

Promises & Partnership

Gary Charness & Martin Dufwenberg*

December 28, 2003

Abstract: We examine, experimentally and theoretically, how communication may mitigate the problem (highlighted in contract theory) of hidden action. Communication dramatically enhances efficient partnership interaction. From the viewpoints of both classical theory and recent social-preference models, this makes for a puzzle. We propose two explanations: dislike of lying and guilt aversion. These explanations link, respectively, words and beliefs about beliefs to motivation, thereby providing the rudiments of a theory why communication matters. Our experimental design admits observation of promises, lies, and beliefs, and we report supporting results.

Keywords: Promises, partnership, contract theory, behavioral economics, hidden action, moral hazard, lies, social preferences, psychological game theory, guilt aversion, reciprocity, fairness

JEL codes: A13, B49, C72, C91, D63, D64, J41

We thank Jon Baron, Jeanette Brosig, Steve Burks, Colin Camerer, Ernst Fehr, Ayelet Fishbach, Guillaume Frechette, Dan Friedman, Drew Fudenberg, Simon Gächter, Uri Gneezy, Brit Grosskopf, David Laibson, Dan Levin, David I. Levine, David K. Levine, Tanya Menon, Matt Parrett, Torsten Persson, David Reiley, Yuval Rottenstreich, David Strömberg, Richard Thaler, Bernd Wittenbrink, Bill Zame, and seminar participants in Antwerp, Berlin (Humboldt), Bonn, Chicago (GSB), Cologne, Dortmund, Gothenburg, Great Barrington (Behavioral Research Council), Harvard, Munich, Örebro, Oslo, Purdue, Santa Barbara, Stockholm, University of Arizona, Wittenberg (GEW-Tagung), and Tampa (Public Choice) for stimulating discussions and helpful suggestions, and the Swedish Competition Authority and the Russell Sage Foundation for financial support.

* **Contact:** Gary Charness, University of California at Santa Barbara, charness@econ.ucsb.edu; Martin Dufwenberg, University of Arizona, martind@eller.arizona.edu.

1. INTRODUCTION

Much of human achievement is produced in partnerships. An extensive body of theoretical research—contract theory—is devoted to understanding which partnerships form, what contracts are signed, what the economic consequences will be. The insights offered pinpoint subtle ways that informational asymmetries, legal issues, and other considerations may matter. Considerable attention has been devoted to environments with *hidden action*, where a contracting party's future choice cannot be regulated by a contract.¹ Contract theorists have shown that if people are rational and selfish (caring only about own income), hidden action is a shoal on which efficient contracting may founder.

We investigate experimentally whether non-binding pre-play communication (*cheap talk*) can be effective in achieving better outcomes in a contract-theoretic one-shot environment. The conventional approach implicitly assumes that such communication is ineffective in promoting partnership formation and cooperation; written contracts bind if supported by the law, but oral agreements (to quote Samuel Goldwyn) "aren't worth the paper they're written on". We feel that this view is at odds with reality, where promises, discussions, handshakes, threats, and other forms of communication are often used when agreements are made. These casual observations justify a suspicion that communication can foster trust and cooperation in settings with hidden action, in contrast to the prediction of conventional contract theory.

Let us clarify: We examine a setting that involves *no repeated play* and in which a conventional approach yields a *unique prediction regardless of whether pre-play cheap talk is incorporated*. We do not consider repeated games (cf, for example, MacLeod & Malcolmson 1989), in which communication may serve as an equilibrium-selection device. Our study explores whether there may be other important psychological aspects related to communication that conventional contract theory has failed to address.

¹ This condition is often referred to as *moral hazard*. For entries to this literature, see e.g. Hart & Holmström (1986), Dutta & Radner (1993) and Salanié (1998, chapter 5).

The general idea that communication can affect strategic interaction in one-shot play is not new.² However, we believe that our tack is novel in many ways. First, the strategic setting we consider is motivated from a contract-theoretic perspective, and we tease out an explicit hidden-action interpretation. Second, we are led to consider a simple ‘trust game’, with a twist involving a move by nature. Our study may be the first to consider the impact of communication in a trust game; previous work has for the most considered prisoners’ dilemmas, coordination games, or bargaining games. Third, and perhaps most importantly, we incorporate a new theoretical perspective, in part building on psychological game theory (see Geanakoplos, Pearce & Stacchetti 1989). We measure beliefs and test a model with belief-dependent utility, and we record messages and examine how ‘statements of intent’ correlate with subsequent choices. All in all, we eventually provide the rudiments of a new theory of why communication matters for overcoming the problems caused by hidden action.

In the experiment, we find that successful cooperation, contrary to the prediction of classical contract theory, is substantial even without communication. This is hardly surprising, given the wealth of experimental studies documenting that subjects seem not to maximize their own income in many games. However, we also find that communication affects behavior quite dramatically, effectively doubling the rate of cooperative, successful outcomes.

We then ask *why* these results obtain. If contract theory needs revision, it is imperative to understand the motivational forces that drive behavior, and why communication may matter in this connection. Such insights are crucial for designing ‘behaviorally-optimal’ contracts, and for deriving testable predictions that allow exploration of how far the results extend to other economic contexts. We examine the predictions of recent models of *social preferences*, which

² Some work is in psychology. Dawes, McTavish & Shaklee (1977) find that ‘relevant’ face-to-face communication leads to more cooperation in a commons dilemma. Orbell, Dawes & van de Kragt (1990) find that promises enhance cooperation if everyone in a game makes them. Communication impact has been discussed concerning social contracts (Blau 1964), psychological contracts (Rousseau 1995), and social norms (Bicchieri 2002). In experimental economics, communication has been found to improve coordination (e.g., Cooper, DeJong, Forsythe & Ross 1990, 1992; Charness 2000) and to matter in bargaining (Valley, Moag & Bazerman 1998; Bohnet & Frey 1999; Ellingsen & Johannesson 2002; Brandts & Charness 2003; Brosig, Ockenfels & Weimann 2003).

make reference to various notions of fairness and reciprocity, regarding behavior in our experimental environment.³ Previous work has not focused on what these models imply regarding the impact of communication. We show that communication should not matter for either *distributional models*, in which decision-makers' preferences depend only the overall distribution of monetary earnings among the reference group, or for belief-dependent models of *kindness-based reciprocity*. These models deliver unique equilibrium predictions for our settings, and therefore cannot allow pre-play cheap talk to matter.

These negative findings lead us to propose, and experimentally test, two alternative models which permit communication to play a role. First, we consider the straightforward idea that people *dislike lying*, which is consistent with recent experimental work by Gneezy (2002). Our experimental design permits an agent (second mover) to send a free-form written message to the principal (first mover) with whom he or she is paired, and we examine how the occurrence of 'promises' (or at least 'statements of intent') correlate with the frequency of trustworthy choices. We find such a connection.

Second, we consider the psychological-games-based explanation of *guilt aversion*: an agent dislikes acting in an untrustworthy fashion to the extent that he or she believes that the principal expects a trustworthy response. Like reciprocity, guilt aversion is a belief-dependent motivation, but (unlike reciprocity) guilt aversion admits multiple equilibria. This creates room for communication to matter; words may influence beliefs, beliefs may influence motivation, motivation permeates the nature of equilibrium play. The nature of guilt-averse utility suggests an experimental test which requires observation of a particular belief about a belief of an agent. A key feature of our design allows us to elicit this second-order belief in an incentive-compatible manner. We find that our data is consistent with guilt aversion.

³ For descriptions of the experimental evidence and the social-preferences literature, see Fehr & Gächter (2000), Sobel (2001), Fehr & Schmidt (2002), and Cox (2004).

Which has better predictive value, guilt aversion or dislike for lying? We point out that the two may both be relevant and may reinforce each other. A sensible theory of the impact of communication may draw on both ideas.

Our paper may be seen as a contribution to a field that may be labeled ‘behavioral contract theory’. Loewenstein (1999) defines behavioral economics as bringing “psychological insights to bear on economic phenomena”. The field of behavioral contract theory should be seen as a sub-field of behavioral economics, wherein one takes into account social and psychological considerations in an attempt to understand partnerships and contracts. Two approaches to behavioral contract theory may be distinguished. First, one might explore (experimentally or theoretically) which contract people choose when there are *many feasible ones*.⁴ Second, one might consider *one specific contractual arrangement* and attempt to understand that environment in some depth; the current paper belongs here. We explore whether and why a given contract is acceptable to two parties, with and without communication. We make no presumption that the contract we look at is ‘optimal’—that would be impossible since we enter the analysis being open-minded with respect to what motivational forces are at work. The objective is to reveal insights about decision-making and motivation that are useful for further developing behavioral contract theory.

