

Broken Promises and Hidden Partnerships: An Experiment

Jingnan (Cecilia) Chen ^{*} Daniel Houser [†]

November 7, 2018

Abstract

Previous research indicates that cheap-talk promises promote cooperation. We extend the empirical study of promises to a three person environment. Doing this enabled us to study two features of promises in the naturally occurring world that have not been studied in the previous literature. For one, we are able to determine the impact of “contingent” promises, which refers to promises that need to be kept (or broken) only in the event that another event first occurs. Second, we were able to study “promise chains”, by which we mean a set of sequentially-made promises. In both cases, we are able to apply existing theories to these novel environments to predict promise keeping behavior. Consistent with theory predictions, we find that promises are not necessary, and do not change behavior, when incentives are aligned. When there is conflict of interest, and when a promise is definite (that is, not contingent), we find that people are indeed willing to forgo monetary payoff to keep their promises. However, when the promise is contingent, promises no longer promote cooperation: people are equally like to choose a selfish action regardless whether they made a promise. Finally, we find that promise chains do not promote trustworthiness, with people at the end of the chain no more likely to honor their promises than people at earlier positions, despite the additional indirect harm that people later in the chain cause by defecting. Our findings offer insights relevant for the behavioral economic theory of guilt aversion.

Keywords: promise, cheap talk, communication, behavioral game theory

^{*}Corresponding author. Economics Department, Business School, University of Exeter, Exeter, UK, EX4 4PU. Email: j.chen2@exeter.ac.uk

[†]Interdisciplinary Center for Economic Science (ICES) and Department of Economics, George Mason University, Fairfax, VA, USA, 22030. Email: dhouser@gmu.edu

1 Introduction

Promises are ubiquitous feature of human society. They occur in both business and personal relationships, and affect the decisions of those who make and those who receive promises. In some cases, promises are between two people, and over behaviors that affect only those two people. In other cases, promises can occur sequentially, with latter promises over behaviors that can impact people involved in previous promises. Despite the importance of sequential promises in multi-agent environments, to our knowledge, all existing theory and evidence on the impact of promises have been limited to two-person environments. The purpose of this paper is to first provide a richer environment for various forms of promises (e.g., contingent promises, sequences of promises, etc.), and second present empirical evidence from a laboratory experiment on the impact of promises among multiple people who have a shared interest in the behaviors over which the promises are made.

Consider, as an example of what we refer to as a “promise chain”, the case of borrowing money from friends or relatives on a handshake. If one needs a downpayment for a house, then she might ask her well-off uncle for help. Her parents may make a promise to her uncle to assure him that the loan will be repaid. Further, the borrower promises to her parents that indeed she will repay the loan so as not to create a difficult situation for them. In this case, the promise chain is the borrower’s promise to her parents and her parents promise to her uncle. In the event that the borrower breaks her promise and defaults on the loan, not only does her uncle lose the money, but her parents may also face social sanctions from the uncle.

In fact, people do often turn to friends and family over banks when looking for a way to finance a house, buy a car, or even fund a business. According to the 2011 report by the National Association of Realtors, 7% of home buyers received a loan from a relative or friend to finance their home. From the 2011 Pepperdine Private Capital Markets Study, more than 50% of small business owners acquired their funding from friends or family.

Another broad example for promise chain is the group lending widely used in micro-finance, e.g., the Grameen lending. Instead of lending money to a specific individual, the micro lenders lend the money to a group of borrowers. Each borrower gets a small loan from the group and the group as a whole is responsible for the loans. The promise chain here is the individual borrower to the group and the group to the micro lender. If an individual borrower defaults, the micro lender does not get the money back, the group may also suffer as future loans may not be available to all the group members

any more.

Another type of promise that may be of great interest is what we call a “contingent promise”, where the opportunity to keep (or break) the promise only arises if other events (unrelated to the promise maker or receiver) first occur. Consider the case of a project manager who promises her client a satisfactory outcome. However the realization of that outcome depends on the technology provided by an independent firm. The project manager’s promise is not relevant in the case where the technology fails. Should one *ex ante* trust a contingent promise as much as a definite one where the promise is followed definitely by action?

The economics literature on promises has focused on two-person environments, often arguing that cheap talk communication tends to discipline dishonest/selfish behaviors. That is, people honor their cheap talk and refrain from cheating even at significant personal monetary cost (see, for example, Charness and Dufwenberg, 2006; Charness et al., 2012, 2013; Charness and Dufwenberg, 2011; Miettinen and Suetens, 2008; Vanberg, 2008; Ellingsen and Johannesson, 2004; Kerr and Kaufman-Gilliland, 1994; Loomis, 1989; Sally, 1995). The focus on two-person environments, however, means that little is yet known about the impact of promises in environments such as those discussed above. This paper fills that gap, by addressing the following related questions.

1. Does one’s position in a promise chain affect the likelihood that one will keep a promise? Would people care beyond their immediate counterpart? In promise chains, one broken promises may impact multiple people in the chain. As a result, are people at the end of the chain more likely to keep their promises than those at the front of the chain?
2. In the two player setting, a promise is followed promptly by an action. However, in a multi-person environment, a promise may need to be kept only in the event that another person keeps (or breaks) their own promise. Is a contingent promise more likely to be kept?

We aim to address the above questions with a novel three-person trust game (we call it the Mistress game), where promise chains and contingent promises occur. Additionally, we allow players to make promises to another player using natural language communication¹.

Our paper makes three key contributions to the literature. First, we provide, to our knowledge, the first empirical evidence on the effect of promises on trustworthiness

¹We also vary the pairings of sender and receiver.

in multi-player environment. Second, we shed light on the extent to which chains of promises and contingent promises can modulate people’s behaviors. We devise four message treatments to systematically explore different paths of communication, by varying the pairing of sender and receiver and the alignment of their monetary interests. Finally, we shed light on the empirical relevance of existing behavioral theories. In particular, we take our data to two types of competing theories: innate preference for honesty models (see, e.g., Ellingsen and Johannesson, 2004; Miettinen and Suetens, 2008; Vanberg, 2008; Gibson et al., 2013; Hurkens and Kartik, 2009), and the consequence based model (see, e.g., Battigalli and Dufwenberg, 2007; Battigalli et al., 2013; Charness and Dufwenberg, 2006).

Our main findings are that: (i) when extended to a three player environment, contingent promises are ineffective as compared to no communication baseline. We observe that people are unlikely to honor a promise when a contingent occurrence is necessary in order for the promise to be realized². (ii) People at the end of the promise chain are no more likely to keep the promises than those at the start of the chain, despite the indirect harm caused to people earlier in the chain as a result of defection by a person later in the chain. We show that this finding has important implications for theories of promise-keeping based on guilt aversion.

The remainder of our paper is organized as follows: Section 2 describes the structure of the game and the various communication treatments. Section 3 details the predictions. The experiment procedure are discussed in Section 4. Section 5 describes our main results. And the final section summarizes and concludes.

2 The Mistress Game

2.1 Baseline Game

To answer the first question from the introduction, we describe a novel environment - the mistress game³ - where various forms of promises can take place. The extensive

²For the detailed support for this result, please refer to Section 5.3 Result 2.

³The name of the game comes from the broad resemblance between the game structure and some extramarital affairs scenarios. our game can be considered as having the following three players: wife (A), husband (B) and mistress (C). The wife chooses whether to get married. If the husband also agrees (chooses *In*), then both parties are better off (both earn a payoff of \$10, rather than the payoff of being single for \$5). Later on, a new player C the mistress appears, and the husband has to choose whether to stay happily married (choose *Out*) or take on a new relationship with the mistress (choose *In*). If the husband decides to take on the mistress, then the mistress chooses whether to keep the affair secret (choose *Left*) or expose the relationship and take a one-time big payoff (choose *Right*). If the mistress chooses to keep the affair secret, then the wife will not discover and is no worse off

form of the game is shown in Figure 1. Using backward induction and assuming risk-neutral selfish players, there is a unique subgame perfect Nash equilibrium (*In*, *Out*, *Right*). This equilibrium is inefficient. Therefore, our game G_1 shares the dilemma common to previously studied trust game variants⁴.

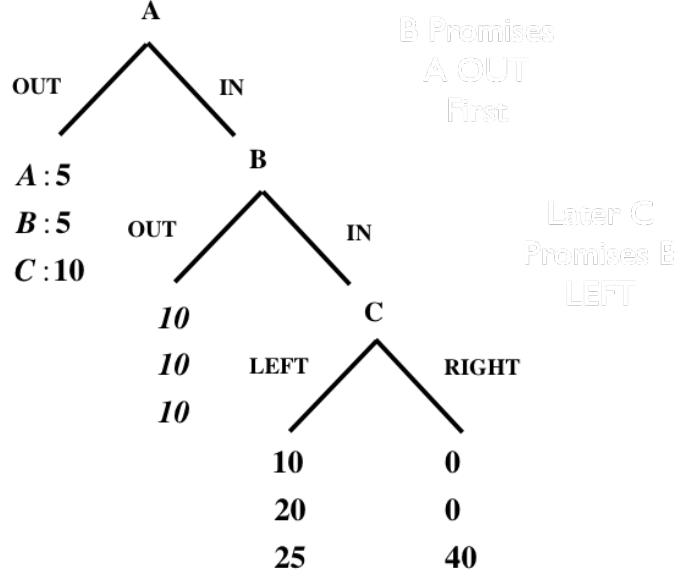


Figure 1: The Mistress Game

A and B consider whether to form a partnership; if no partnership occurs, then both parties receive the outside option payoff of \$5. In this case, C receives \$10. If a partnership is formed, a trust relationship emerges, and the payoffs to this relationship depend on the B's decision. B is faced with a dilemma – to stay with the current partnership (corresponding to B's *Out* option) or form an additional trust relationship with a third person and enjoy a potentially higher payoff (corresponds to B's *In* option). Note that A is NO better off (perhaps even worse off) by B's choosing *In*; thus, A would always prefer B to choose *Out* and maintain an exclusive partnership. If B chooses to stay with A (corresponding to the strategy profile (*In*, *Out*, *Left/Right*), both A and B are better off (with the payoff of \$10 for each), and C (who has no move) again earns the outside option of \$10. The strategy profile (*In*, *Out*, *Left/Right*) corresponds to

(since her payoff is the same as when the husband chooses *Out*), both the husband and mistress enjoy a higher payoff. If the mistress chooses to expose the affair and take the one-time big payoff, then the marriage collapses and both the husband and wife receive nothing.

⁴Such related games are Charness and Dufwenberg (2006) – two-person trust game with a hidden action; Sheremeta and Zhang (2013) & Rietz et al. (2012) – sequential three person trust game; and Cassar and Rigdon (2011) – three person trust game with one trustee two trustor or one trustor two trustee, finally Bigoni et al. (2012) – two person trust game with an add-on dominant solvable game between the trustee and a third player.

the situation where an exclusive partnership contract is enforceable. However, such a contract may not be enforceable. Indeed, B's choice may not be observable to A, depending on the Cs decision. Our game captures this as discussed below.

For this case, if B chooses to form a new partnership with C (corresponding to B's *In* option), C can either be cooperative and reciprocal by choosing *Left*, or defect by choosing *Right*. Note that if C chooses *Left*, B's behavior is unknown to A (B's original partner). However, if C chooses *Right*, not only does B receive nothing from the newly initiated partnership (C takes all), A is also impacted and receives nothing. In this case, A knows B's choice. Note that A may foresee such outcomes and choose not to enter the partnership with B. The players' choices *Out*, *In* and *Right* describe those possibilities.

There are several points to note about our game. First, it resembles sequential trust game at first glance. Closer look reveals that payoffs in our G_1 differ from sequential trust games with constant multipliers across trustees (e.g., Sheremeta and Zhang (2013) or Rietz et al. (2012)). The multipliers in G_1 double with the second trustee and this makes it much more profitable to establish the second partnership. Second, its structure is closely related to that described by Charness and Dufwenberg (2006). The key difference is that we add a strategic player in place of chance, yet maintain unobservable actions, i.e., B could choose *In* without A's discovery, so long as C chooses *Left*. However, unlike Charness and Dufwenberg (2006), where defection necessarily reduces the trustees payoff (B chooses Don't Roll, A's payoff is drastically reduced), in G_1 defection (for example, B chooses *In*) may have no payoff consequences to the trustor (A's payoff can stay the same even if B defects and chooses *In*). This set-up is therefore necessary to capture the possibility of a hidden partnership.

