of Vanberg's contributions reveals some interesting remaining unchartered research territory: First, Vanberg identified a potential confound to C&D's result, namely CBE. Second, he also developed a design-tool – partner-switching – that allows testing for CBE. Third, he found support for CBE in a relevant context (with messages back-and-forth). However, he did not run a test of the relevance of CBE in C&D's context (with unilateral messages). While he identified a potential confound to C&D's result, he did not test the relevance of that confound in C&D's setting.

[1] EBE 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, 2021).

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

[3] Di Bartolomeo *et al*. (2019b) report similar results from a related design; we elaborate nuances in section 2.

Since C&D' study has garnered much interest, and given our hunch that the difference between messages back-and-forth and unilateral messages may be psychological relevant, we propose that running such a test is of interest. In this paper we report results from a design that accomplishes this. Using Vanberg's partner-switching technology, we evaluate CBE in C&D's trust game setting with unilateral promises.

We also report results regarding EBE. Vanberg's switching feature allows a relevant test, although there are also limits to which extent this can be done (as noted by Ederer & Stremitzer 2017 and Di Bartolomeo *et al.* 2019). We postpone a discussion of details until section 2.

Section 2 provides in-depth scientific background: hypotheses, designs (C&D's, Vanberg's, ours, and also that of Di Bartolomeo *et al.*'s 2019b which helps add perspective), and other related literature. Section 3 describes our procedures. Section 4 explains what we found. Section 5 offers an ex-post 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.

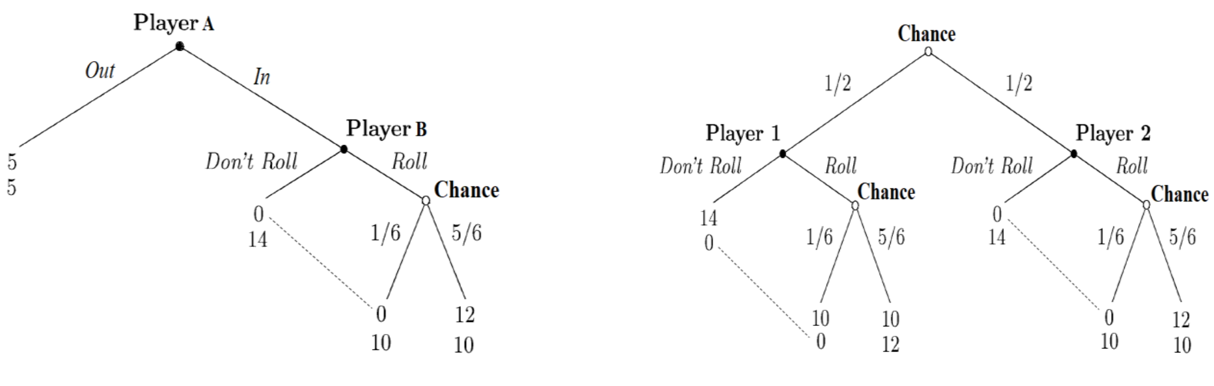

Figure 1 – The game trees: C&D's (to the left) and Vanberg's (to the right)

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*.[4] 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.

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 the empirical relevance of CBE, enabling him to draw robust causal inference regarding the impact of a promise. Namely, he proposed that if subjects *i* and *j* formed a chatting pair and then *i* was chosen to be the dictator, then with 50% probability, *j* would be "switched" and replaced by another subject *k* who previously chatted with 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 were no switch.[5]

Vanberg's results. Vanberg did not base his test of CBE on C&D's game, but rather on the game to the right in Figure 1. That is, if a subject *i* was selected to be the dictator (by the initial chance move), then the subject *j* 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-

---

[4] 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) and Rimbaud (2021) for surveys.

[5] A methodologically attractive feature of Vanberg's switching methodology is that randomization is really at the individual level, not at the session level (which is otherwise common in many experiments, although results are at times unjustifiably interpreted as if the randomization was at the individual level). We thank a referee encouraging us to emphasize this.

shot messages from one player to the other, Vanberg allowed the two players to engage in four rounds of back-and-forth messaging. Based on this design, Vanberg reported support for CBE.

Vanberg furthermore elicited subjects' (first- and second-order) beliefs that a dictator would *Roll*, and he documented that these beliefs increase when subjects that received a promise are involved. Given this pattern of data, Vanberg's design admits the following clean test of EBE: Consider *i, j, k, l* as described in the previous paragraph, and suppose that *i* made a promise to *Roll* to *j*. EBE implies that *i* should be more likely to *Roll* if *l* made a promise to *Roll* to *k* than if *l* did not make such a promise. Vanberg did not find support for this prediction.