Our presentation alternates between theory (section 2 and part of section 4) and experimental evidence (sections 3 and part of section 4). Section 5 concludes.

2. A PARTNERSHIP MODEL AND THE HIDDEN-ACTION GAMES

In this section we present a simple model of a partnership (2.1), explain why this model is fit to address aspects of hidden action (2.2), and finally derive the game that will be experimentally tested, with and without communication (2.3, 2.4). The partnership model

⁴ Examples of this approach include Anderhub, Gächter & Königstein (2002), Cabrales & Charness (2000), Fehr, Klein & Schmidt (2001), Güth, Klose, Königstein & Schwalbach (1998).

highlights the contract-theoretic backdrop to our design, but may also have independent value by providing a framework for other experimental studies. Although we elicit a mono-contractual setting incorporating hidden action, the model could alternatively be used to consider hidden information (adverse selection), multiple contracts, or richer strategic settings.

2.1 A Nash bargaining benchmark

A principal and an agent consider forming a partnership in which a project is carried out. If no partnership is formed, then no contract is signed, no project is carried out, and the parties get their outside-option payoffs x for the principal and y for the agent (measured in dollars). If the project is carried out, then the contract specifies a ‘wage’ w that the principal must pay to the agent. The project generates revenue for the principal. There can be two outcomes: poor or good. A poor outcome generates revenue $r > 0$, while a good outcome involves an additional bonus of $b > 0$, so that total revenue is $r+b$. The probability of these outcomes depends on the choice and characteristic of the agent; we assume the agent chooses ‘effort’, $e \in [0,1]$, and has a given ‘talent’, $t \in [0,1]$, and that the probability of a good outcome is et .⁵ The agent experiences increasing ‘effort cost’, measured in dollars and equal to ce , where $c > 0$.

In order to derive a benchmark, consider the Nash bargaining solution for risk-neutral and selfish players, assuming that effort and wage is contractible and that all the other parameters are commonly known to the two parties. Suppose the project is carried out, the agent's talent is t , the principal pays the wage w to the agent, and the agent chooses effort e . Following Nash (1950), one sees that the solution will be the wage-effort combination (w,e) that maximizes

$$[(r-w+etb)-x] \cdot [(w-ce)-y] \tag{1}$$

⁵ This specification is inspired by Dufwenberg & Lundholm's (2000) model; see their Sect. 1.1 and Figure 1.

whenever it is possible to choose (w,e) so that each factor of this product is positive (otherwise, no partnership would form). In this paper, we assume that the agent's talent is high enough and the cost of effort low enough that the expected return from exerting effort exceeds the cost of the effort. This requires that $etb > ce$ if $e > 0$, or equivalently $t > c/b$. We also assume that it is possible to choose (w,e) so that each factor of (1) is positive. A sufficient condition is that $r+tb-c > x+y$. The Nash bargaining solution is then given by (2) & (3):

$$w^* = (r + tb - x + y + c)/2 \quad (2)$$

$$e^* = 1 \quad (3)$$

2.2 Hidden action and the chance move

The reader may wonder why we have bothered to include in the above (otherwise rather spare) model a move by nature that determines the success of the project, rather than just replace that move with its expected outcome. We have done so in order to prepare the grounds for consideration of circumstances where the contract cannot be conditioned on the agent's choice of effort. A typical justification for such contractual limits, often stressed by contract theorists, is that the agent's effort is not observable to the principal, or at least to third parties. Thus contractual clauses about effort choices are not enforceable in court (see Holmström, 1979).

If, however, outcomes were perfectly correlated with the effort choice, then the agent's choice could nevertheless be *inferred* with certainty, and thus (arguably) be enforceable in court. The move by nature is then essential for making conceptual sense of our exercise. With this move, if a project fails due to low effort, the agent can *claim* that he exerted high effort but that he had bad luck. The chance move ensures that it cannot be proven in court that he lied, once effort is not directly observable by third parties. This is the essence of hidden action.

2.3 A hidden-action game

A major issue in modern contract theory is the choice of contract when a partnership is influenced by hidden action. The assumption is typically maintained that the principal and the agent are perfectly selfish, and an optimal contract is derived based on that premise. Our goal is to incorporate hidden action to the model of section 2.1, but we will *not* examine which contract out of many feasible ones would be agreed upon given a particular motivation. Rather we stay open-minded with respect to the nature of the motivation and examine, *for a given contract*, how serious the problems caused by hidden action are in the first place. That contract corresponds to the above Nash bargaining benchmark, for the following specific set of parameters:

$$\begin{array}{lll} r = 14 & c = 4 & x = 5 \\ b = 12 & t = 5/6 & y = 5 \end{array} \quad (4)$$

If both wage and effort were contractible, using (2) and (3), we would get $(w^*, e^*) = (14, 1)$. However, we will now remove the assumption that the effort can be regulated in the contract and instead assume that it remains for the agent to *choose* his effort level; this incorporates hidden action. We restrict the agent to two possibilities: $e \in \{0, 1\}$. Will the outcome corresponding to the Nash bargaining solution still obtain?

There are two basic reasons why it may not. First, the agent may choose $e = 0$ instead of $e = 1$, keeping the contractual wage $w^* = 14$ while opportunistically saving himself the effort cost. Second, the principal may foresee such a turn of events, dislike it, and not agree to form a partnership. The following extensive-form game incorporates these two possibilities:

FIGURE 1

First, the principal decides whether to say *Yes* or *No* to the contract according to which he must pay $w^* = 14$ to the agent. If he says *Yes*, then a partnership is formed and the project is

carried out, but it remains for the agent to choose his effort level $e = 0$ or $e = 1$. The payoffs are derived using (4) and the assumptions spelt out in section 2.1.⁶ The game has a ‘trust structure’ with the added twist of the chance move, which we view as conceptually crucial given the underlying contract-theoretic interpretation (cf. section 2.2).

If the players are selfish the game has an obvious (backward induction) solution: If called upon to play, the agent would exert low effort. The principal's best response is to choose *No*, not agreeing to form a partnership. This outcome is inefficient, since both parties receive a higher expected payoff with the (*Yes*, $e = 1$) outcome than when no partnership is formed; this is a simple illustration of how conditions of hidden action may undermine efficient contracting. However, in light of previous work on trust games (e.g. Berg, Dickhaut & McCabe 1995, Dufwenberg & Gneezy 2000), it is natural to expect play to sometimes differ from the backward induction solution for selfish players. The novelty of our approach, beyond proposing a related test, and beyond making explicit links to a contract-theoretic background model, is to examine the role of communication in this context. We discuss this next.

2.4 Incorporating communication

Our experiment is built around the game in Figure 1, but we consider two treatments that differ according to whether or not a communication opportunity is present. In the *no-communication treatment* the experimental design corresponds directly to the game in Figure 1. In the *communication treatment* we consider an augmented version, with an added preceding communication stage where the agent may transmit a message to the principal. The agent might, for example, promise to exert high effort. Figure 2 portrays the resulting situation.

⁶ In case the principal chooses *No*, no partnership is formed and each of the prospective partners earns her or his outside option of 5 ($= x = y$). In case the principal chooses *Yes* and the agent chooses $e = 0$ the project will fail ($et = 0 \cdot 5/6 = 0$), so the principal gets $r - w^* = 14 - 14 = 0$ and the agent gets $w^* - ce = 14 - 4 \cdot 0 = 14$. In case the principal chooses *Yes* and the agent chooses $e = 1$ the outcome corresponds exactly to the Nash bargaining solution given (4).

FIGURE 2

If the players are selfish the presence of the communication stage has no impact on the analysis. The above argument, leading to the (*Yes*, $e = 1$) outcome in the game of Figure 1, goes through unchanged for the post-communication subgame in Figure 2. Hence, with selfish players, again the theoretical prediction is that no partnerships are formed and that the outcomes will be inefficient. On the other hand, if other concerns than selfishness are motivating the players, perhaps communication will matter. This is what we will explore.