2.2 Communication

We next focus on treatments that differ by whether a pre-play communication opportunity is available and how such opportunity is presented. Among all the communication treatments, one player transmits a message to the other player(s) before they play game G_1 . If we maintain the assumption that players are selfish, then all the pre-game cheap talk communication should have no effect, and the strategy profile (*In*, *Out*, *Right*) remains the unique sub-game perfect solution. However, if there are other concerns that incentivize players, as detailed in the next section, communication will have an impact on behaviors.

We consider the following communication treatments denoted as B-A, C-B, C-A,

and Double treatment, respectively.

B-A Treatment

Prior to the game play, player B can transmit a message to A. Notice that *Out* is already the dominant strategy for B and A also prefers B to choose *Out*. This treatment captures the situation where incentives are aligned between the promise sender and receiver.

C-B Treatment

In a similar fashion, player C can transmit a message to B prior to the game in the C-B treatments. In this treatment, there is a conflict of interest, where *Right* is the option that maximizes C's payoff but minimizes B's payoff. In addition, the promise from C to B is a *definite* promise where the action follows the communication directly between the promise sender and receiver.

C-A Treatment

Similar to C-B treatment, player C can transmit a message to A in the C-A treatment. There is also a conflict of interest between promise sender C and receiver A. Additionally, the promise from C is a *contingent* promise that relies on B's decision.

Double Treatment

In the Double treatment, we introduce a promise chain. It is common information from the start that role B can send a message to A, and the experimenters collect those messages and pass them on to their matched partners. At this point, for player B, the Double treatment is exactly the same as the B-A treatment. However, after all messages are received, we announce a surprise communication opportunity where role C has the chance to transmit a message to B, and after messages from Cs are transmitted, players play the game. Note that the Double treatment is designed such that we can compare B's behavior in the B-A and Double treatments, and Cs behavior in the C-B and Double treatments holding the players communication opportunity and the contents of the communication constant.

3 Predictions

Standard economic models of self-interest utility maximization emphasize the role of outcome in dictating agents choices. As a result, these models predict no effect of

cheap-talk communications on behaviors, since cheap-talk is not an enforceable commitment and is therefore impossible to verify. One of the important assumptions in these standard models is that self-interested agents will have no problem lying or defaulting on their words as long as the resulting outcome is preferred. However, honesty and promise-keeping behaviors are frequently reported not only on the news (*e.g.*, the whistleblowers) but also as observations in lab and field experiments. To account for these seemingly puzzling behaviors, researchers have come up with two main types of models: intrinsic preference for honesty model and consequence-based model.

The intrinsic preference for honesty model assumes that different people incur some level of costs while caught lying (*e.g.*, Lundquist et al. (2009); Ellingsen and Johannesson (2004); Miettinen and Suetens (2008); Vanberg (2008); Gibson et al. (2013)). We define player i as those who make a promise. For those who choose not to communicate or send non-verifiable communication, their decision problems are modeled with standard self-interest maximization. And $(In, Out, Right)$ remains the unique backward-induction solution. Formally, player i has the following simplified utility function:

$$u_i = \begin{cases} m_i - l_i & \text{if player } i \text{ lies} \\ m_i & \text{otherwise.} \end{cases} \quad (1)$$

where m_i denotes agent i 's immediate monetary payoff, l_i indicates the utility loss player i endures when she breaks the promise or lies and $l_i \in [15, +\infty]$ ⁵. Note that player i incurs the cost of lying as soon as she breaks the promise, regardless of the consequences to the promise receiver. It could be the case that the promise receiver may actually gain from the broken promise. However, the cost of lying would still refrain i from breaking the promise.

On the other hand, the consequence-based model assumes that the cost of lying comes from the guilt of letting someone down, and the level of guilt a player suffers depends on the level of harm he imposes on others relative to what the others believe they will suffer, i.e., the difference between the players actual action and the action the player believes others believe he would take (*e.g.*, Geanakoplos et al. (1989); Dufwenberg and Kirchsteiger (2004); Charness and Dufwenberg (2006); Battigalli and

⁵For simplicity, we assume the cost to lying are large enough to make a behavioral difference. According to Ellingsen and Johannesson (2004), the estimated l is around 21.43 SEK \approx USD 2.64. If we apply this estimates to our game, $l = 2.64 < 15$, we should expect no effects at all across treatments.

Dufwenberg (2007)). Player i has the following modified utility:

$$u_i = \begin{cases} m_i - \gamma_i \cdot \tau_i^j \cdot \Delta m_j, & i \neq j \quad \text{if player } i \text{ lies and deviates from the promise} \\ m_i & \text{otherwise.} \end{cases} \quad (2)$$

where γ_i denotes player i 's sensitivity to guilt, and it is independent from τ_i ; $\gamma_i \in [0, +\infty)$; τ_i^j denotes player i 's belief about j 's belief about i , $\tau_i^j = E(\mu_j)$, where μ_j is the probability player j assigns to i 's move, $\mu_j \in [0, 1]$; Δm_j denotes j 's monetary loss between what i thinks j thinks that i would do and what i actually does.

In case of chained promises (see Figure 2 below), the utility function (2) can be modified to reflect further indirect guilt player i may incur. Imagine the following scenario: player j first makes a promise to player k , and later on j receives another promise from player i . Notice that player i is at the end of the promise chain. If player i breaks her promise to j , it will affect j directly but also impact k indirectly (k may receive different payoff from the amount that is promised by j , due to i 's broken promise). As a result, player i has the following modified utility for chained promises:

$$u_i = m_i - \gamma_i \cdot (\tau_i^j \cdot \Delta m_j + \tau_i^k \cdot \Delta m_k), \quad i \neq j \neq k \quad (3)$$

The term $\tau_i^j \cdot \Delta m_j$ represents the direct guilt that player i experiences when letting down player j (deviating from the promised action). The term $\tau_i^k \cdot \Delta m_k$ describes the indirect guilt player i receives when letting down player k indirectly. τ_i^k denotes player i 's belief about k 's belief about j 's action upon receiving the promise from j . And the term Δm_k denotes the payoff difference between the amount promised by j and the actual amount received after i 's decision.

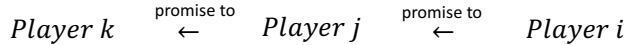


Figure 2: An Example of Chains of Promises

Prediction 1 *When there is no conflict of interest between message sender and receiver, communication/promises will have little effect on players' behavior. Player B will choose Out as frequently in the B-A treatment as in the Baseline (no communication). Player C will choose Right as frequently in the B-A treatment as in the Baseline (no communication).*

In the Baseline and B-A treatments, it is always the best response for B to choose *Out*, anticipating C to prefer *Right*. As a result, promises will have limited effect on Bs.

Prediction 2 *When the incentives are misaligned between message sender and receiver, communication and promises will promote the cooperative outcome. Compared with Baseline (no communication), player B will choose Out (non-cooperative option) less frequently in C-A, C-B and Double treatments. Furthermore, compared with Baseline (no communication), player C will choose Right (non-cooperative option) less frequently in C-A, C-B and Double treatments.*

In the Double treatment, choosing *In* can become the best response for some message-receiving Bs (at least for those Bs who do not send but receive messages from Cs) when there is increased probability of C choosing *Left*. Similarly, in the C-B or C-A treatment, choosing *In* can be the best response for ALL message-receiving Bs. As a result, we expect fewer Bs to choose *Out* in both Double, C-B and C-A treatment than the Baseline.

It is worth emphasizing that actions do not follow promises immediately in C-A treatment. Player C makes a promise at the beginning of the game and does not take an action until player B makes a move. However, the theories above predict that player Cs will choose *Left* more often in C-A treatment than baseline (no communication), i.e., promises work even when actions do not follow immediately after. Anticipating the promise-sending Cs to choose *Left*, more Bs will choose *In* in C-A treatment than baseline. As a result, we predict B will choose *Out* more frequently in the Baseline than C-A treatment.

Prediction 1 and 2 above are made based on comparisons between treatments with communication and the baseline (no communication). In order to make comparisons among treatments with communication, we need to make one (testable) assumption:

Assumption *The frequency of promises from each player type is constant across treatments.*

This assumption implies that: 1) the percentage of Bs sending promises in the B-A treatment is comparable to that in the Double treatment; 2) the percentage of Cs sending promises in the C-B treatment is the same as that in the Double treatment; and 3) the percentage of Bs receiving promises in the C-B treatment is the same as that in the Double treatment.

Prediction 3 and 4 below discuss comparisons between treatments with communication. In particular, Prediction 3 focuses on the comparison between C-A and C-B treatment, while Prediction 4 emphasizes the comparison between C-B and Double treatment. Since different theories have various predictions with regard to treatment comparisons, we use a, b or c to differentiate predictions under different models.

Prediction 3a (*Intrinsic preference model*) *Both player B and C will behave similarly in C-A and C-B treatment.*

Following the intrinsic preference model, promises work the same way for Cs in both C-A and C-B treatment. As a result, we would anticipate both player C and B to behave similarly.

Prediction 3b (*Consequence based model*) *Player B will behave similarly in C-A and C-B treatment. Player C will choose Right (non-cooperative option) more frequently in C-A than C-B treatment.*

Following the consequence based model, the cost of lying occurs when unexpected harm is done to the person who receives the promise. The unexpected harm is greater for player B than player A (A may lose 10 while B loses 20). we would predict more Cs choose *Right* in C-A than C-B treatment.

Prediction 4a (*Intrinsic preference model*) *Player B will choose Out (non-cooperative option) more frequently in Double treatment where chains of promises emerge than in C-B treatment. Player C will choose Right (non-cooperative option) in C-B treatment as frequent as Double treatment.*

According to intrinsic preference model, once a promise is sent, it will be kept. In C-B treatment, only those Bs who receive a promise will choose *In*,⁶ anticipating their matched Cs to keep their promises. In Double treatment, only those Bs who receive a promise but do not send one will choose *In*. Since we assume that the rate of promises exchanges is constant across different roles, we would predict more Bs to choose *Out* in chained promises. Player C will behave similarly in both Double treatment with chained promises and C-B single promise treatment.

Prediction 4b (*Consequence based model - directly guilt only*) *Player B will choose Out (non-cooperative option) as frequently in C-B treatment as Double treatment. Player C will choose Right (non-cooperative option) as frequently in C-B treatment as the Double*

⁶This prediction is made under the assumption that all players in the game are rational, self-interested and utility maximizing. Without a promise, the unique subgame perfect Nash equilibrium (In, Out, Right) will be played.

treatment where chains of promises appear.

According to consequence based model with direct guilt only, one only incurs the cost of lying (direct guilt) from the most immediate counterpart, when he/she causes unanticipated harm to the counterpart with whom the promise is made. In Double treatment, *In* is still the best response for Bs who receive promises but do not send promises to As. Further, *In* also becomes the best response for Bs who receive promises and send promises to As. As long as their paired Cs keep their promises, player As' promised payoffs of \$10 will remain even if Bs choose *In*. As a result, those Bs who receive promises from Cs and send promises to As no longer suffer from the cost of lying since there are no real consequences (to A). Similarly, in C-B treatment, all Bs who receive promises from Cs will also choose *In*. Since we assume that the rate of promises from Cs are the same in Double and C-B treatment, we predict Bs to choose *Out* as frequently in C-B treatment as Double treatment.

Since player C only suffers from direct guilt from breaking promises to the immediate counterpart B, C-B and Double treatment are identical for player C. Therefore, we predict that the *Out* rate for Cs would be the same in both C-B and Double treatments.

Prediction 4c (*Consequence based model - direct and indirectly guilt*) *Player B will choose Out (non-cooperative option) as frequently in C-B treatment as Double treatment. Player C will choose Right (non-cooperative option) more frequently in C-B treatment than Double treatment where chains of promises appear.*

According to consequence based model with direct and indirect guilt, one in chained promises incurs both direct and indirect cost of lying when breaking a promise. In Double treatment, player Bs are at the start of the chain and only suffers from direct guilt. As a result, the prediction for Bs is the same as in Prediction 4b: player B will choose *Out* as frequent in C-B treatment as Double treatment.