Unchartered territory. It is natural to wonder whether Vanberg's key results – support for CBE, and not support for EBE – would obtain also in C&D's setting. The back-and-forth nature of Vanberg's communication protocol may generate experiences that look like informal agreements and so have a rather different flavor than one-sided promises. While informal agreements as well as one-sided promises both evoke issues of keeping-one's-word, the modes of exchange are conceptually distinct. They may relate differently as to how and why trust and cooperation may be induced.

Our design. We apply Vanberg's partner-switching feature to C&D's original game, thus providing new independent tests of CBE and EBE in C&D's setting. 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.

Di Bartolomeo *et al*. (2019b) (DDPP). To clarify our contribution further it is useful to compare with our previous study DDPP (the extra P is Francesco Passarelli's). DDPP tweak Vanberg's design so that the switching probability is not 50% but rather (depending on treatment) 25% or 75%. This induces exogenous variation not only in whether a player who sent a promise was then paired with whoever received that promise (as needed for Vanberg's test of CBE) but it also induces exogenous variation in players' beliefs.[6] This new feature allows DDPP to run several new tests of EBE and guilt aversion, as well as of CBE. Their main results align with Vanberg's, supporting CBE but not EBE. Our current design differs in two respects from DDPP's. First, we work with C&D's game rather than Vanberg's. Second, we use Vanberg's switching probability rather than those of DDPP. In other words, we conduct tests a la Vanberg in C&D's game, whereas DDPP modify Vanberg's tests while staying close to his game.

Less closely related literature. So far in this section, we referenced the studies that most closely relate to (and in fact directly motivate) us. All this work is part of a broader experimental literature on communication in various trust games. In order to provide readers with a richer context and backdrop, we now offer the following brief comments:

Ellingsen & Johannesson (2004) are pioneers among studies that considered preferences for promise-keeping in trust games (which they derived from hold-up problems). Much of their focus, unlike ours, is on theories that refer to fair distributions of material rewards.

C&D's trust game is derived in a contract-theoretic setting, emphasizing "hidden action." Charness & Dufwenberg (2011) is a parallel study, which, however, emphasizes "hidden information," and therefore ends up exploring rather different forms of trust games. Charness & Dufwenberg (2010), on the other hand, is concerned with C&D's game, but explore a different type of (pre-fab and "bare") promise.

Ismayilov & Potters (2016) explore two alternative motivations for promise-keeping, which they label "internal consistency" (desired by the party issuing a promise) and "social obligation" (felt because someone received one's promise). Investigating causality, they argue that promises do not induce trustworthiness, and instead they introduce two types and argue that promises are more likely to be sent by one of them. In another paper, Ismayilov & Potters (2017) study the impact of promise elicitation by trustors from trustees on trust and trustworthiness.

---

[6] DDPP explain that "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, then their second-order beliefs should vary by switching probability in the same direction as recipients' first-order beliefs."

Ederer & Stremitzer (2017) propose a novel design which includes an "unreliable random device," creating exogenous variation in players' expectations and allowing clean tests of EBE. Their exercise may be seen as parallel to Vanberg's, except that the primary focus is on EBE rather than EBE. Their paper relates more closely to DDPP (as discussed above) than to our current work, and we refer to DDPP for more discussion.

Nielsen *et al.* (2019) compare the behavior of individuals and two-person teams in trust games with pre-play communication. Di Bartolomeo *et al*. (2019a) compare active and passive non-communication, i.e., situations where agents stay silent even if promises are viable to a counterfactual where communication is not feasible. Ederer & Schneider (2020) investigate the relation between time and trust with and without pre-play communication (mainly promises).

Finally, Kessler & Leider, Dufwenberg *et al.* (2017) and Krupka *et al*. (2017) explore various aspects of informal agreements and (in the case of Krupka *et al*.) social norms.

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.[7] Upon arrival at the lab, subjects were randomly assigned to isolated computer terminals.[8] 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).

[7] Across the 8 sessions, there were 28 subjects in session 1; 22 in 2; 30 in 3; 30 in 4; 26 in 5; 30 in 6; 30 in 7; 30 in 8, for a total of 226 participants.

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

3. **A's action.** A reads B's message, and then A has to choose *In* or *Out*.
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.[9]
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.[10]
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.[11]

Eliciting first- and second-order beliefs is common in the experimental literature on guilt aversion (see footnote 4). 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.[12]

---

[9] Only in pairs where A chose *In*, switches were possible (with 50% probability). In other pairs, where A chose *Out*, the game finished in that round. Note that when pairs choosing *In* were odd-numbered, half-plus-one of pairs were switched, so the probability of being switched was actually slightly higher than 0.5.

[10] 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.

[11] Participants A 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*.