3. THE EXPERIMENT

In this section we present our experimental design (3.1), state two initial hypotheses we test (3.2), and give the related results (3.3). Further hypotheses and results appear in section 4.

3.1 Design

Our experimental design corresponds to the games of Figures 1 and 2. The instructions did not refer to game trees, and we used different labels than in section 2 for the players and actions (cf. Appendix A). A subject in the principal's position was referred to as person 'A'; a subject in the agent's position was referred to as person 'B'. Instead of *Yes* and *No*, the subjects chose IN or OUT. Instead of $e = 1$ and $e = 0$, subjects chose to ROLL or DON'T ROLL a 6-sided die. We use a one-shot design to avoid potential reputation and supergame issues.

The game was described using the following chart, presented to each of the participants:

	A receives	B receives
A chooses OUT	\$5	\$5
A chooses IN, B chooses DON'T ROLL	\$0	\$14
A chooses IN, B chooses ROLL, die = 1	\$0	\$10
A chooses IN, B chooses ROLL, die = 2,3,4,5, or 6	\$12	\$10

The treatment without messages began with each A choosing IN or OUT. Next, each B indicated (without knowing A's actual choice) whether he or she wished to choose ROLL or

DON'T ROLL, *contingent upon A having chosen IN*, as B's choice is immaterial if A has chosen OUT. We thus obtain an observation for every B.⁷ The outcome corresponding to a successful project occurred if and only if the die came up 2, 3, 4, 5, or 6 after a ROLL choice.

In the message treatment, each B had an option to send a non-binding message to A prior to A's decision concerning IN or OUT. Each B received a sheet, on which any (non-identifying) message could be written, if desired. B could also decline to send a message by circling the letter B at the top of the otherwise-blank sheet.⁸

Participants were recruited at UCSB by sending out an e-mail message to the campus community. We conducted six sessions, three where messages were feasible and three where they were not. Sessions were conducted in a large classroom that was divided into two sides by a center aisle, and people were seated at spaced intervals. The number of participants in a session ranged from 24 to 36, with 90 people in the sessions without communication and 84 people in the sessions with communication; each person could only participate in one of these sessions. Average earnings were \$16, including a \$5 show-up fee; each session was one hour in duration.

A coin was tossed to determine which side of the room was A (principal) and which side was B (agent). Identification numbers were shuffled and passed out face down, and participants were informed that these numbers would be used to determine pairings (one A with one B) and to track their decisions. After answering questions, the experimenter chose individuals at random to state the outcome for all possible cases, starting when it seemed clear that everyone understood the rules. After the decisions had been collected, a 6-sided die was rolled for each agent; this was made clear to the participants in advance, to avoid the anticipated loss of public anonymity for agents who chose DON'T ROLL. This roll was determinative if and only if (IN, ROLL) had been chosen.

⁷ Although somewhat controversial, this *strategy method* (Selten 1967) is used extensively in experimental economics and may be best suited to games with few decision nodes.

⁸ The law distinguishes between written (legally binding) and verbal contracts; this might be relevant for certain written messages, like promises. In the experiment all written statements were *non-binding*, thus in the (legal) spirit of verbal promises. It was not feasible to have verbal promises without sacrificing anonymity.

Our design also incorporated a belief elicitation scheme, to be discussed in section 4.3.

3.2 Two initial hypotheses

If an agent has purely selfish preferences, it is a dominant strategy to choose zero effort when effort is not observable. Anticipating this, the principal will reject the contract. Our first hypothesis is that this simple pattern of behavior will occur universally:

H1: *No agent chooses high effort. No principal ever accepts the contract.*

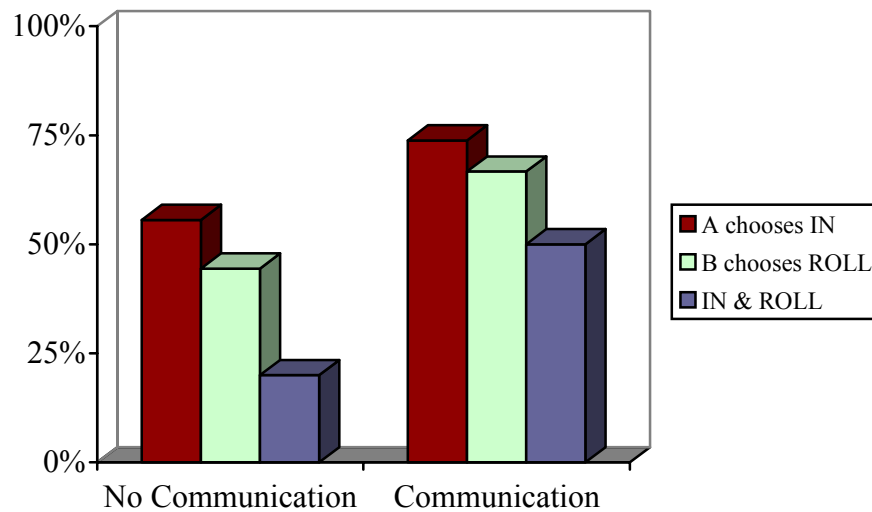
In accordance with findings in many experimental studies, where many people seem not to maximize own money, it is not unreasonable to expect that H1 might be rejected. In that case, it is an open question whether or not communication can move behavior. We test:

H2: *The possibility of communication will increase neither the proportion of principals accepting contracts nor the proportion of agents who choose high effort. (IN, ROLL) outcomes are not more common when messages are feasible.*

3.3 Results on H1 and H2

Figure 3 summarizes the overall choices of A's and B's. In the no-communication treatment, 20/45 (44%) of B's choose to ROLL; this compares to 28/42 (67%) in the communication treatment. A's were more likely to choose IN in the communication treatment, 31/42 (74%) to 25/45 (56%). The (IN, ROLL) choice occurred 20% of the time (9 of 45 pairs) without communication, compared to 50% (21 of 42 pairs) with communication.

Figure 3 - Observed Choices



H1: This is the classical hypothesis; we reject it for both the no-communication and communication treatments, for both principals and agents. For the principals, the test of the difference of proportions (see Glasnapp & Poggio, 1985) gives $Z = 5.88$ and 7.01 for the respective treatments, with $p < 0.000001$; for the agents, this test gives $Z = 5.07$ and 6.48 , with similarly high statistical significance. Own monetary reward is not the only motivation present.

H2: This hypothesis states that the possibility of communication will not affect behavior. We can reject it for both principals and agents. In the no-communication treatment, 55.6% A's chose IN, while in the communication treatment this proportion increased to 73.8%. The test of the difference of proportions gives a test statistic of $Z = 1.78$, with significance at $p = 0.038$.⁹ Without communication 44.4% B's chose ROLL, while with communication this proportion increased by half to 66.7%. The test of the difference of proportions gives a test statistic of $Z = 2.08$, with significance at $p = 0.019$. The likelihood of a successful partnership more than

⁹ Except where otherwise indicated, we use one-tailed tests to reflect our *ex ante* directional hypotheses.

doubles, and the test of proportions gives $Z = 2.94$, $p = 0.002$. We may conclude that messages have a major influence on behavior and outcomes.

4. UNDERSTANDING THE IMPACT OF COMMUNICATION

Our rejection of H2 suggests that communication may help foster trust and cooperation so that problems of hidden action can be partly overcome. It remains to explain *why* this occurs.

In order to apply the non-self-interested behavior observed in experiments to economic settings such as consumer response to price changes or employee response to changes in wages and employment practices, researchers have begun to develop formal models of social preferences that assume people are not solely selfish, but also care in some way about others. For descriptions of the experimental evidence and the ‘social-preferences’ literature it has inspired, see Fehr & Gächter (2000), Fehr & Schmidt (2002), and Sobel (2001). To a person familiar with this literature, it will come as no surprise that these models can explain why problems of hidden action may be overcome. However, previous work has not focused on what these models imply regarding the impact of communication. We extend the analysis in this direction.

It turns out that certain social-preference models (viz. models of *inequity aversion*, and models of *kindness-based reciprocity*) that have received a lot of attention do not effectively capture the impact of communication. We exhibit this insight in section 4.1.

This prompts us, in section 4.2, to introduce two alternative motivational forces that can potentially explain the impact of communication: *dislike of lying* and *guilt aversion*. These ideas have for the most part not been considered in the social-preferences literature. We develop some further experimental tests that can shed light on the predictive success of dislike of lying and guilt aversion, and we report the results in section 4.3.