For player C, those Cs in Double treatment will suffer direct guilt from letting B down, and they also suffer from indirect guilt from letting player A down. The increased cost of breaking promises makes *Right* a less attractive option in Double treatment than C-B treatment. As a result, we anticipate C to choose *Right* more often in C-B than Double treatment.

4 Experimental Procedure

The experimental sessions were conducted in the experimental laboratory of the Interdisciplinary Center for Economic Science (ICES) at George Mason University. All subjects were recruited using ICES recruiting software from the undergraduate students pool. We conducted 17 sessions with 9-12 participants per session (total of 273 subjects).⁷ All experimental sessions were run with pen and paper. Participants could only participate in one session. Average earnings were \$17 (including \$5 show-up fee); sessions lasted approximately 90 minutes.

We used within subjects design in our experiment. In each session, subjects played three games (treatments) in random order. Subjects were fully aware that for each game they were matched with complete strangers with whom they had never previously interacted before. Only one of the games was randomly selected at the end of the experiment to be a paid game. The session and corresponding game played are shown in Table 1⁸. Note that the Double treatment was never run together with either the B-A or C-B treatments; the reason is that the Double treatment effectively combines the B-A and C-B treatment. If we had run the Double treatment together with the B-A and/or or C-B treatments, participants would have had to write messages to the same role twice, which could have potentially confounded treatment differences (such as different contents in the message).

Table 1: Session – Game information

| Session No. | Game Played (in random order for each session) |
|-------------|--|
| 1-6 | Baseline, B-A treatment, C-B treatment |
| 7-13 | Baseline, Double treatment, C-A treatment |
| 14-17 | Baseline, Double treatment, A-C treatment |

Upon arrival, participants were asked to privately draw from a stack of cards with a letter (“A”, “B” or “C”) and a number written on them. The letter indicated the participant’s role in the session. Participants’ roles stayed constant throughout the experiment. During each session, participants were referred to as A or B or C (as in the games in section 2). Thus participants roles in the experiment were randomly determined at the beginning of the experiment. Then subjects were seated at their

⁷All the sessions were run in two weeks in 2012.

⁸Notice that in Table 1, A-C treatment isn’t discussed in this paper since it is less relevant for the purpose of this paper. Although around 80% of As sent a message to C, the players behavior in A-C treatment are quite similar to that in baseline.

private stations marked with the corresponding number of the card they draw. A quiz was administered before the actual games were played to ensure subjects' understanding of the tasks in the experiment. We adopted the strategy method for all players in conducting our experiment, as did Charness and Dufwenberg (2006)⁹. In the Baseline treatment, no messages were allowed. In all other treatments with communication, each potential message-sender had the option to send a nonbinding messages to their matched partner prior to their partners' decision; they were given a blank sheet of paper on which a message can be hand written, but could decline to send a message by circling the letter (A, B or C) that indicated their role in the experiment at the top of the otherwise-blank sheet. Then messages were transmitted to the respective partners. Upon completion of the message transmission, participants played the game; B made his/ her choice of *In* or *Out* without knowing A's actual choice of *In* or *Out* (similarly, C made his/her decision without know the actual decision of B), but the instruction explained that B's choice would be immaterial if A chose *Out*. We therefore obtained an observation for every B and C.

At the end of each game, we also elicited players' beliefs. After we collected the strategic decisions, we gave out decision sheets that invited subjects to make guesses about the choices of other players in the game and offered to reward good guesses. We elicit from the subjects both the first order beliefs, e.g., player As were asked to guess the proportion of Bs who chose *In*, as well as the second order belief for the players who can send messages. For instance, in C-B treatment, player Bs were asked to guess 1) the proportion of As who chose *In* (first order belief) and 2) the average guess made by As who chose *In* about the proportion of Bs who chose *In* (second order belief). If a guess was within five percentage points of the realization, we rewarded the guesser with \$2.

5 Results

5.1 Overview of The Messages

What messages were sent? Free-form messages can be classified in various ways¹⁰. To simplify the analysis, we assume that a player can make a statement regarding

⁹This is an effort to make the results more comparable for theory testing purposes. Also Amdur and Schmick (2012) suggest that there is no behavioral difference between the use of strategy method and direct response for our type of game with communication.

¹⁰For more detailed analyses on the contents of those messages, please refer to Chen and Houser (2016).

his/her planned action or stay silent. For instance, Player B can make a statement regarding his or her planned action (*In* or *Out*) and Player C can make a statement (*Left* or *Right*). Staying silent indicates two things: no messages are transmitted or the message shows no indication of the player’s planned action¹¹. Here, we denote a statement of planned action as Promises(P), and staying silent as No Message (NM). From the messages that we collected, Promises from B always involve a statement indicating the action *Out*, while all Promises from C involve a statement indicating the action *Left*.

Table 2 below summarizes the rate of messages and the frequency of promises for each of the treatments. The difference between the two indicates the percentage of messages that belongs to NM. The data generally support our assumption that both the rate of communication and the frequency of promises are constant for the same role as long as there is no contingency that must occur before the promise is realized.

As shown in Table 2, Player B sends messages about 88 percent of the time in both the B-A and Double treatments ($z = 0.11, p = 0.92$). Player C sends messages about 80 percent of the time in both the C-B and Double treatments ($z = 0.83, p = 0.40$); however, in the C-A treatment, player Cs are significantly less likely to send messages ($z = 2.60, p = 0.01$); less than half of them sent messages to As.

Table 2: Communication Summary by Treatment

| | % send a Message | | % send a Promise | |
|-----------|------------------|----------------|------------------|----------------|
| Treatment | Player B | Player C | Player B | Player C |
| B-A | 21/24 (88%) | | 13/24 (54%) | |
| C-B | | 20/24 (83%) | | 16/24 (67%) |
| C-A | | 13/27 (48%) | | 7/27 (26%) |
| Double | 38/43 (88%) | 32/43 (74%) | 22/43 (51%) | 29/43 (67%) |

Player B sends Promises about 53 percent of time in both the B-A and Double treatments ($z = 0.23, p = 0.81$). Player C promises around 67 percent of the time in both the C-B and Double treatments ($z = 0.06, p = 0.95$). In the C-A treatment, consistent with the pattern for messages, Cs are significantly less likely to send messages

¹¹We combined no messages and no planned action into one category - stay silent, as there were very few messages without planned action.

to As compared with C-B and Double ($z = 2.89, p = 0.00; z = 3.36, p = 0.00$, respectively); around a quarter of them sent promises to As. However, we can reject the null hypotheses ($p = 0.00$) that there are no promises exchanged in the C-A treatment as compared to Baseline.

5.2 Overview of the Observed Behaviors

Figure 3 summarizes the choices of players A, B and C for the various treatments. In the Baseline, where there is no communication opportunity, 81% of As chose *In*, 76% of Bs chose *Out*, and 73% of Cs chose option *Right*. The behaviors we observed are well described by the unique backward induction Nash equilibrium strategy profile (*In*, *Out*, *Right*).

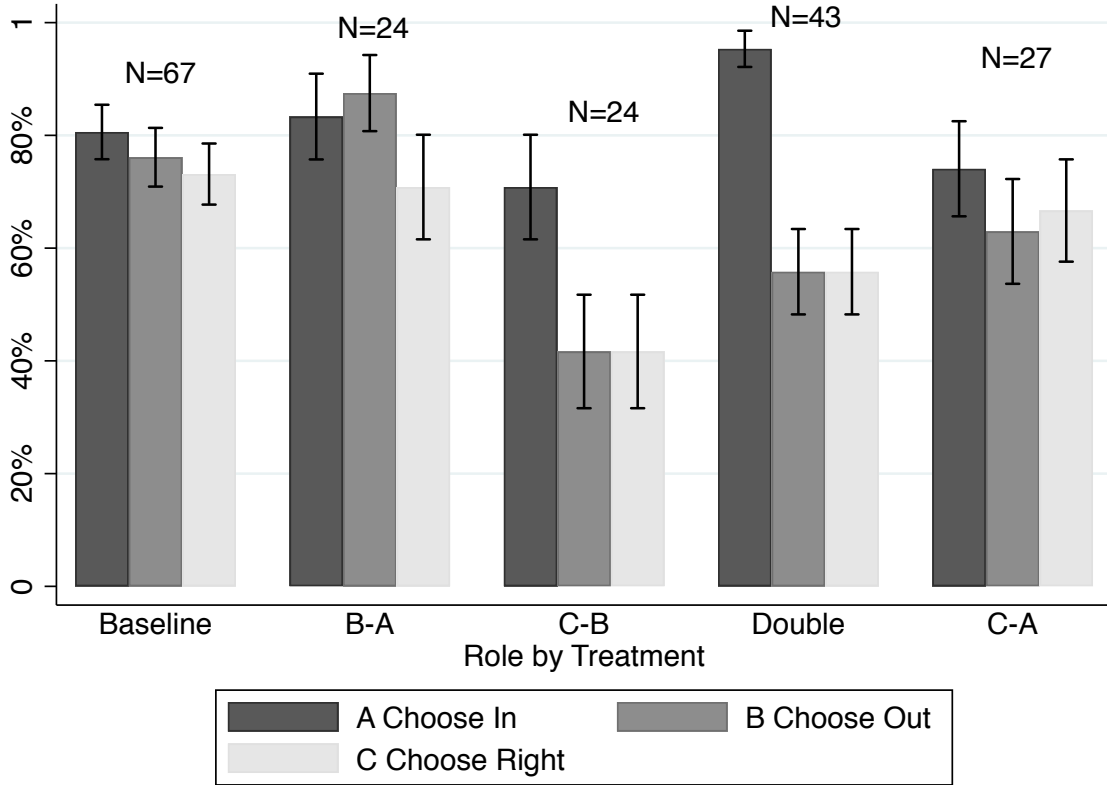


Figure 3: Choices By Role Across All Treatments

In the B-A treatment, we observe a similar percentage of As choosing *In* (83%), even more Bs - 87% - choosing *Out*, and a similar percentage of Cs choosing *Right* (71%). Compared with Baseline, more Bs chose *Out*, although the percentage is not

statistically significant. The promises from Bs reduce Bs willingness to explore potential Pareto improving opportunities, i.e., the outcome (*In, In, Left*).

In the C-B treatment, 71% of As chose *In*, half of Bs switched and chose *Out* (42%), and more than half of Cs (58%) chose the cooperative action - *Right*. When B receives a promise from C, not only does B choose to trust C, C also reciprocates. The promise from Cs are effective in binding Cs behaviors despite the misalignment of monetary interest between B and C.

In the Double treatment, 95% of As chose *In*, only 56% of Bs chose *Out*, and 56% of Cs chose *Right*. B starts the chain of promises by a promise to A indicating the willingness to choose *Out*; however, when new opportunities arise (C sends a promise conveying the willingness to cooperate), significantly less Bs ($p = 0.00$) chose *Out* (an action that is consistent with their promises). In this environment, promises are less effective in binding peoples behaviors.

When C could send a message to A (C-A treatment), 74% of As chose *In*, 63% of Bs chose *Out*, and 67% of Cs choose *Right*. The behaviors in this treatment resemble those in Baseline, which suggests that indirect promises are ineffective in promoting trust and reciprocity.

We also use the bootstrap method to compare the frequency of Nash equilibrium strategy profile (*In, Out, Right*) and Pareto efficient strategy profile (*In, In, Left*) among different treatments. We find that the B-A treatment has the highest frequency of Nash strategy profile, while C-B treatment has the highest frequency for Pareto efficient strategy profile. Overall, C-B treatment has the highest level of expected efficiency, partly due to the high frequency of Pareto efficient strategy being played. All communication treatment outperformed the baseline, with the exception of B-A treatment. For more detailed discussion, please refer to the Appendix A2.

5.3 Main Results

Result 1 *Our data support Prediction 1. Although promises were sent by Bs in B-A treatment, player B chooses Out as frequently in B-A treatment as in Baseline (no communication). And player C chooses Right the same rate in both B-A treatment and the Baseline.*

Tables 3 below present the Wilcoxon-Mann-Whitney two-sided test results with the null hypotheses that the frequency of Bs choosing *Out* is the same in the Baseline and the treatments (B-A, C-B, C-A or Double treatment). Similarly, Table 4 presents the test results with the null hypotheses that the frequency of Cs choosing *Right* is the

same in the Baseline as in the treatments. As predicted, both Bs' and Cs' behaviors in the B-A treatment are statistically indistinguishable from the Baseline. It suggests that in B-A treatment promises have limited effect on behaviors.¹² Regression analyses in Table 5 further support this result.