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

4. Results

We report our main findings, related to CBE and EBE, in Section 4.2. For the relevance of those tests, it is critical to first document that, in our design, promises on balance have the effect of raising subjects' expectations and *Roll* rates. We establish that in Section 4.1.

4.1 Promises, second-order beliefs, and Roll rates

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*).[13] This way we obtained 527 promises out of 1130 messages (47%). 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 choose *In* after receiving a promise is significantly higher than the proportion of As who choose *In* after not receiving a promise (76% vs. 49%: Z=2.52, p=0.012).

Let us now focus on second-order beliefs.[14] As said before, we observe that As choose *In* in 691 cases, so we also have 691 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, the Bs' beliefs about the probability that an A subject believed that B would roll the die. Table 1 also reports standard deviations and number of observations (in brackets). 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 (a and c) refer to the case where Bs read a promise, while even ones (b and d) 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).

[13] Following Vanberg, we asked two assistants to code the messages, ex-ante having decided to use the code of only one of them. Assistants were unaware of our choice. This way, we were able to check the robustness of the codification. The correlation between the two codes of messages is 0.89.

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

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 send a promise (70% vs. 55%: Z=2.38, p=0.017).[15] 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 regarding whether a promise was transmitted to the eventual partner at play, among subjects who made a promise.

Table 1 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).

The second-order beliefs of Bs who did not send a promise (all those in row (2)) are not statistically different.[16] This is surprising since promises do affect beliefs (as we just reported). This

[15] 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.

[16] These beliefs 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)).

may suggest some form of heterogeneity between agents who self-select to send promises and those who do not (reminiscent of differences reported by Ismayilov & Potters 2016). We had not expected this, but the observation is of limited relevance to our main results, to be reported in Section 4.2, which concern effects of promises made rather than promises that are not made.[17]

4.2 Main results: CBE & EBE

Table 2 reports the average *Roll* rates of Bs, standard deviations and number of observations. The structure is otherwise like that of Table 1. 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 whether or not a switch occurred.

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-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

[17] We acknowledge the concerns, pointed out be a referee, that who makes or does not make a promise may depend on the experimental protocol (e.g., C&D's vs. Vanberg's) as well as cultural differences between the subject pools (Vanberg's subjects came from a university in Thuringia, whereas our subjects were recruited in Rome.) If different cultural norms govern behavior, attempts to fully compare results may require a full replication of each design in the same society.

necessarily imply causation. We need to further investigate the issue by using our exogenous variation in promises.

Among 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 result 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 Bs who read a promise with those who did not. Do they have different second-order beliefs? Looking at Table 2, 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 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). This finding supports EBE. Our result here is different from that of Vanberg's who ran an analogous test but found no statistically significant difference and hence no support for EBE.

5. Discussion

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

Vanberg deviated from C&D's design as regards choice of game as well as communication protocol. It is natural to wonder if his main result extends also to a setting that uses C&Ds game and communication protocol. This is what we have explored. Our results are the opposite of Vanberg's, supporting EBE, and not supporting CBE.

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.[18]

As regards CBE, there is no tension between our results and Vanberg's. If his study (as we just suggested) is interpreted as documenting evidence for a preference for honoring informal agreements, then this 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. 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, after observing the results, we have the following reflections: Different games may trigger different thinking. Humans are motivated in many ways (e.g., by reciprocity, 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 one designated to choose whether or not to *Roll*. 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). [19]

[18] Very little research seems to have been done on the impact of informal agreements; but see Miettinen (2013) and Dufwenberg *et al.* (2017) for some theory and Kessler & Leider (2012), Dufwenberg *et al.* (2017), and Krupka *et al.* (2017) for experiments.

[19] Our remarks here are narrowly focused, as reflected by the following remark by a referee: "While the comparison between the setup of [Vanberg] and C&D is an important step in our understanding of why humans want to keep their word, the current study remains incomplete without (a) asking why people make promises in the first place under the different situations and (b) considering cultural differences as a source of the observed discrepancies (c) exploring the ability for empathy and its role in promise keeping." (Compare with footnote 17 above.)

We hope that future research may take inspiration from these speculative remarks and develop new related design that may be useful for testing the empirical relevance of our ideas.

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).[20]

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

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

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%**	**63%**
	(0.31/204)	(0.31/194)
(2) A does not receive a PROMISE	**55%**	**54%**
	(0.28/120)	(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 (2021), "Belief-dependent motivations and psychological game theory," *Journal of Economic Literature*, forthcoming.

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.

Nielsen, K., P. Bhattacharya, J. Kagel, A. Sengupta (2019), "Teams promise but do not deliver," *Games and Economic Behavior*, 117: 420-432.

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.

Rimbaud, C. (2021), "Introduction," in *Three Essays on Guilt Aversion: Theory and Experiments*. Ph.D. Thesis, GATE-LAB, University of Lyon.

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