4.1 Models that fail to capture the impact of communication

Inequity aversion

An important class of social-preference models defines preferences with reference to the final monetary distribution of payoffs. We will focus in particular on models of *inequity aversion*, as presented in Fehr & Schmidt (1999) and Bolton & Ockenfels (2000), in which people are presumed to like money but also to dislike disparities in payoffs.¹⁰ Models of inequity aversion have proven quite successful in explaining a variety of experimental data; applied to our games, they allow the agent to prefer $e = 1$ to $e = 0$, if he is sufficiently averse to getting a higher payoff than the principal.

The inequity-aversion models are non-standard in that decision makers may care for overall distributions beyond their personal gain. Unlike some competing models (to be discussed) they abstract away from ways that decision makers may care about the *process* leading up to a final allocation. The inequity-aversion models have the great virtue of being simple to apply and to test, and it is therefore important to assess their range of applicability. We check one particular aspect: Can models of inequity aversion capture how communication influences strategic interaction in our setting with hidden action?

The answer is *no*. To explain why, we exemplify using the Fehr & Schmidt model. Consider any two-player game in which the players with $i, j = 1, 2, i \neq j$, get (expected) monetary payoffs m_i and m_j . The Fehr/Schmidt utility of player i is given by

$$m_i - \alpha_i \cdot \max\{m_j - m_i, 0\} - \beta_i \cdot \max\{m_i - m_j, 0\} \quad (5)$$

where β_i and α_i are parameters satisfying $\beta_i \leq \alpha_i$ and $0 \leq \beta_i < 1$. The three terms in (5) capture how the agent cares for own income, how he suffers from disadvantageous inequality, and how

¹⁰ Nevertheless, our conclusions here stand for any model with purely distributional preferences (e.g., the social-welfare preferences in the basic (non-reciprocity) form of the Charness & Rabin, 2002 model).

he suffers from advantageous inequality. We can apply (5) to incorporate inequity aversion to the games in Figures 1 and 2. Given any collection of parameters α_i, β_i , the resulting games look just like those in Figures 1 and 2 except that the payoffs are transformed. These new games are solvable by backward induction. For example, with the agent as player 2, if $\beta_2 > 2/7$ the backward induction solution will be (*Yes*, $e = 1$).¹¹ The logic of the argument does not depend on whether or not we work with Figures 1 or Figure 2, however. This illustrates that although models of inequity aversion are capable of explaining how a problem of hidden action is overcome, they also predict that *communication does not matter*.¹²

Kindness-based Reciprocity

We next consider models in which decision-makers wish to be kind to those they believe to be kind, and to be unkind to those they believe to be unkind. The classic reference for such a model of *kindness-based reciprocity* is Rabin (1993). His notion of kindness takes into account what a person believes he accomplishes; his notion of believed kindness takes into account what a person believes about the kindness of others. Intentions, and intentions ascribed to others, matter, and preferences end up being complicated to describe, as they depend on beliefs and beliefs about beliefs. The analysis requires the toolbox of psychological game theory (Geanakoplos et al, 1989), which extends standard game theory to allow for such considerations.

Rabin's model is developed for games represented in normal form, and seems meant to highlight the key qualitative features of reciprocity. As Rabin notes (p. 1296), it is not well-suited for applied work involving games that have (as here) a non-trivial dynamic structure.

¹¹ The derivation is as follows: Using (5), one sees that with agent will choose $e=1$ if $10-\alpha_2 \cdot \max\{10-10, 0\} - \beta_2 \cdot \max\{10-10, 0\} > 14-\alpha_2 \cdot \max\{0-14, 0\} - \beta_2 \cdot \max\{14-0, 0\}$, which is equivalent to $\beta_2 > 2/7$.

¹² Three comments may be warranted: (i) We do not claim that models of inequity aversion do not permit communication to matter in all other games. In games with multiple equilibria, communication may play a role by facilitating coordination. (ii) Communication may furthermore matter if the *degree* of inequity aversion is made a function of whether there is communication. We do not think much of this idea; the key idea of distributional models is that one need *only* make reference to distributions. Assuming communication-sensitive inequity aversion destroys that virtue. (iii) If there is incomplete information about the players' degree of inequity aversion (which may seem reasonable, and is in fact assumed in Fehr et al., 2002), then a principal's behavior could be affected by a message that may signal the agent's 'type'. However, this could never affect the agent's behavior in the subgame.

Dufwenberg & Kirchsteiger (forthcoming) (DK) therefore develop a model of reciprocity for extensive-form games,¹³ and we shall rely on that model to derive a prediction for our games.

Consider any two-player game in which the players get (expected) monetary payoffs m_i and m_j , with $i, j = 1, 2$, $i \neq j$. The DK utility of player i is given by

$$m_i + Y_i \cdot \kappa_{ij} \cdot \lambda_{iji} \quad (6)$$

The two terms in (6) capture how the agent cares for own income, and how he is motivated by reciprocity. Y_i is a non-negative parameter describing i 's sensitivity to reciprocity. κ_{ij} represents i 's kindness to j ; κ_{ij} is positive if i is kind, and negative if i is unkind. λ_{iji} represents i 's belief about how kind j is to i ; this parameter is positive if i believes j is kind to i , and negative if i believes j is unkind to i . The specification captures reciprocity by making it in i 's interest to make the sign of κ_{ij} match the sign of λ_{iji} .

κ_{ij} and λ_{iji} may depend on player i 's beliefs. Consider for example the kindness of the principal when he chooses *Yes*. The choice of *Yes* brings about a higher payoff to the agent than the choice of *No*, so *Yes* is kinder than *No*. However, the exact degree of kindness of *Yes* depends on the principal's beliefs. Let τ denote the probability with which the agent chooses $e = 1$; let τ' denote the principal's expectation of τ ; let τ'' denote the agent's expectation of τ' (everything measured at the agent's decision node). The principal's kindness when choosing *Yes* depends negatively on τ' . The logic is that the lower is τ' , the more money the principal believes she believes she gives to the agent, so the kinder she is. Similarly, the agent's beliefs about the principal's kindness depends negatively on τ'' ; the principal's kindness depends negatively on τ' , so the lower the agent's belief of τ' , the kinder the agent believes the principal is.

¹³ Segal & Sobel (1999) present another reciprocity model. There are also models that combine distributional preferences and reciprocity; see Falk & Fischbacher (1998), Charness & Rabin (2002), Cox & Friedman (2002).

DK develop a solution concept called *Sequential Reciprocity Equilibrium (SRE)*, which incorporates a sequential rationality requirement. In many games the equilibrium set is large, with different equilibria supported by very different beliefs. One may reasonably suspect that the theory allows communication to matter in the games of Figures 2 and 3, by admitting different equilibria with and without communication. However, this turns out *not* to be the case; with the agent as player 2, the DK equilibrium depends on Y_2 and is (essentially) *unique*. For low values of Y_2 the SRE is (*No*, $e = 0$); for high values of Y_2 the SRE is (*Yes*, $e = 1$); for intermediate values of Y_2 some mixing is predicted. See Appendix C for the exact derivation.

The equilibrium analysis does not depend on whether or not one works with Figure 1 or Figure 2. This illustrates that although kindness-based reciprocity can explain how a problem of hidden action is overcome, and although the model incorporates belief-dependent utilities, the model nevertheless predicts that, in equilibrium, *communication does not matter*.

4.2 Models that succeed in capturing the impact of communication

In light of the findings of section 4.1, we next propose two alternative models that permit communication to play a role.

Dislike of Lying

One way to rationalize our finding that cooperation enhances cooperation is to assume that people do not like to mislead others. The idea is supported by the Gneezy (2002) finding that people experience a cost of lying and results by Brandts & Charness (2003) indicating that people dislike deception *per se*.

The idea can be modeled in a simple way, by allowing the agent to choose whether or not to make a promise, and to thereby select which one of two different subgames should be entered. If no promise is made, the subgame looks exactly as the game in Figure 1. If a promise is made, the subgame looks exactly as *the game in Figure 1 except that a cost of lying is deducted following the profile (Yes, $e = 0$)*. If the cost of lying is high enough, the unique

subgame-perfect equilibrium will see the agent issuing a promise, thereby creating a credible commitment to choose $e = 1$ so that the principal's best response is to choose *Yes*.