Table 3: Wilcoxon-Mann-Whitney two-sided test results for Player B

| Frequency of Bs Choosing <i>Out</i> | | | | |
|-------------------------------------|----------------|-----------------|----------------|-----------------|
| Baseline | B-A | C-B | C-A | Double |
| 51/67 (76%) | 21/24 (87%) | 10/24 (42%)* | 17/27 (63%) | 24/43 (56%)* |
| | $p = 0.24$ | $p = 0.00$ | $p = 0.20$ | $p = 0.03$ |

Note: *, **, and *** indicate $p < 0.10$, 0.05 and 0.01, respectively, two-sided tests.

Table 4: Wilcoxon-Mann-Whitney two-sided test results for Player C

| Frequency of Cs Choosing <i>Right</i> | | | | |
|---------------------------------------|----------------|-----------------|----------------|-----------------|
| Baseline | B-A | C-B | C-A | Double |
| 49/67 (73%) | 17/24 (71%) | 10/24 (42%)* | 18/27 (67%) | 24/43 (56%)* |
| | $p = 0.83$ | $p = 0.01$ | $p = 0.53$ | $p = 0.06$ |

Note: *, **, and *** indicate $p < 0.10$, 0.05 and 0.01, respectively, two-sided tests.

Result 2 *Our data provide partial support for Prediction 2. As predicted, there are significantly less Bs choosing Out and significantly less Cs choosing Right in C-B and Double treatment than Baseline. Contrary to Prediction 2, we did not find any behavioral difference between C-A treatment and Baseline for both player B and C.*

As shown in table 3, 4 and 5, we observe significantly fewer Bs and Cs choosing *Out*(or *Right*) in both the C-B and Double treatments than Baseline. This result is largely in line with the literature: people are willing to honor their non-binding promises at a personal cost.

¹²It is possible that the promises from Bs aren't so credible to A. However, it is B's belief about how A would perceive those promises that matters. We elicited first order belief data from B on the average probability of A choosing IN. In the Baseline, Bs believed that on average 87% of A chose IN. In B-A treatment, Bs believed that on average 96% of A chose IN. The difference is statistically significant at 10% ($P = 0.07$). It suggests that at least Bs believed that As were more likely to believe their promises and choose IN in B-A than Baseline.

Table 5: Probit Analyses of Treatment Effects

| | Player B's Choice | Player C's Choice |
|--------------|---------------------|---------------------|
| | (1) | (2) |
| B-A | -0.440 (0.33) | 0.068 (0.28) |
| C-B | 0.921*** (0.29) | 0.827*** (0.32) |
| C-A | 0.379 (0.27) | 0.186 (0.28) |
| Double | 0.564*** (0.19) | 0.471** (0.22) |
| Constant | -0.710*** (0.17) | -0.617*** (0.17) |
| Observations | 185 | 185 |

Note: Standard errors are clustered by subjects. * indicates significance level at 10%, ** at 5%, *** at 1%

Contrary to the theoretical prediction, we fail to find any communication treatment differences between the C-A treatment and the Baseline as shown in tables 4 and 5. This implies that contingent promises are ineffective in binding people's behaviors. To better understand this phenomenon, we first look at the rates of messages. As shown in Table 2, there is still a significant proportion of Cs who send messages (48%) and promises (26%). As a result, message frequency is unlikely to explain this result. Then we look further into the player Cs' beliefs. Specifically, we look at player C's first order belief and the second order belief in the C-A treatment. Note that player C's first order belief refers to C's belief about the percentage of player As to choose *In*, and the second order belief refers to C's belief about the average guess made by As who chose *In* about the proportion of Cs who chose *Left*. The belief comparison results are reported in Table 6.¹³ We can see that the first order beliefs do not differ for those sent messages or not ($p = .88$), while the second order beliefs are significantly different for those who sent messages and those who did not ($p = .005$). Similar results are found for promises. It suggests that messages and promises do change Cs second order belief. However, this change in second order belief did not translate into behaviors, as

¹³Note that in Table 6, the belief with promises is slightly lower than that with all messages. This is likely due to the fact that our categorization method may not match the perception of the message receivers perfectly. For more details, please refer to Chen & Houser(2016).

Cs in the C-A treatment behave the same way as in the Baseline. The reason may be that Cs do not perceive their messages and promises will make any difference in As' behaviors since the beliefs about As' *In* rate are the same for both message/promise sending and non-message/promise sending Cs. Indeed, when we compare As' actual *In* rate in C-A and Baseline, we do not observe any difference ($p = 0.49$).

Table 6: Player Cs' Beliefs In C-A Treatment

| First Order Belief | | | Second Order Belief | | |
|--------------------|------------|---------|---------------------|------------|---------|
| Msg | No Msg | P value | Msg | No Msg | P value |
| 55.62 | 55.62 | 0.88 | 63.31 | 15.71 | 0.005 |
| (8.94) | (7.37) | | (11.40) | (4.19) | |
| $n = 13$ | $n = 14$ | | $n = 13$ | $n = 14$ | |
| Promise | No Promise | P value | Promise | No Promise | P value |
| 51.14 | 54.95 | 0.72 | 61.14 | 30.75 | 0.08 |
| (15.16) | (5.78) | | (15.58) | (7.91) | |
| $n = 7$ | $n = 20$ | | $n = 7$ | $n = 20$ | |

Note: Wilcoxon-Mann-Whitney test, two-sided.

Result 3 *Our data support the Prediction 3b (Consequence based model). We find that Bs behave similarly in C-A and C-B treatments, however, Cs are significantly more likely to choose Right in C-A treatment than C-B treatment.*

Table 7 presents the Wilcoxon-Mann-Whitney two-sided test results with the null hypotheses that the *Out* and *Right* rate are comparable among C-B, C-A and Double treatment. There is significant difference in Cs behavior between the C-A and C-B treatments, which is contrary to what theories predict. However, this difference may result from the fact that there are significantly fewer messages and promises exchanged in the C-A treatment than in the C-B treatment. We performed the similar analysis with Cs first order and second order belief for the C-B treatment. Interestingly, we observe a significant difference in *both* first order and second order belief among those who send and those who do not send messages ($p = 0.01$ for first order belief and $p = 0.007$ for second order belief). This observation may offer additional explanations as to why *contingent* promises (C-A promises) do not work, while *definite* promises (C-B promises) do. Although the messages and promises do change people's second order beliefs, it may not necessarily change people's behaviors unless they believe their promises would ex-ante change the receivers' behaviors. Imagine a case where a friend invites you to a party, you know that the friend will be at the party whether or not you

go. And you promise the friend that you would go. You probably will not encounter a huge psychological cost when you break your promise and decide not to go.

Result 4 *Our data are more in line with Prediction 4b (consequence model with direct guilt only). We find that Bs are equally likely to choose Out in C-B and Double treatment. Similarly, the Right rate is comparable in both C-B and Double treatment.*

From table 7 and 8, we cannot reject the null hypothesis that *out* and *Right* rate are the same for Bs and Cs in both C-B and Double treatment. It indicates that behaviors are better prescribed by consequence based model with direct guilt only. In chained promises where one broken promise may impact multiple players, people do not seem to incur indirect guilt, rather the guilt stops at the most immediate counterpart. As a result, people at the end of the chain are as likely to break their promises as those at the front of the chain, even if the potential harm is greater for those at the end of the chain.

Table 7: Wilcoxon-Mann-Whitney Two-sided Test Results For Player B and C (Within Treatment)

| Frequency of Bs Choosing <i>Out</i> | | | Frequency of Cs Choosing <i>Right</i> | | |
|--|------------|------------|--|------------|------------|
| C-B | C-A | Double | C-B | C-A | Double |
| 10/24 | 17/27 | 24/43 | 10/24 | 18/27 | 24/43 |
| (42%) | (63%) | (56%) | (42%) | (67%)* | (56%) |
| | $p = 0.13$ | $p = 0.27$ | | $p = 0.08$ | $p = 0.27$ |

Note: *, **, and *** indicate $p < 0.10$, 0.05 and 0.01, respectively, two-sided tests.

Table 8: Probit Analyses of Between Treatment Effects

| | Player B's Choice | Player C's Choice |
|--------------|-------------------|-------------------|
| | (1) | (2) |
| C-A | -0.541 (0.36) | -0.641* (0.36) |
| Double | -0.357 (0.32) | -0.357 (0.32) |
| Constant | 0.210 (0.26) | 0.210 (0.26) |
| Observations | 94 | 94 |

Note: The control group in the regressions is C-B treatment. Standard errors are clustered by subjects. * indicates significance level at 10%, ** at 5%, *** at 1%

6 Discussion and Conclusion

We extended the empirical study of promises to a three-person environment. Doing this enabled us to study two features of promises in the naturally occurring world that have not been studied in the previous literature. For one, we are able to determine the impact of “contingent” promises, which refers to promises that need to be kept (or broken) only in the event that another event first occurs. Second, we were able to study “promise chains”, by which we mean as a set of sequentially-made promises. In both cases, we are able to apply existing theories to these novel environments in order to predict promise keeping behavior.

The two theories we considered were intrinsic preference for honesty and consequence-based models, specifically guilt aversion. Neither model is able to account for all the patterns in our data. In particular, “contingent” promises have essentially no effect on behavior, while both models predict promises in these cases should promote cooperation. The behavior we observe in promise chains, on the other hand, can be well-described by models of guilt-aversion under the condition that people only experience guilt towards their immediate counterpart, and not any others who are at earlier positions in the chain and might be harmed by the breaking of a promise. On balance, consequence based model with direct guilt best characterizes the behaviors we observed in the data.

Broadly, we find, as have many others, that communication, especially a promise, impacts behavior in the game we study. When there is little conflict of interest, as

predicted, we do not observe significant effect of promises. When there is conflict of interest and when the promise is *definite* (i.e., not contingent), we find that people are indeed willing to forgo monetary payoff to keep their promises. However, when the promise is *contingent*, contrary to the theory predictions, promises no longer work in that people are equally likely to choose the selfish action whether they make a promise or not.¹⁴ Further investigation into the players' beliefs sheds some light on some possible explanations for this result. In order for the players to incur the psychological cost of breaking promises, the promise sender has to not only experience a shift in her second order belief, but also ex-ante anticipate her promises to shift the receiver's behavior. As detailed in the example provided in the the previous section, one may not necessarily experience the psychological cost of breaking one's promise and failing to appear at a friend's party, if she knows that her friend will attend the party regardless of her promise. In comparison, one is more likely to experience the psychological cost of breaking the promise if she knows that her friend will only go to the party because of her promise. It may be useful to incorporate such insights into guilt aversion models. Another possible explanation for the ineffectiveness of the contingent promise may be diminishing sensitivity to guilt. As the number of contingencies grow, one becomes less susceptible to guilt. In our game, C's decision on whether to keep the promise is only made after both A and B's actions. The additional level of contingency may have reduced the effect of the promise. Further research is needed to better inform those possibilities.

Finally, we find that promise chains do not seem to improve the trustworthiness, in that people at the end of the the chain are no more likely to honor their promises than people at earlier positions. This observation may have further implications for application of guilt aversion model, i.e., people only experience direct guilt toward their immediate counterpart despite the indirect harm to others that their actions may cause. This observation may help guide further modelling of guilt-averse preferences.

Our conclusions are limited by the fact that the people in our games were strangers and anonymous. While we did not find promise chains to be effective in increasing the likelihood of promise-keeping, such chains may be more effective if the participants were known to each other. Examples might include micro-lending networks, where others have found such chains to reduce default rates. Our conclusions on contingent promises are also limited by the use of strategy method. Although the conclusions

¹⁴The results we highlight here are based on the comparison between C-A treatment and baseline, as well as C-B treatment and baseline. Although there is significant difference in Cs behavior between C-A and C-B treatment ($P=0.08$, two-sided test), the payoff asymmetry between A and B may have contributed to some of this result.

based on treatment comparisons are internally valid, as strategy method is implemented throughout all treatment conditions, it is an open question whether our results using strategy method can be extended to situations with direct response. There is some literature suggesting that our results may be applicable to scenarios with direct response.¹⁵ Further theoretical and empirical research informing the value of promise chains and the behavioral difference between strategy method and direct response in different contexts would be quite valuable.

¹⁵Amdur and Schimick(2012) argues that there is no statistically significant difference in trustworthiness measured by strategy method and direct response. Our game shares the similar payoff structure and it is possible that our results may be applicable to situations with direct response. However, it is worth to note that Casari and Cason (2009) found significant difference in trustworthiness measured by direct response and strategy method in a standard trust game setting.

References

- David Amdur and Ethan Schmick. Does the direct-response method induce guilt aversion in a trust game? 2012.
- P Battigalli and Martin Dufwenberg. Guilt in games. *The American economic review*, 97(2):170–176, 2007. URL <http://www.jstor.org/stable/10.2307/30034441>.
- M Bigoni, S Bortolotti, and Marco Casari. Trustworthy by Convention. 2012. URL http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2057607.
- Marco Casari and Timothy N. Cason. The strategy method lowers measured trustworthy behavior. *Economics Letters*, 103(3):157–159, 2009. ISSN 01651765. doi: 10.1016/j.econlet.2009.03.012.
- Alessandra Cassar and Mary Rigdon. Trust and trustworthiness in networked exchange. *Games and Economic Behavior*, 71(2):282–303, mar 2011. ISSN 08998256. doi: 10.1016/j.geb.2010.04.003. URL <http://linkinghub.elsevier.com/retrieve/pii/S0899825610000643>.
- G. Charness and M. Dufwenberg. Promises and partnership. *Econometrica*, 74(6):1579–1601, 2006. URL <http://onlinelibrary.wiley.com/doi/10.1111/j.1468-0262.2006.00719.x/abstract>.
- Jingnan Chen and Daniel Houser. Promises and lies: can observers detect deception in written messages. *Experimental Economics*, pages 1–24, jul 2016.
- Martin Dufwenberg and Georg Kirchsteiger. A theory of sequential reciprocity. *Games and Economic Behavior*, 47(2):268–298, may 2004. ISSN 08998256. doi: 10.1016/j.geb.2003.06.003. URL <http://linkinghub.elsevier.com/retrieve/pii/S0899825603001908>.
- Tore Ellingsen and Magnus Johannesson. PROMISES , THREATS AND FAIRNESS *. *The Economic Journal*, 114:397–420, 2004.
- Ernst Fehr and Klaus M Schmidt. A Theory Of Fairness, Competition, and Cooperation. *Quarterly Journal of Economics*, 114(3):817–868, 1999. ISSN 00335533. doi: 10.1162/003355399556151. URL <http://www.mitpressjournals.org/doi/abs/10.1162/003355399556151>.

- J Geanakoplos, David Pearce, and E Stacchetti. Psychological games and sequential rationality. *Games and Economic Behavior*, 1(1):60–79, 1989. URL <http://www.sciencedirect.com/science/article/pii/0899825689900055>.
- Rajna Gibson, Carmen Tanner, and Alexander F Wagner. Preferences for Truthfulness: Heterogeneity among and within Individuals. *American Economic Review*, 103(1): 532–548, feb 2013. ISSN 0002-8282. doi: 10.1257/aer.103.1.532. URL <http://pubs.aeaweb.org/doi/abs/10.1257/aer.103.1.532>.
- U. Gneezy. Deception: The Role of Consequences. *The American Economic Review*, 95(1):384–394, 2005. URL <http://www.jstor.org/stable/10.2307/4132685>.
- Sjaak Hurkens and Navin Kartik. Would I lie to you? On social preferences and lying aversion. *Experimental Economics*, 12(2):180–192, 2009. ISSN 13864157. doi: 10.1007/s10683-008-9208-2. URL <http://www.springerlink.com/index/10.1007/s10683-008-9208-2http://www.springerlink.com/index/h8078018898516w0.pdf>.
- Tobias Lundquist, Tore Ellingsen, Erik Gribbe, and Magnus Johannesson. The aversion to lying. *Journal of Economic Behavior & Organization*, 70(1-2):81–92, may 2009. ISSN 01672681. doi: 10.1016/j.jebo.2009.02.010. URL <http://linkinghub.elsevier.com/retrieve/pii/S0167268109000523>.
- T. Miettinen and S. Suetens. Communication and Guilt in a Prisoner’s Dilemma. *Journal of Conflict Resolution*, 52(6):945–960, jul 2008. ISSN 0022-0027. doi: 10.1177/0022002708322325. URL <http://jcr.sagepub.com/cgi/doi/10.1177/0022002708322325>.
- Thomas a. Rietz, Roman M. Sheremeta, Timothy W. Shields, and Vernon L. Smith. Transparency, efficiency and the distribution of economic welfare in pass-through investment trust games. *Journal of Economic Behavior & Organization*, oct 2012. ISSN 01672681. doi: 10.1016/j.jebo.2012.09.019. URL <http://linkinghub.elsevier.com/retrieve/pii/S0167268112002004>.
- Roman M. Sheremeta and Jingjing Zhang. Three-Player Trust Game With Insider Communication. *Economic Inquiry*, pages no–no, mar 2013. ISSN 00952583. doi: 10.1111/ecin.12018. URL <http://doi.wiley.com/10.1111/ecin.12018>.
- Christoph Vanberg. Why Do People Keep Their Promises? An Experimental Test

of Two Explanations. *Econometrica*, 76(6):1467–1480, 2008. ISSN 0012-9682. doi: 10.3982/ECTA7673. URL <http://doi.wiley.com/10.3982/ECTA7673>.

Appendix

A.1 Theoretical Predictions

Among the models of preference for honesty¹⁶, there are two main varieties. One is the homogeneous aversion to lying model suggested by Ellingsen and Johannesson (2004), Miettinen and Suetens (2008), and Vanberg (2008), where the model assumes that people incur a similar fixed cost in their utilities when caught lying; the other is the heterogeneous cost of lying model, where different people might incur different costs while caught lying (*e.g.*, Gibson et al. (2013)).

Homogeneous aversion to lying model

Ellingsen and Johannesson (2004) proposed a modified model (based on Fehr and Schmidt (1999)) (we now call EJ model) with an added universal cost component l to reflect the universal cost of lying. If there can be no communication, there is no cost of lying. Thus, in the no-communication baseline game, the predictions correspond to the case with selfish preferences.

We define player i as those who communicate and state a verifiable contract/commitment¹⁷. Formally, player i has the following simplified utility function:

$$u_i = \begin{cases} m_i - l & \text{if player } i \text{ lies} \\ m_i & \text{otherwise.} \end{cases} \quad (4)$$

where m_i denotes agent i 's immediate monetary payoff, $-l$ denotes the utility loss from lying. Notice that l is invariant to players therefore implies the assumption that people share a homogeneous cost to lying¹⁸. G_2 , G_3 , G_4 in figure 1, 2, 3 model this for B-A, C-B, and Double treatment for players who communicate and indicate a verifiable promise.

For all models under consideration, the C-B and C-A treatments are treated exactly the same (although player B has more direct decision relevance to player C than player A). Consequently, in the following sections, we do not explain in depth the predictions for the C-A treatment, they are exactly the same as for the C-B treatment. Notice

¹⁶In some papers, it also called lying aversion, or the cost of lying model (for example, Lundquist et al. (2009)).

¹⁷For those who choose not to communicate or send non-verifiable communication, their decision problems are modeled with standard self-interest maximization. And (*In*, *Out*, *Right*) remains the unique backward-induction solution.

¹⁸Similarly, Miettinen (2008) introduces similar invariant fixed cost to lying to model the effects of pre-play agreements in contracts. Vanberg (2008) provides supporting laboratory evidence suggesting that lack of lying behaviors in his experiment can be better explained with a simple cost of lying model.

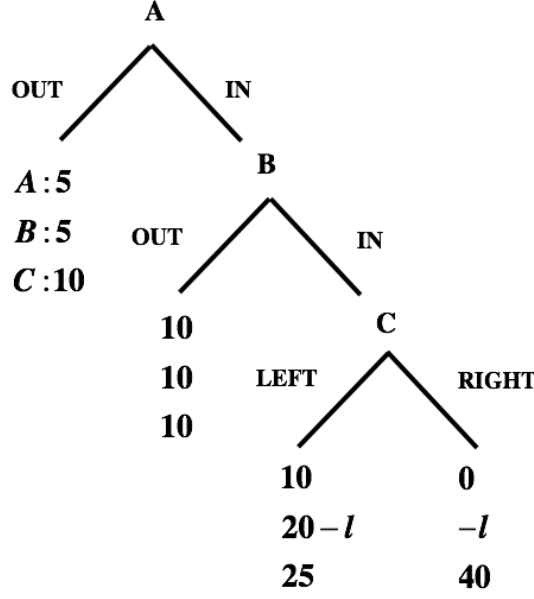


Figure 1: G_2 - B-A treatment

that game $G_2 - G_4$ is a nonstandard game in which the utilities are not just numbers (m_i) at the end nodes but rather reflect the adjusted utility (u_i) (this applies to all the games in the following sections).

We denote $\varphi_{p,t,s}$ the percentage of players p in treatment t choosing strategy s , while $p \in \{B, C\}$, $t \in \{Base, BA, CB, CA, Double\}$, $s \in \{Out, R\}$, where *Base*, *BA*, *CB*, *CA*, *Double* represent baseline, B-A, C-B, C-A and Double treatment respectively, and *In* and *R* represent strategy *In* and *Right* respectively. For example, $\varphi_{B,Base,In}$ denotes the percentage of player B choosing *In* in Baseline treatment.

Comparing all communication treatments with Baseline, we have the following:

Proposition 1 (*EJ model*)¹⁹ If $l < 15$, $\varphi_{p,Base,s} = \varphi_{p,j,s}$, where $j \in t$ and $j \neq Base$; if $l \geq 15$, **for player B**, $\varphi_{B,Base,Out} = \varphi_{B,BA,Out}$, $\varphi_{B,Base,Out} > \varphi_{B,CB,Out} = \varphi_{B,CA,Out}$, $\varphi_{B,Base,Out} > \varphi_{B,Double,Out}$; **for player C**, $\varphi_{C,Base,R} = \varphi_{C,BA,R}$, $\varphi_{C,Base,R} > \varphi_{C,CB,R} = \varphi_{C,CA,R}$, $\varphi_{C,Base,R} < \varphi_{C,Double,R}$.

Proof. 1) In the B-A treatment, *Right* is still the dominant strategy for C, *In* and *Out* remain the best responses for B and A respectively in G_2 as in G_1 . Therefore, we expect no treatment difference between B-A and Baseline.

¹⁹According to Ellingsen and Johannesson (2004), the estimated l is around 21.43 SEK \approx USD 2.64. If we apply this estimates to our game, $l = 2.64 < 15$, we should expect no effects at all across treatments.

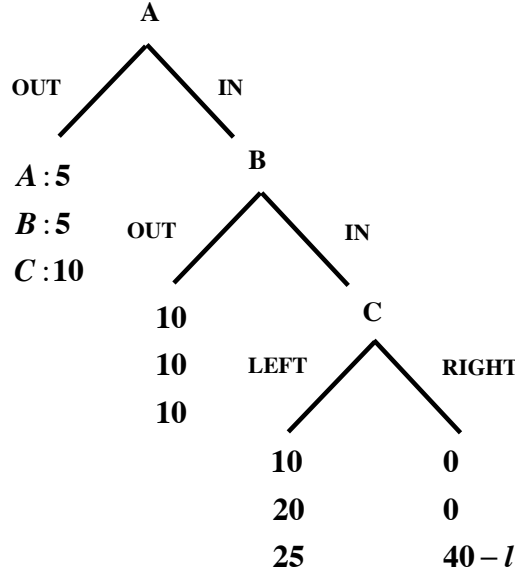


Figure 2: G_3 - C-B(C-A) treatment

2) In the C-B (C-A) treatment, if $l \geq 15$, *Left* becomes the dominant strategy for C, (In, In) is the best response strategy profile for both A and B. As a result, we expect to observe a higher percentage of Bs choosing *In* and a higher percentage of Cs choosing *Left* than the Baseline. If $l < 15$, however, *Right* once again becomes the dominant strategy for C. For A and B, the best responses are *In* and *Out*. In this case, we expect no treatment differences between the C-B and Baseline treatments.