Guilt Aversion

A different idea is that decision-makers suffer from guilt if they believe they let others down. Such persons, whom we label 'guilt averse', shy away from such choices. Guilt aversion has not received much attention by scholars working on social preferences, but related ideas appear in some applied theoretical work by Huang & Wu (1994) (on remorse in corruption) and by Dufwenberg (2002) (on guilt in marriage) and in experimental work by Dufwenberg & Gneezy (2000), Bacharach, Guerra & Zizzo (2002), and Charness & Rabin (2001). Related ideas appear also in social psychology (although no mathematical modeling is done). See, *e.g.*, Baumeister, Stillwell & Heatherton (1994, 1995) who (on the basis of autobiographical narratives) suggests that people suffer from 'guilt' if they inflict harm on others. One prominent way to inflict harm is to let others down. In the words of Baumeister *et al.* (1995, p. 173): "Feeling guilty [is] associated with ... recognizing how a relationship partner's standards and expectations differ from one's own". It is such a sentiment we now propose to model.

Simple as the idea may seem, it calls for non-standard modeling. Consider the following example, chosen to illustrate this in the simplest possible way, rather than because of its immediate relevance to contract theory:

Björn feels guilty if he lets others down. In restaurants, this influences his tipping. If the waitress does a decent job, the more he believes that the waitress believes she will receive as a tip, the more he will tip. Björn gives just as much as he believes his waitress believes she will get, in order to avoid the feelings of guilt that will plague him if he gives less. (When Björn goes abroad, he inquires at the airport about 'tipping customs'.)

Perhaps surprisingly, conventional game theory cannot model Björn's motivation and behavior. Consider a standard game where Björn (player 1) chooses a tip, and the waitress (player 2) has no choice (her strategy set is modeled as a singleton). Björn's choice of tip

determines a full strategy profile. In game theory, payoffs are defined only on strategy profiles, so Björn's set of best choices must be independent of his belief of the waitress's belief. This contradicts the example. Hence, Björn's preferences cannot be described using conventional game theory; as with kindness-based reciprocity, psychological game theory must be used.

To make this clear, let $t \geq 0$ denote Björn's tip, t' denote the waitress's belief (her expectation of t) of the tip Björn will give her, and t'' denote Björn's belief (expectation) of t' . The assumption of guilt from letting the waitress down can be modeled such that Björn's utility following the strategy profile where he chooses tip t and believes t'' is

$$-t - \gamma \cdot \max\{t'' - t, 0\},$$

where $\gamma > 0$ is a constant measuring Björn's sensitivity to guilt. Björn's utility depends negatively on how much money he gives away, and on the extent to which he believes he does not live up to the waitress' expectations. We make several observations: (i) This leads to a psychological game rather than a standard game; although t determines a strategy profile, the payoff $-t - \gamma \cdot \max\{t'' - t, 0\}$ is not a number but rather a function of t'' . (ii) If $\gamma > 1$, Björn's optimal choice would be $t = t''$. (iii) Incorporating guilt this way implies a positive correlation between t and t'' . (iv) A hypothesis corresponding to (iii) can be tested with experimental data on t and t'' .

The guilt-aversion concept can be applied in much the same way to the games in Figures 1 and 2. Let τ , τ' , and τ'' be defined as in the case of kindness-based reciprocity. Guilt aversion can now be captured by assuming that the agent's utility following his choice $e = 0$ is decreasing in τ'' . A psychological game results, as depicted in Figure 3:

FIGURE 3

The parameter $\gamma > 0$ again scales the agent's guilt sensitivity. If γ is large enough, and if the agent believes sufficiently strongly that the principal believes he will choose $e = 1$ (i.e., if τ'' is large enough), then $14 - \gamma\tau'' < (5/6) \cdot 12 + (1/6) \cdot 0 = 10$ holds, so the agent will choose $e = 1$.

In this connection, the potential role of communication is remarkable: it may be that communication influences beliefs, beliefs influence motivation, and motivation influences behavior. For example, suppose the agent says “I promise to choose $e = 1$ ”. If the agent believes that the principal believes him, this will make the agent *more* inclined to choose $e = 1$. This in turn gives the principal a reason to believe the agent's statement. For an agent capable of feeling guilty, truth-telling is self-enforcing. By issuing a promise, the agent can gain commitment power regarding the exercise of his choice $e = 1$.¹⁴

To give these ideas formal substance we need to develop a solution concept that can capture guilt aversion in the psychological game of Figure 3. To this end, note first that *given* τ'' , the psychological game of Figure 3 has real numbers characterizing payoffs at each end node. In this sense, it reduces to a standard game. Call this game $\Gamma(\tau'')$. An equilibrium solution must fulfill two requirements: First, the players must optimize at all decision nodes given their beliefs and choices. Second, the beliefs must be consistent with what is actually happening. The following notion, in which (σ, τ) denotes the strategy profile in which the principal chooses *Yes* with probability σ and the agent chooses $e = 1$ with probability τ , imposes these requirements.¹⁵

DEFINITION: The profile (σ, τ) , together with beliefs τ' and τ'' , is a *guilt-aversion equilibrium* if

- $\tau'' = \tau' = \tau$
- (σ, τ) is a subgame-perfect equilibrium in the standard game $\Gamma(\tau'')$

¹⁴ We focus on communication by the agent, but note that with an agent prone to feeling guilty the principal also has an incentive to indicate that he has high expectations. Similarly, in the tipping example, the waitress or the restaurant owners may have analogous incentives. For example, if Björn visited the *Crab House* restaurant at Pier 39 in San Francisco, his waitress would give him a plastic card which reads (in six languages): “Thank you for dining with us. Many guests ask us about tipping. We want you to know that no additional tip or service charge has been added to your bill. In the United States, quality service is rewarded with a tip, or gratuity, of at least 15%.”

¹⁵ The solution concept is similar to Geanakoplos et al.’s notion of “subgame-perfect psychological equilibrium”. It differs in that τ'' is recorded at the node where the agent moves, while Geanakoplos et al only allow initial beliefs (before the principal moves) to influence payoffs. Moreover, Geanakoplos et al. impose explicitly that equilibrium profiles be common knowledge. It simplifies the presentation, and affects no conclusion in the current context, to be explicit only about those beliefs that have a direct bearing on some players' payoff perception. Only τ'' appears at an end node in our case, and therefore $\tau'' = \tau' = \tau$ is the only explicit restriction on beliefs that is made.

If $\gamma < 4$, the agent's guilt sensitivity is insufficient to sustain any cooperation, and the prediction is that $(\sigma, \tau) = (0, 0)$ with $\tau'' = \tau' = 0$ (*i.e.* the principal chooses *No* and the agent chooses $e = 0$). However, there are multiple equilibria with guilt aversion if $\gamma \geq 4$. In particular, we have:

1. $(\sigma, \tau) = (0, 0)$ with $\tau'' = \tau' = 0$
2. $(\sigma, \tau) = (1, 1)$ with $\tau'' = \tau' = 1$

Thus, there are both 'good' and 'bad' equilibria, with the second equilibrium being obviously 'better'.¹⁶ In the second equilibrium *both* the principal and the agent gain relative to the first (*No*, $e = 1$) outcome. The key idea we stress is that *communication may help bring about the second equilibrium*. A happy ending may be brought about in a way which reflects on Leith & Baumeister's (1998, p. 1) assertions that "guilt serves many adaptive, beneficial, and prosocial functions", and that "guilt helps strengthen and maintain close relationships" (p. 2).

4.3 Further tests of dislike of lying and guilt aversion

Our rejection of H2 is consistent with dislike for lying as well as guilt aversion. In this section we devise additional tests that may provide further support for these ideas.

Dislike of Lying

We have access to the messages sent by agents, so we can check directly whether or not promises make for higher cooperation. Since messages can have nearly any form, this requires a classification of the messages. We have used three categories: promises, empty talk, and no

¹⁶ There are additional equilibria in mixed strategies: **3.** $\gamma \in [4, 8]$, $(\sigma, \tau) = (1, 4/\gamma)$, $\tau'' = \tau' = 4/\gamma$; **4.** $\gamma = 8$, $(\sigma, \tau) = (\sigma, 4/\gamma)$, $\tau'' = \tau' = 4/\gamma$, $\sigma \in [0, 1]$; **5.** $\gamma \geq 8$, $(\sigma, \tau) = (0, 4/\gamma)$, $\tau'' = \tau' = 4/\gamma$.

message.¹⁷ The promises category is broad, including any ‘statement of intent’ that we could find. Our classification is given in Appendix B, along with the raw data on individual choices.