3) For the Double treatment, if $l \geq 15$, *In*, *Out*, *Left* describe the best responses for players A, B, C respectively; if $l < 15$, *In*, *Out*, *Right* are the best responses instead. Therefore, if $l < 15$, we should not expect to see treatment differences from Baseline, and communication in the form of cheap talk has no effect on the behavior of all players. If $l \geq 15$, we should observe an increased rate of player B choosing *In* for Double compared with the Baseline Treatment; for player C, we should observe an decreased frequency of choosing *Right* in the Double treatment compared with Baseline. ■

In order to make within-treatment comparisons, we have to make one simple assumption:

Assumption 1 *The frequency of informal contract/commitment exchanged is constant from the same role.*

Assumption 1 implies that: 1) the percentage of Bs sending informal contracts in the B-A treatment is comparable to that in the Double treatment; 2) the percentage of

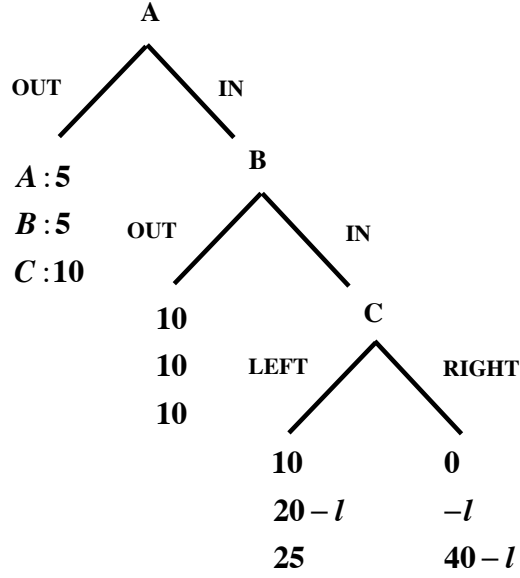


Figure 3: G_4 - Double treatment

Cs exchanging informal commitment in the C-B treatment is comparable to that in the Double treatment; and 3) the percentage of Bs receiving informal contracts in the C-B treatment is comparable to that in the Double treatment. We denote that b and c percent of Bs and Cs send informal contracts respectively, where $b, c \in [0, 1]$.

Proposition 2 (*EJ model*) *If $l < 15$, for player B, $\varphi_{B,BA,Out} = \varphi_{B,Double,Out} = \varphi_{B,CB,Out}$; for player C, $\varphi_{C,BA,R} = \varphi_{C,CB,R} = \varphi_{C,Double,R}$. If $l \geq 15$, for player B, $\varphi_{B,CB,Out} < \varphi_{B,Double,Out} < \varphi_{B,BA,Out}$; for player C, $\varphi_{C,BA,R} > \varphi_{C,CB,R} = \varphi_{C,Double,R}$.*

Proof. 1) In both the B-A and Double treatments, if B is rational and sends a contract, he will only choose *In* if $p(20 - l) - (1 - p)l > 10$, i.e., $20p - l > 10$, where p is Cs probability of choosing *Left*. Specifically, in the B-A treatment, $p = 0$, regardless of the value of l , it is B's best response to choose *Out* (whether or not they send informal contracts).

2) In the Double treatment, if $l < 15$, Cs always choose *Right* whether they send informal commitment or not, i.e., $p = 0$. In response to that, all Bs choose *Out* whether they receive informal contracts or not. If $l \geq 15$, the situation is more complicated. $1 - c$ percent of Bs do not receive informal contracts, i.e., $p = 0$, the best response them is *Out*. For the remaining Bs who receive informal contracts, $b \cdot c$ percent of Bs also send informal contracts, in this case, $p = 1$, and $20 - l \leq 5 < 10$, the best response is *Out*; $(1 - b) \cdot c$ percent of Bs do not send informal contracts but receive one from C, $p = 1$

and $10 < 20p$, their best response is instead *In*. In sum, $\varphi_{B,Double,Out} = 1 - (1 - b) \cdot c$.

3) In C-B treatment, if B is rational and send informal contracts, he will only choose *In* if $20p > 10$, i.e., $p > 0.5$. If $l < 15$, *Out* is B's best response since $p = 0$. If $l \geq 15$, c percent of Bs receive informal contracts from C ($p = 1$), *In* is the best response. For the remaining $1 - c$ percent of Bs who do not receive informal contracts from C, *Out* is the best response. And $\varphi_{B,CB,Out} = 1 - c$. Since $c > c(1 - b) > 0$, we have $\varphi_{B,CB,Out} < \varphi_{B,Double,Out} < \varphi_{B,BA,Out}$.

4) If C is rational and send informal contracts, he/she will only choose *Left* if $25 > 40 - l$. In both C-B and Double treatment, the decision problem for C is the same. Therefore, we expect no differences between the two treatments. And the rate of C choosing *Right* in both C-B and Double treatment will be smaller than B-A treatment where there is no incentive for C to choose *Left* at all. ■

Combine both Proposition 1 and 2, we have:

Proposition 3 (*EJ model*) If $l < 15$, $\varphi_{p,i,s} = \varphi_{p,j,s}$, where $i, j \in t$ and $i \neq j$; If $l \geq 15$, **for player B**, $\varphi_{B,Base,Out} = \varphi_{B,BA,Out} = \varphi_{B,Double,Out} < \varphi_{B,CB,Out}$; **for player C**, $\varphi_{C,Base,R} = \varphi_{C,BA,R} = \varphi_{C,Double,R} > \varphi_{C,CB,R}$.

In the EJ model, everyone suffers the same cost from lying; therefore, if the benefit of lying outweighs the cost, any communication is futile since words said will never be kept. However, if the cost of lying outweighs the benefit, we expect people to keep their contracts (if they send one). Thus, in both the B-A and Double treatments where player B may send a contract, we hypothesize a higher rate of *In* than in the C-B (C-A) treatments. Similarly, we expect to see a higher rate of cooperative action from player C choosing *Left* in both the C-B (C-A) and Double treatments than in any other treatments.

Heterogeneous cost to lying model

Gibson et al. (2013) propose and test a heterogeneous preference for honesty model (we call GTF model). The GTF model is very similar to the EJ model in that the cost of lying is independent from any type of belief. The difference is that instead of a fixed l for all players, the GTF model assumes that each individual might have a different cost l_i associated with lying. The utility function is as follows:

$$u_i = \begin{cases} m_i - l_i & \text{if player } i \text{ lies} \\ m_i & \text{otherwise.} \end{cases} \quad (5)$$

where l_i indicates the utility loss player i endures when he(or she breaks the promise

or lies and $l_i \in [0, +\infty]$.

Game G_5 , G_6 , G_7 in Figure 4, 5, 6 incorporate those. And we have Proposition 4 below.

Proposition 4 (*GTF model*) **For player B**, $\varphi_{B,Base,Out} = \varphi_{B,BA,Out} > \varphi_{B,Double,Out} > \varphi_{B,CB,Out} = \varphi_{B,CA,Out}$; **for player C**, $\varphi_{C,Base,R} = \varphi_{C,BA,R} > \varphi_{C,Double,R} = \varphi_{C,CB,R} = \varphi_{C,CA,R}$.

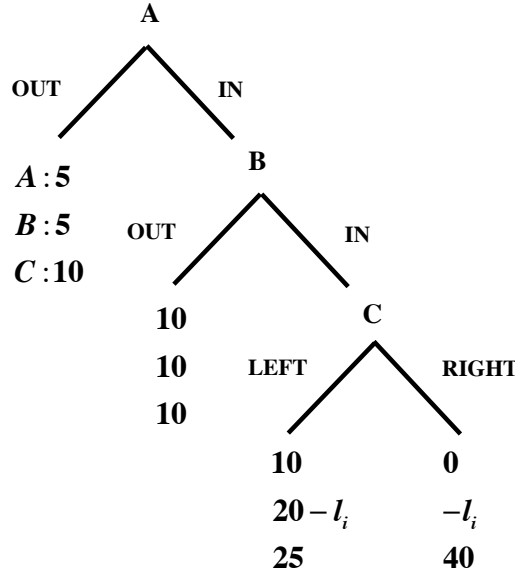


Figure 4: G_5 - B-A treatment

Proof. 1) Similarly, in the B-A treatment, all Bs choose *Out* given that *Right* is the dominant strategy for Cs.

2) In the Double treatment, for the $1 - c$ percent of Bs who do not receive informal contracts, *Out* is the best response. For the $b \cdot c$ percent of Bs who both send and receive informal contracts from C, they will only choose *In* if $p(20 - l_i) - (1 - p)l_i \geq 10$. $p = 1$ if $l_j \leq 15$, and $p = 0$ otherwise. Assume that $Pr(l_j \geq 15) = w$ and $Pr(l_i \leq 10) = k$, only $w \cdot k \cdot b \cdot c$ percent of B (whose cost to default on informal contracts is small enough, while his/her partners cost is big enough) choose *In* and $(1 - w \cdot k)b \cdot c$ choose *In*. For the remaining $(1 - b)c$ percent of Bs who do not send but receive informal contracts from C, they will only choose *In*, if $20p \geq 10$. We have $w \cdot (1 - b)c$ percent of Bs choosing *In*, and $(1 - w)(1 - b)c$ percent of Bs choosing *Out*. In total, $w \cdot k \cdot b \cdot c + w \cdot (1 - b)c$ percent of Bs choose *In* and the rest choose *Out*.

3) In the C-B treatment, similar to the Double treatment, the $1 - c$ percent of Bs who do not receive informal contracts choose *Out*. The rest will choose *In* only if $20p \geq 10$.

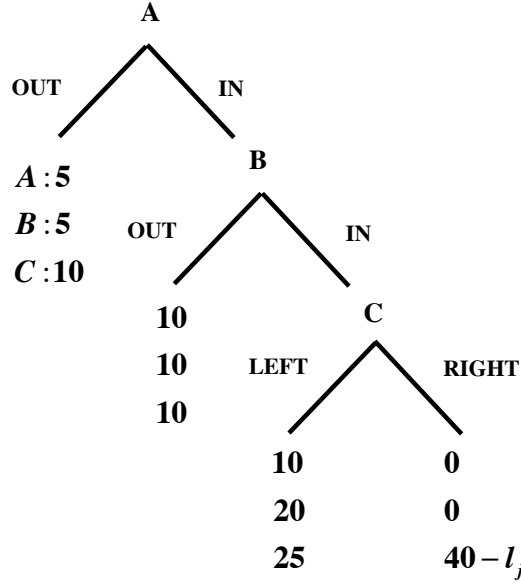


Figure 5: G_6 - C-B(C-A) treatment

Since $p = 1$ with probability w , we have $w \cdot c$ percent of Bs choosing *In* and $1 - w \cdot c$ percent of Bs choosing *Out*. Since $\varphi_{B,Double,In} = w \cdot k \cdot b \cdot c + w \cdot (1 - b)c$ is smaller than $\varphi_{B,CB,In} = w \cdot c$, we should expect a higher frequency of Bs choosing *In* from C-B to the Double treatment. ■

For player B's behavior, the GTF model hypothesizes that B may choose *In* more frequently in the C-B treatment than in the Double treatment and also more frequently in Double treatment than B-A and Baseline treatments. The intuitions are as follows: in the B-A and baseline treatments, it is always the best response for B to choose *Out*, anticipating C to prefer *Right*. In the Double treatment, however, choosing *In* can become the best response for some message-receiving Bs when there is increased probability of C choosing *Left* (for those Cs with $l_j > 15$). In the C-B treatment, choosing *In* can be the best response for ALL message-sending Bs when their matched Cs are with $l_j > 15$.

Consequence-based Model

Charness and Dufwenberg (2006) proposed a guilt from blame model built on the psychological game theory framework developed by Geanakoplos et al. (1989), furthered by Dufwenberg and Kirchsteiger (2004) and formalized in Battigalli and Dufwenberg (2007). In this model, the cost of lying comes from the guilt of letting someone down, and the level of guilt a player suffers depends on the level of harm he imposes on

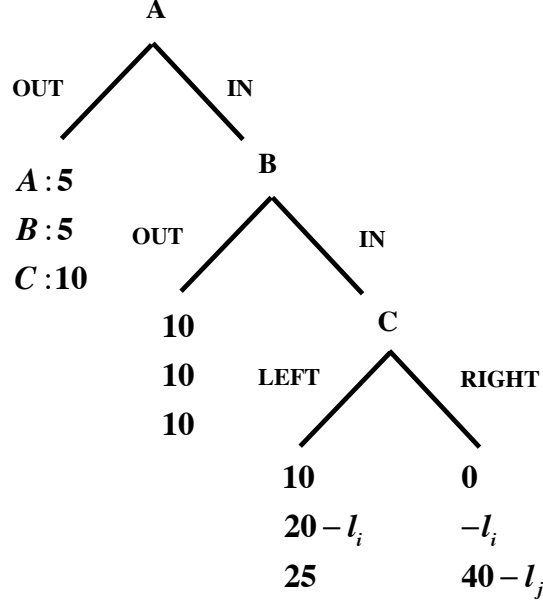


Figure 6: G_7 - Double treatment

others relative to what the others believe they will suffer (i.e., the difference between the players actual action and the action the player believes others believe he would take). In a sense, this model is a different take on social preference models, where the degree one cares about others also depends on the belief one holds about others belief about him/her.

According to the Guilt Aversion Model (we call CD model from now onwards), player i has the following modified utility:

$$u_i = m_i - \gamma_i \cdot \tau_i \cdot \Delta m_j, \quad i \neq j \quad (6)$$

where γ_i denotes player i 's sensitivity to guilt, and it is independent from τ_i ; $\gamma_i \in [0, +\infty)$; τ_i denotes player i 's belief about j 's belief about i , $\tau_i = E(\mu_j)$, where μ_j is the probability player j assigns to i 's move, $\mu_j \in [0, 1]$; Δm_j denotes j 's monetary loss between what i thinks j thinks that i would do and what i actually does. In light of our game with B-A treatment where B sends a contract to A, Δm_A in this case would be \$10, which is what A would get given B's informal contracts (this is also the monetary payoff B would expect A to believe B would be able to give him/her) minus 0 (if, instead of choosing *Out*, B deviates from his/her contract and chooses *In*. since the best strategy for C is to choose *Right*, A would get 0 given B and Cs (*In*, *Right*) choices) equals 10.

G_8 , G_9 , G_{10} in Figure 8, 9, and 10, respectively, incorporate the CD model for all

treatments. Notice that in the B-A treatment, player B doesn't suffer from guilt, when he/she chooses *In* (violating informal contracts) and player C chooses *Left*. The reason is that when player C plays *Left*, A receives \$10, which is the same if B has chosen *Out*. In other words, violating the contract has no monetary consequences to A if C chooses *Left*, thus B doesn't feel any letting down guilt. Similarly, in the Double treatment, B doesn't suffer utility loss from guilt choosing *In* and breaking informal contracts as long as C chooses *Left*.

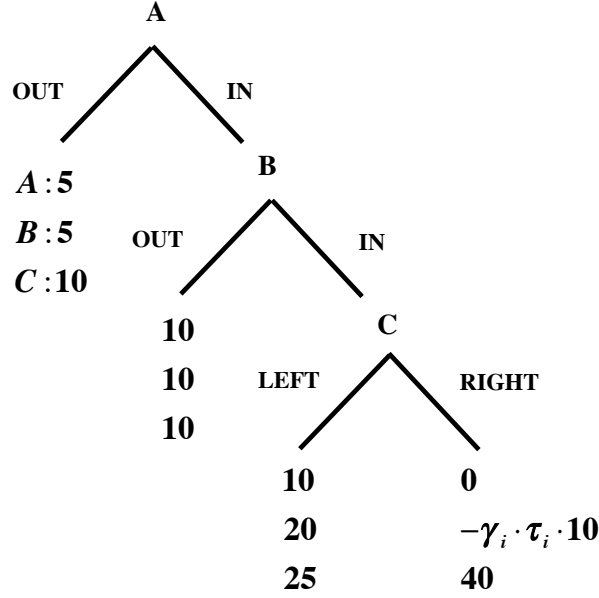


Figure 7: G_8 - B-A treatment

Proposition 5 (*CD model*) **For player B**, $\varphi_{B,Base,Out} = \varphi_{B,BA,Out} > \varphi_{B,CB,Out} = \varphi_{B,CA,Out} = \varphi_{B,Double,Out}$; **for player C**, $\varphi_{C,Base,R} = \varphi_{C,BA,R} = \varphi_{C,Double,R} > \varphi_{C,CB,R} = \varphi_{C,CA,R}$.

Proof. 1) Similar to the analysis for the EJ model, b percent of Bs in B-A treatment will only choose *In*, if $20p - \gamma_i \cdot \tau_i \cdot 10 \cdot (1 - p) \geq 10$. Anticipating that *Right* is Cs dominant strategy, $p = 0$, *Out* is the best response for all Bs.

2) In the Double treatment, for the $1 - c$ percent of Bs who do not receive informal contracts, *Out* is the best response. For the $b \cdot c$ percent of Bs who both send and receive informal contracts from C, they will only choose *In* if $20p - \gamma_i \cdot \tau_i \cdot 10(1 - p) \geq 10$. $p = 1$ if $25 \geq 40 - \gamma_j \cdot \tau_j \cdot 20$, and $p = 0$ otherwise. Assume that $Pr(25 \geq 40 - \gamma_j \cdot \tau_j \cdot 20) = v$, we have $v \cdot b \cdot c$ percent of B choose *In* and $(1 - v)b \cdot c$ choose *Out*. For the remaining $(1 - b)c$ percent of Bs who do not send but receive informal contracts from C, they

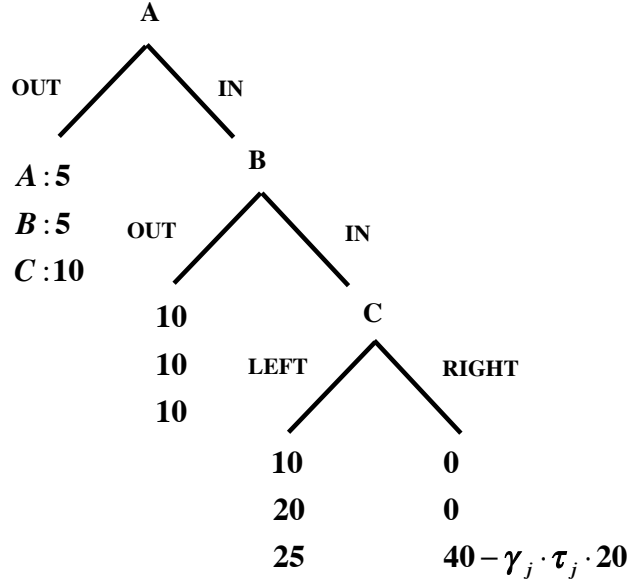


Figure 8: G_9 - C-B(C-A) treatment

will only choose *In*, if $20p \geq 10$. We have $v \cdot (1 - b)c$ percent of Bs choosing *In*, and $(1 - v)(1 - b)c$ percent of Bs choosing *Out*. In total, $v \cdot c$ percent of Bs choose *In* and $1 - v \cdot c$ percent of Bs choose *Out*, i.e., $\varphi_{B,Double,Out} = 1 - v \cdot c$.

3) In the C-B treatment, similar to the Double treatment, the $1 - c$ percent of Bs who do not receive informal contracts choose *Out*. The rest will choose *In* only if $20p \geq 10$. Since $p = 1$ with probability v , we have $v \cdot c$ percent of Bs choosing *In* and $1 - v \cdot c$ percent of Bs choosing *Out*, which is the same as in Double treatment. ■

The CD model offers the same predictions on player Cs behavior as EJ and GTF models; likewise, the intuition (i.e., lying is costly, although in the CD model defaulting on informal contracts is costly due to the fact that lying always lowers a partners payoff) is similar among all three models. For player B, the CD model differs from the EJ model in that each player may have varying costs from lying, and it is different from GTF model in that the different costs are from “letting-down” resulted guilt. Also, the level of guilt depends on guilt sensitivity and potential harm that may incur to others while in GTF model, the cost of lying is innate and independent from consequences.

The key prediction difference between the CD and EJ models (when $l > 15$) is player B’s behavior in the Double treatment: in the EJ model, breaking informal contracts is costly, and consequently player B in the Double treatment behaves the same way in the B-A treatment; in the CD model, however, breaking informal

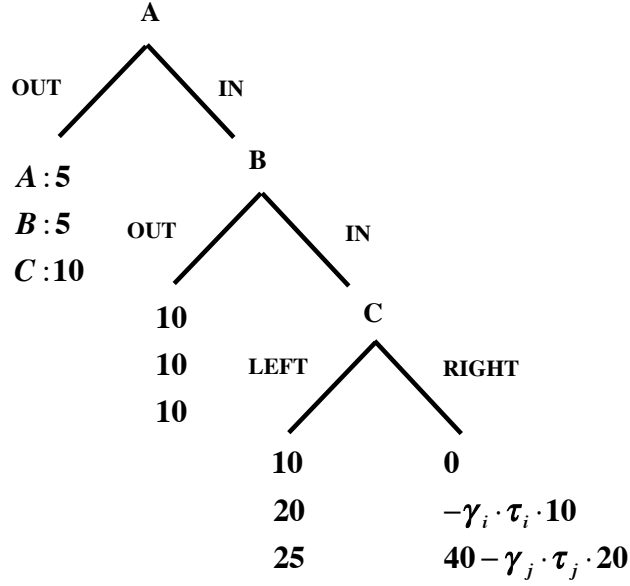


Figure 9: G_{10} - Double treatment

contracts can be costless for player B as long as there is no foreseeable harm to player A (which is clearly the case if C chooses *Left*); as a result, we should observe an increased rate of B choosing *In* for the Double treatment versus the B-A treatment.

Simple Type Model

Hurkens and Kartik (2009) put forward a simple type model that can make sense of the observations in Gneezy (2005) (we call it the HK model). The HK model assumes that there are two types of people, one with infinite cost of lying (honest type), and the other with zero cost of lying (economic type).

$$u_i = m_i - i \cdot L \quad (7)$$

where i indicates players' type, $i \in \{0, 1\}$, if $i = 0$, player i is economic type with no cost to lying, while if $i = 1$, i is honest type with infinite cost to lying; L denotes an enormous cost to breaking informal contracts and $L \rightarrow +\infty$. Assume that in the population ρ percent are honest types, and $1 - \rho$ percent are economic types. The implications of the HK model for our game are detailed in Proposition 6.

Proposition 6 (*HK model*) **For player B**, $\varphi_{B,Base,Out} = \varphi_{B,BA,Out} > \varphi_{B,Double,Out} = \varphi_{B,CB,Out} = \varphi_{B,CA,Out}$; **for player C**, $\varphi_{C,Base,R} = \varphi_{C,BA,R} > \varphi_{C,Double,R} = \varphi_{C,CB,R} = \varphi_{C,CA,R}$.

Proof. 1) As with all other theories, the HK model predicts that all Bs choose *Out* given that *Right* is the dominant strategy for Cs in the B-A treatment.

2) In the Double treatment, for the $1 - c$ percent of Bs who do not receive informal contracts, *Out* is the best response. For the $b \cdot c$ percent of Bs who both send and receive informal contracts from C, $\rho \cdot b \cdot c$ of them are the honest type, and will always choose *Out* as they state in the informal contracts. Among the remaining $(1 - \rho)b \cdot c$ economic types, they will only choose *In* if $20p \geq 10$. $p = 1$ if they receive informal contracts from an honest type, which happens with probability of ρ , and $p = 0$ otherwise. Therefore, we have $\rho(1 - \rho)b \cdot c$ percent of Bs choosing *In* and $(1 - \rho)^2 b \cdot c$ choosing *Out*. For the remaining $(1 - b)c$ percent of Bs who do not send but receive informal contracts from C, they will only choose *In*, if $20p \geq 10$. We have $\rho \cdot (1 - b)c$ percent of Bs choosing *In*, and $(1 - \rho)(1 - b)c$ percent of Bs choosing *Out*. In total, $\rho \cdot c(1 - b \cdot \rho)$ percent of Bs choose *In* and the rest choose *Out*.