Twenty-four of the 42 agents (57%) made promises to ROLL, 14 agents (33%) sent messages with no promise, and four agents (10%) sent no message. We test the following null, which we reject for both principal behavior and agent behavior:

H3: *Neither the proportion of principals accepting contracts nor the proportion of agents choosing high effort will be increased if a sent message contains a statement of intent (a promise).*

Twenty-two of the 24 principals (92%) who received promises chose IN. For principals who did not receive a promise, nine of 18 in the communication treatment (50%), or 34 of 63 (54%) overall, chose IN. The difference in principal behavior is quite significant for each comparison ($Z = 3.04, p = 0.001$ or $Z = 4.22, p = 0.00001$, respectively).

With respect to agents’ behavior, 18 of the 24 promises (75%) were actually kept (the agent chose ROLL). This compares with 10 of the 18 (56%) non-promising agents in the communication treatment, or 30 of the 63 non-promising agents (48%) overall, choosing ROLL. Comparing the ROLL rate with a promise to the rate without a promise, we see that the difference is weakly significant in the first case ($Z = 1.32, p = 0.093$) and quite significant ($Z = 2.30, p = 0.011$) with the larger sample.

We can also compare across treatments to see if not promising in the communication treatment (where promising is possible) yields the same outcomes as in the no-communication treatment (where it is not). Neither the 7/14 ROLL rate nor the 10/18 ROLL rate is significantly different than the 20/45 ROLL rate in the no-communication treatment ($Z = 0.36$ and 0.80 , for the respective comparisons); similarly, the 8/14 or 9/18 IN rates with empty talk or non-promises

¹⁷ It is common in social psychology to code responses according to various classifications. While we only consider the classification in the text, we provide (in Appendix B) the complete messages for those readers who wish to consider alternative coding. Some of the messages are rather colorful, and serve well to enliven proceedings in seminars. Consider, e.g., message 7 in session 3, which contains a poem by Samuel Francis Smith and fictitious references to desires and advice from some famous persons....

are nearly the same as the 25/45 IN rate in the no-communication treatment ($Z = 0.10$ and 0.40 , for the respective comparisons). When a promise was made, the successful partnership outcome (IN, ROLL) resulted in 67% of the cases (16/24). This compares to 28% successful outcomes (5/18) for non-promises, and an expected rate of 25% in the no-communication treatment.¹⁸ It seems clear that the observed difference between communication and no-communication treatments is driven not by messages *per se*, but rather by the promises made by agents.

Although our results indicate that promises foster cooperation, honesty is not ubiquitous. Six participants (out of 24) made a promise that they didn't keep. This accords well with findings in the literature on self-serving deception in (signaling or bargaining) games with asymmetric information, which shows that some people engage in deception in certain situations.¹⁹

Guilt Aversion

Guilt aversion depends on beliefs about beliefs. One may devise a related experimental test, which requires elicitation of the relevant beliefs. Our experimental design is set up to achieve this, and we report the results in this section. The testing strategy avoids incorporating an assumption of equilibrium play; in equilibrium beliefs are required to be correct, and this by itself implies certain belief-behavior correlations beyond the (motivational) idea of guilt aversion. The proposed test (which is reminiscent of observation (iv) following the tipping example in section 4.2) refers only to the agent and his utility function.

Recall that in Figure 3, the parameter $\gamma > 0$ scales the agent's guilt sensitivity. If γ is large enough, and if the agent believes sufficiently strongly that the principal believes he will choose $e = 1$ (i.e., if τ is large enough), then the inequality $14 - \gamma\tau < (5/6) \cdot 12 + (1/6) \cdot 0$ holds, and the agent

¹⁸ The average (Principal, Agent) earnings were (7.08, 10.58) with promises, (5.28, 8.39) with non-promises, and (4.69, 9.01) in the no-communication treatment. Both principals and agents earn more when a promise is made. The average total payoffs of 17.66, 13.67, and 13.70 translate to efficiency rates of 76.6%, 36.7%, and 37.0%, considering that the minimum total payoff is 10 and the expected maximum is 20.

¹⁹ See Blume, DeJong, Kim & Sprinkle (2001), Forsythe, Lundholm & Rietz (1999), Boles, Croson & Murnighan (2000), and Croson, Boles & Murnighan (forthcoming). Croson (2002) reviews the results.

will indeed choose $e = 1$. Alternatively put, the lower is τ , the higher γ must be in order for the agent to prefer $e = 1$. If γ differs among individuals and is independent of τ , there will be positive correlation between the likelihood of an $e = 1$ choice and τ .²⁰ This provides a way to test for the importance of guilt aversion.²¹

Our experimental design involves an attempt to measure τ . After we collected the strategic decisions made, we passed out decision sheets that invited participants to make guesses about the choices of their counterparts, and offered to reward good guesses. A's were asked to guess the proportion of B's who chose ROLL.²² B's were analogously asked to guess the average guess made by A's who chose IN. If a guess was within five percentage points of the realization, we rewarded the guesser with \$5 (we also told participants that we would pay \$5 for all B guesses if no A's had chosen IN).

We chose this belief-elicitation protocol mainly because it is simple and rather easy to describe in instructions. Our method, which invites the subjects to make certain maximum likelihood guesses, at the cost of the exclusion of (rational) guesses of less than 5% or greater than 95% which may have been rational under the alternative of (more complicated) quadratic-scoring rules. Our working hypothesis is that we get a rough but meaningful ballpark estimate of subjects' 'degrees of beliefs'. As our game is one-shot and we didn't mention guesses until after strategies were chosen, the belief elicitation should not affect participants' prior choices.

Our null hypothesis rules out the belief-behavior correlation described above:

H4: *There is no correlation between the agent's expectation of the principal's expectation of a favorable response and the frequency of trustworthy responses in the experiment.*

²⁰ Tangney (1995) asserts that "there are stable individual differences in the degree to which people are prone to shame and guilt".

²¹ If we were to apply the analogous idea developed in this paragraph to the DK reciprocity model, *i.e.* if instead of looking at equilibrium we assumed that Y_i differed among individuals and were independent of τ , then the DK model would predict a *negative* correlation between the likelihood of an $e = 1$ choice and τ .

²² We did not ask A's to guess the likelihood that the paired B would choose ROLL, as we don't observe this likelihood. The observed binary choice would make this simply a Yes or No guess.

H4 is rejected, as we find a significant positive correlation in both treatments. In the no-communication case, agents who chose DON'T ROLL guessed that principals who chose IN guessed on average that 40.4% of agents would ROLL, compared to the 53.2% guessed by the B's who chose ROLL. A Wilcoxon-Mann-Whitney rank-sum test (see Siegel & Castellan, 1988) finds that the guesses of the ROLL group are significantly higher ($Z = 1.99$, $p = 0.046$, two-tailed test).

This gap across agents widens considerably in the communication treatment, where agents who chose DON'T ROLL guessed 45.1% and agents who chose ROLL guessed 73.2% ($Z = 3.20$, $p = 0.001$, two-tailed test), suggesting that messages have some effectiveness in focusing expectations. Thus H4 is rejected in favor of the alternative hypothesis with positive correlation, consistent with guilt aversion.

We also find that agents who make a promise and choose ROLL exhibit higher guesses about principals' guesses than do non-promising agents in the communication treatment ($Z = 1.70$, $p = 0.045$, one-tailed Wilcoxon test) or non-promising agents overall ($Z = 2.29$, $p = 0.011$, one-tailed Wilcoxon test). One possible interpretation is that promises foster high expectations, which coupled with guilt aversion creates a commitment device for agents to ROLL.

5. DISCUSSION

We examine the impact of communication on cooperation in a game designed to capture the essence of hidden action, as treated in much of contract theory. Hidden action is less of a stumbling block in the lab than classical contract theory predicts, even without communication. When we allow an agent to send a free-form message to a principal, the problem is mitigated to a much further degree.

To know which motivational forces are at work and why and how communication influences the interaction is essential for deriving insights regarding optimal contracting, and for formulating testable predictions in other settings to which the results may extend. We examine whether several approaches can explain our findings.