3) In the C-B treatment, similar to the Double treatment, $1 - c$ percent of Bs who do not receive informal contracts choose *Out*. The rest will choose *In* only if $20p \geq 10$. Since $p = 1$ with probability ρ , we have $\rho \cdot c$ percent of Bs choosing *In* and $1 - \rho c$ percent of Bs choosing *Out*, which is the greater than Double treatment. ■

The predictions under the HK model are indistinguishable from the GTF model in our game settings, and the reasoning behind the hypotheses is quite similar. In the GTF model, an honest type never lies, whereas an economic type has no cost of lying if the outcome from lying is preferred. Communication thus only has an effect on the behaviors of an honest type who chooses to send a message. As for player C, the changes in aggregate behavior only come from honest types who communicate. For player B, anticipating that some Cs would switch and choose *Left*, all Bs from the C-B treatment, all economic-type Bs, and some honest-type Bs who choose not to send a message from the Double treatment have the incentive to switch to *In*. And in both B-A and baseline treatment, there is no incentive for any type of Bs to deviate from choosing *Out*.

Following the analysis in section 2, Table 1 and 2 below summarizes hypothesis under different existing theories. The tables can be read as follows: the inequality (equal) sign represents the comparison outcome between the row treatment and the column treatment. For example, the equal sign on row 3 column 2 implies that the frequency of Bs choosing *Out* is expected to be the same between Baseline and B-A treatments. Notice that all theories except for EJ model with $l < 15$ predict the same treatment effects compared with Baseline for both players B and C. However, the predictions dif-

fer when we compare between treatments. Hypothesis 1–4 investigates the treatment effects compared with Baseline, and hypothesis 5–6 focuses on the between-treatment differences under which existing theories offer different predictions.

Table 1: Frequency of Bs Choosing *Out*: $\varphi_{B,t,Out}$

| | EJ model ($l < 15$) | | | | EJ model ($l \geq 15$)/GTF model/HK model | | | | CD model | | | |
|----------|-----------------------|-----|-----|--------|---|-----|-----|--------|----------|-----|-----|--------|
| | B-A | C-B | C-A | Double | B-A | C-B | C-A | Double | B-A | C-B | C-A | Double |
| Baseline | = | = | = | = | = | > | > | > | = | > | > | > |
| B-A | | = | = | = | | > | > | > | | > | > | > |
| C-B | | | = | = | | | = | > | | | = | = |
| C-A | | | | = | | | | > | | | | = |

Note: the inequality (equal) sign represents the comparison outcome between the row treatment and the column treatment.

Table 2: Frequency of Cs Choosing *Right*: $\varphi_{C,t,R}$

| | EJ model ($l < 15$) | | | | EJ model ($l \geq 15$)/GTF model/HK model/CD model | | | |
|----------|-----------------------|-----|-----|--------|--|-----|-----|--------|
| | B-A | C-B | C-A | Double | B-A | C-B | C-A | Double |
| Baseline | = | = | = | = | = | > | > | > |
| B-A | | = | = | = | | > | > | > |
| C-B | | | = | = | | | = | = |
| C-A | | | | = | | | | > |

Note: the inequality (equal) sign represents the comparison outcome between the row treatment and the column treatment.

A2. Bootstrap Analyses on Strategy Profiles Distribution

We use the bootstrap method to compare the frequency of Nash equilibrium strategy profile (*In*, *Out*, *Right*) and Pareto efficient strategy profile (*In*, *In*, *Left*) among different treatments.

We find that the B-A treatment has the highest frequency of Nash strategy profile ($p = .26$ compared with Baseline, $p = .00$ with C-B, $p = .06$ with Double, $p = .00$ with C-A), and on average, this strategy profile is played about half of the time. B-A is followed by Baseline ($p = .12$ compared with Double treatment, $p = .00$ with C-B, $p = .03$ with C-A) and the Double treatment ($p = .21$ compared with C-A treatment,

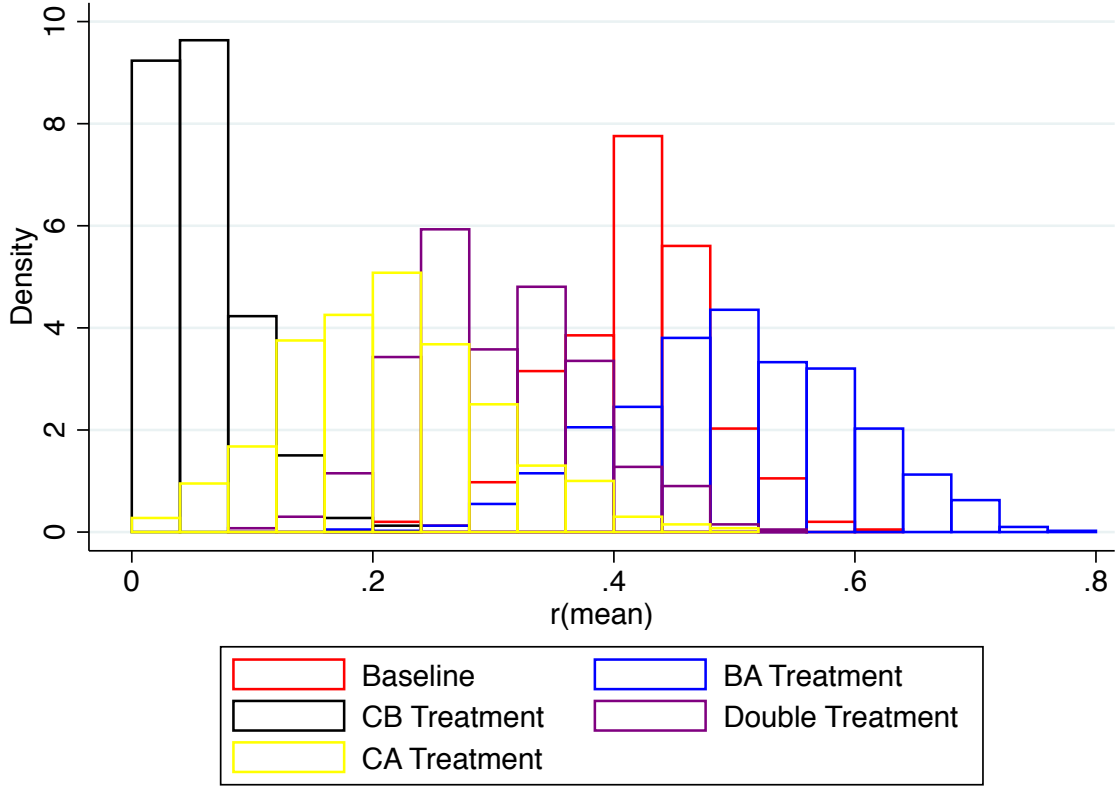


Figure 10: Bootstrap distribution of frequency of Nash strategy profile (*In*, *Out*, *Right*) by treatments (Resample $N = 999$)

$p = .00$ with C-B), while the C-B treatment has the lowest frequency of Nash play where the Nash profile is played around 4% of the time on average.

As for the frequency of Pareto efficient strategy profile, we observe that the C-B treatment has the highest frequency for (*In*, *In*, *Left*) strategy profile ($p = .00$ compared with Baseline, $p = .00$ with B-A, $p = .08$, with Double, $p = .02$ with C-A), and the Pareto efficient strategy is played 29% of the time on average. (*In*, *In*, *Left*) is played in the Double treatment about 14% of the time ($p = .19$, compared with C-A treatment, $p = .01$ compared with Baseline, $p = .00$ with B-A). The B-A treatment has the lowest frequency of Pareto efficient strategy profile, and it is almost never played.

Finally, we look into the expected efficiency across all treatments. Efficiency here is measured as the total payoff of players. There are four possible payoff outcomes: 20 (*Out*, *In/Out*, *Left/Right*), 30 (*In*, *Out*, *Left/Right*), 55 (*In*, *In*, *Left*), and 40 (*In*, *In*, *Right*). Table 3 below displays the average frequency of the payoffs being realized as well as the expected efficiency. The expected efficiency by treatment ranked from highest to lowest is as follows: C-B (35.93), Double (35.54), C-A (32.19), Baseline

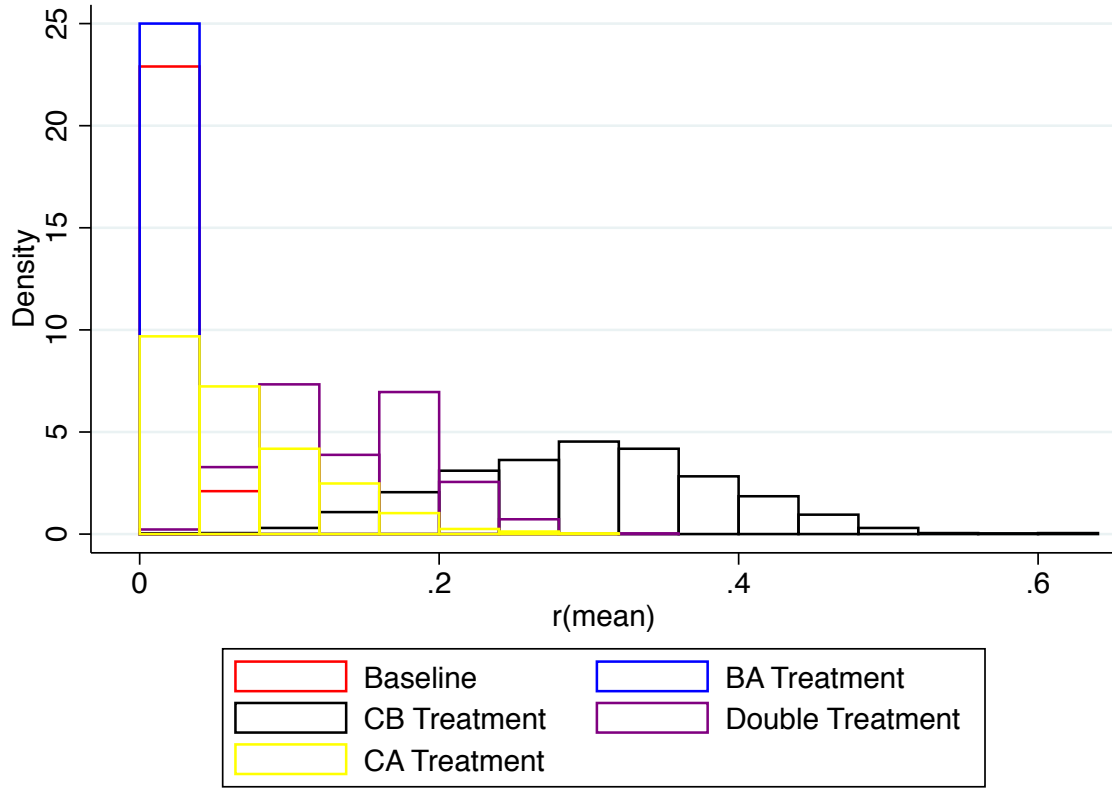


Figure 11: Bootstrap distribution of frequency of Pareto improving strategy profile $(In, In, Left)$ by treatments (resample $N = 999$)

(30.79) and B-A (29.64). Overall, C-B treatment has the highest level of expected efficiency, partly due to the high frequency of Pareto efficient strategy being played. All communication treatment outperformed the baseline, with the exception of B-A treatment.

Table 3: Bootstrap expected efficiency by treatment

| Average Probability (resample $N = 999$) | | | | | Expected Efficiency |
|---|-----------------------|--------------------|--------------------|-----|------------------------|
| <i>(Out, In/Out, L/R)</i> | <i>(In, Out, L/R)</i> | <i>(In, In, L)</i> | <i>(In, In, R)</i> | | |
| Baseline | 20% | 60% | 5% | 16% | 30.79 (0.03) |
| B-A | 16% | 71% | 0% | 13% | 29.64 (0.04) |
| C-B | 30% | 25% | 29% | 17% | 35.93 (0.09) |
| C-A | 26% | 37% | 7% | 30% | 32.19 (0.06) |
| Double | 5% | 56% | 14% | 26% | 35.54 (0.04) |

Note: standard error of the mean is included in the parentheses.