We first look at models in which decision-makers care only about the overall final distribution of monetary payoffs in a game, and ask whether such models can capture the impact of communication in our setting. The answer is negative. The equilibrium prediction is unique and independent of whether or not communication is allowed. We then turn to models of kindness-based reciprocity, with belief-dependent utilities. One might suspect that communication would have an impact, if beliefs were sensitive to messages. However, there is again a unique equilibrium prediction (in DK's model of sequential reciprocity), so again the impact of communication is not explained. In sum, models of distributional preferences as well as of kindness-based reciprocity can explain behavior in *either* the no-communication or the communication treatment, but cannot explain the behavioral differentials *across* treatments.

We then turn to two alternative motivational forces: dislike for lying and guilt aversion. The first idea links words to motivation, the second idea links beliefs about beliefs to motivation. Both approaches can potentially explain the differences in principal and agent behavior in our treatments, and we present experimental evidence which supports this.

Dislike for lying may seem an especially attractive assumption because of its simplicity. However, dislike of lying cannot explain the observed significant positive correlation between the likelihood of the agent's high-effort choice and his belief about the principal's belief of that choice when messages (and therefore promises and lies) are not feasible. Dislike for lying may explain part of our results, but the idea cannot serve as the sole explanation.

Consider instead guilt aversion. In order to focus on and isolate the impact of this motivational force, we define a solution concept illustrating how communication may help select an equilibrium with cooperative play when the agent is guilt averse. Promises may play an important role, as commitment devices to carry out trustworthy choices. The psychological mechanism by which this happens is different when an agent dislikes lying and when he is guilt averse. With guilt aversion the agent lives up to his promise because he believes it is believed, not because he hates to lie. Our experimental design allows us to observe the belief (about a belief) that is central to guilt aversion, and we find positive correlation between this belief and

the trustworthy action as implied. Guilt aversion can also explain what goes on in the treatment without communication, so it is the only model that is capable of explaining our data by itself.

Using guilt aversion to explain behavior does not imply disregard of previous experimental findings regarding cooperation in games that incorporate an element of trust. Rather, it is a reinterpretation that is suggested. For example, results in gift-exchange games are usually taken to illustrate reciprocal forces at work (see Fehr & Gächter, 2000).²³ Guilt aversion implies that the more effort a ‘worker’ expects that his ‘firm’ expects him to exert the more effort he will exert, because if he did not he would feel guilty letting the firm down. If firms believe that higher wages trigger higher effort, and agents believe this, the oft-observed positive wage-effort relationship will result.

An intriguing aspect of dislike of lying and guilt aversion is the potential these motivational forces have for *a theory of why communication matters*. By making a promise to behave in a trustworthy fashion the agent strengthens the principal's degree of belief that the agent can be relied upon, because by the force of his very word, or by the force of his own changed second-order belief, the agent creates an incentive for himself to live up to his promise.²⁴ A promise feeds a beneficial, self-fulfilling circle of beliefs, beliefs about beliefs, and trustworthy behavior. Truth-telling becomes ‘self-enforcing’; in a trust situation there will be no incentive to lie or renege on a promise.

Although more work is needed to delineate more exactly the range of situations in which dislike of lying and guilt aversion plays a role,²⁵ we propose that the ideas be seriously

²³ Note, however, that reciprocity has two sides, positive reciprocity, where a player is kind in return to another's kind choice, and negative reciprocity, where a player is unkind in return to another's unkind choice. In our game, the only way the principal can be unkind is by not agreeing to the partnership. The agent then has no subsequent choice, so we cannot observe negative reciprocity. However, negative reciprocity seems important in other games; see e.g., Kahneman, Knetsch & Thaler (1986), Blount (1995), Charness (1996), Offerman (2002), Brandts & Charness (1999) Andreoni, Brown & Vesterlund (2002), Kagel & Wolfe (2001), Charness & Rabin (2002).

²⁴ This insight for a *psychological* game is reminiscent of ideas explored in the literature on *cheap talk* in *standard* games; see Farrell & Rabin (1996) and Crawford (1998) for surveys, and Jamison (2000) for a recent model.

²⁵ Clearly that is not *always* the case; it is hard, e.g., to imagine poker players feeling guilty, no matter how they deceive. Another issue concerns expectations that may be deemed ‘unreasonable’ or ‘obnoxious’. How would this be determined? Moreover, guilt aversion may matter more when the party making a promise knows that the other

considered as fundamental for understanding communication, partnerships, contracts, and human interaction quite generally. Examples range from everyday experiences, like tipping in a café, to many kinds of partnerships, including husband & wife, lawyer & client, procurement agency & contracted firm, inventor & producer, talented young golfer & rich sponsor, co-owners of firms, employer & employee, etc. The importance of guilt aversion in these situations does moreover not necessarily have to do only with promises and lies or other one-sided messages. Why and how do people discuss, argue, and debate? How do such exchanges influence group decisions, formation of and adherence to social norms, partnership formation, and contracting? Perhaps a key aspect is that these exchanges lead up to commonly-expected standards of behavior that, once in place, are shared and not violated by guilt-averse people.

This suggests a promising lode to explore in future research, and some issues may be pursued within the general framework we developed earlier in this paper (section 2). For example, what is the impact, on behavior and beliefs, for example, varying details of the communication protocol, say, having messages from the principal or having interactive responses between principals and agents?²⁶ Or, how do matters change in settings plagued by hidden information? One may reasonably suspect that again communication fosters cooperation. Agents might say, “I promise I am a person with high talent”, and perhaps people with low talent find it unbearable to lie about such matters. Examining this proposition could shed light on whether or not hidden information is more or less problematic than hidden action.

We close the paper by stressing that there is also ample scope for theoretical work. Contract theory has a history of basking in the light of tremendous intellectual achievement, and incorporating communication, guilt aversion, and dislike for lying into the analysis would extend

party will find out whether a promise is kept (which may still be consistent with a hidden-action environment, if these facts are not provable in court). Charness & Grosskopf (forthcoming) find that observability may matter.

²⁶ Comparisons with related work on other games is then possible. For example, a form of one-sided communication by principals appears in some recent gift-exchange studies: firms offer contracts consisting of wage and desired effort; see Fehr, Gächter, and Kirchsteiger (1997), Fehr & Gächter (2002), and Fehr, Klein & Schmidt (2001). The last two studies report positive correlations between desired and actual effort (though not always significant). Beliefs were not measured, but the findings seem to rhyme rather well with guilt-aversion predictions, if statements of desired effort shape beliefs and beliefs about beliefs.

this tradition. As experimentalists accumulate insights regarding which behavioral ideas have bearing on understanding partnerships and contracts, it makes sense to develop new theory based on these ideas. How might one, for example, characterize optimal contractual arrangements when agents are affected by guilt aversion? To answer such questions seems to us an exciting challenge in behavioral contract theory.

REFERENCES

- Anderhub, Vital, Simon Gächter & Manfred Königstein (2002), "Efficient Contracting and Fair Play in a Simple Principal-Agent Experiment", *Experimental Economics*, **5**, 5-27.
- Andreoni, James, Paul Brown & Lise Vesterlund (2002), "What Produces Fairness? Some Experimental Evidence", *Games & Economic Behavior*, **40**, 1-24.
- Bacharach, Michael, Gerardo Guerra & Daniel Zizzo (2001), "Is Trust Self-Fulfilling? An Experimental Study" mimeo.
- Baumeister, Roy, Arlene Stillwell & Todd Heatherton (1994), "Guilt: An Interpersonal Approach", *Psychological Bulletin*, **115**, 243-67.
- Baumeister, Roy, Arlene Stillwell & Todd Heatherton (1995), "Personal Narratives about Guilt: Role in Action Control and Interpersonal Relationships", *Basic & Applied Social Psychology*, **17**, 173-98.
- Berg, Joyce, John Dickhaut & Kevin McCabe (1995), "Trust, Reciprocity, and Social History," *Games & Economic Behavior*, **10**, 122-42.
- Bicchieri, Cristina (2002), "Covenants without Swords", *Rationality & Society*, **14**, 187-222.
- Blau, Peter (1964), *Exchange and Power in Social Life*, New York: John Wiley.
- Blount, Sally (1995), "When Social Outcomes Aren't Fair: The Effect of Causal Attributions on Preferences", *Organizational Behavior & Human Decision Processes*, **LXIII**, 131-44.
- Blume, Andreas, Douglas DeJong, Yong-Gwan Kim & Geoffrey Sprinkle (2001), "Evolution of Communication with Partial Common Interest", *Games & Economic Behavior*, **37**, 79-120.
- Bohnet, Iris and Bruno Frey (1999), "The Sound of Silence in Prisoner's Dilemma and Dictator Games", *Journal of Economic Behavior & Organization*, **38**, 43-57.

- Boles, Terry, Rachel Croson, Keith Murnighan (2000), "Deception and Retribution in Repeated Ultimatum Bargaining", *Organizational Behavior & Human Decision Processes*, **83**, 235-59.
- Bolton, Gary & Axel Ockenfels (2000), "ERC: A Theory of Equity, Reciprocity, and Competition", *American Economic Review*, **90**, 166-93.
- Brandts, Jordi & Gary Charness (2003), "Truth or Consequences: An Experiment", *Management Science*, **49**, 116-30.
- Brosig, Jeannette, Axel Ockenfels & Joachim Weimann (2003), "The Effect of Communication Media on Cooperation", *German Economic Review*, **4**, 217-41.
- Cabrales, Antonio & Gary Charness (2000), "Optimal Contracts, Adverse Selection, and Social Preferences: An Experiment," mimeo.
- Charness, Gary (forthcoming), "Attribution and Reciprocity in an Experimental Labor Market", *Journal of Labor Economics*.
- Charness, Gary (2000) "Self-serving Cheap Talk and Credibility: A Test of Aumann's Conjecture", *Games & Economic Behavior*, **33**, 177-94.
- Charness, Gary & Brit Grosskopf (forthcoming), "Cheap Talk, Information, and Coordination: Experimental Evidence", *Economics Letters*.
- Charness, Gary & Matthew Rabin (2001), "Expressed Preferences and Behavior in Experimental Games", mimeo.
- Charness, Gary & Matthew Rabin (2002), "Understanding Social Preferences with Simple Tests", *Quarterly Journal of Economics*, **117**, 817-69.
- Cooper, Russell, Douglas DeJong, Robert Forsythe & Thomas Ross (1990), "Selection Criteria in Coordination Games: Some Experimental Results", *American Economic Review*, **53**, 218-33.
- Cooper, Russell, Douglas DeJong, Robert Forsythe & Thomas Ross (1992), "Communication in Coordination Games", *Quarterly Journal of Economics*, **53**, 739-71.
- Cox, James (2004), "How to Identify Trust and Reciprocity", *Games & Economic Behavior*, **46**, 260-81.
- Cox, James & Daniel Friedman (2002), "A Tractable Model of Reciprocity & Fairness", mimeo.
- Crawford, Vincent (1998), "A Survey of Experiments on Communication via Cheap Talk", *Journal of Economic Theory*, **78**, 286-98.
- Croson, Rachel (2002), "Game-Theoretic and Experimental Perceptions of Deception", mimeo.

- Croson, Rachel, Terry Boles & J. Keith Murnighan (forthcoming), “Cheap Talk in Bargaining Experiments: Lying and Threats in Ultimatum Games”, *Journal of Economic Behavior & Organization*.
- Dawes, Robyn, Jeanne McTavish & Harriet Shaklee (1977), “Behavior, Communication, and Assumptions about Other People’s Behavior in a Commons Dilemma Situation,” *Journal of Personality & Social Psychology*, **35**, 1-11.
- Dufwenberg, Martin (2002), “Marital Investment, Time Consistency, and Emotions”, *Journal of Economic Behavior & Organization*, **48**, 57-69.
- Dufwenberg, Martin & Uri Gneezy (2000), “Measuring Beliefs in an Experimental Lost Wallet Game”, *Games & Economic Behavior*, **30**, 163-82.
- Dufwenberg, Martin & Georg Kirchsteiger (forthcoming), “A Theory of Sequential Reciprocity”, *Games & Economic Behavior*.
- Dufwenberg, Martin & Michael Lundholm (2000), “Social Norms and Moral Hazard”, *Economic Journal*, **111**, 506-25.
- Dutta, Prajit K. & Roy Radner (1993), “Moral Hazard”, in Robert Aumann & Sergiu Hart (eds.), *Handbook of Game Theory*, vol. 2.
- Ellingsen, Tore & Magnus Johannesson (2002), “Promises, Threats, and Fairness”, mimeo.
- Falk, Armin & Urs Fischbacher (1998) “A Theory of Reciprocity,” mimeo.
- Farrell, Joseph & Matthew Rabin (1996), “Cheap Talk”, *Journal of Economic Perspectives*, **10**, 103-18.
- Fehr, Ernst & Simon Gächter (2000), “Fairness and Retaliation: The Economics of Reciprocity”, *Journal of Economic Perspectives*, **14**, 159-81.
- Fehr, Ernst & Simon Gächter (2002), “Do Incentive Contracts Undermine Voluntary Cooperation?”, mimeo.
- Fehr, Ernst, Simon Gächter, Georg Kirchsteiger (1997), “Reciprocity as a Contract Enforcement Device - Experimental Evidence”, *Econometrica*, **64**, 833-60.
- Fehr, Ernst, Alexander Klein & Klaus Schmidt (2001), “Fairness, Incentives and Contractual Incompleteness”, mimeo.
- Fehr, Ernst & Klaus Schmidt (1999), “A theory of fairness, competition, and cooperation”, *Quarterly Journal of Economics*, **114**, 817-68.

- Fehr, Ernst & Klaus Schmidt (2002), "Theories of Fairness and Reciprocity – Evidence and Economic Applications," in: M. Dewatripont, L. Hansen & S. Turnovsky (Eds.), *Advances in Economics and Econometrics – 8th World Congress, Econometric Society Monographs*, Cambridge University Press.
- Forsythe, Robert, Russel Lundholm, Thomas Rietz (1999), "Cheap Talk, Fraud, and Adverse Selection in Financial Markets: Some Experimental Evidence", *Review of Financial Studies*, **12**, 481-518.
- Geanakoplos, John, David Pearce & Ennio Stacchetti (1989), "Psychological Games and Sequential Rationality", *Games & Economic Behavior*, **1**, 60–79.
- Glasnapp, Douglas & John Poggio (1985), *Essentials of Statistical Analysis for the Behavioral Sciences*, Columbus, Merrill.
- Gneezy, Uri (2002), "Deception: The Role of Consequences", mimeo.
- Güth, Werner, Wolfgang Klose, Manfred Königstein & Joachim Schwalbach (1998), "An experimental study of a dynamic principal-agent relationship", *Managerial & Decision Economics*, **19**, 327-341.
- Hart, Oliver D. & Bengt Holmström (1986), "The Theory of Contracts", in Truman Bewley (ed.), *Advances in Economic Theory*, Cambridge University Press.
- Holmström, Bengt (1979), "Moral Hazard and Observability", *Bell J. of Economics*, **10**, 74-91.
- Huang, Peter & Ho-Mou Wu (1994), "More Order without More Law: A Theory of Social Norms and Organizational Cultures", *Journal of Law, Economics & Organization*, **10**, 390-406.
- Jamison, Julian (2000), "Valuable cheap talk and equilibrium selection", mimeo.
- Kagel, John & Katherine Wolfe (2001), "Tests of Fairness Models Based on Equity Considerations in a Three-Person Ultimatum Game", *Experimental Economics*, **4**, 203-19.
- Kahneman, Daniel, Jack Knetsch & Richard Thaler (1986), "Fairness and the Assumptions of Economics", *Journal of Business*, **59**, S285-S300.
- Leith, Karen P. and Baumeister, Roy F. (1998), "Empathy, Shame, Guilt, and Narratives of Interpersonal Conflicts: Guilt-Prone People Are Better at Perspective Taking", *Journal of Personality*, **66**, 1–37.
- Loewenstein, George (1999), "Experimental Economics from the Vantage-Point of Behavioral Economics", *The Economic Journal*, **109**, F25-F34.
- MacLeod, W. Bentley & James Malcolmson (1989), "Implicit Contracts, Incentive Compatibility, and Involuntary Unemployment", *Econometrica*, **57**, 447-80.

- Nash, John (1950), "The Bargaining Problem", *Econometrica*, **18**, 155-62.
- Offerman, Theo (2002), "Hurting Hurts More than Helping Helps: The Role of the Self-serving Bias", *European Economic Review*, **46**, 1423-37.
- Orbell, John, Robyn Dawes & Alphons van de Kragt (1990), "The Limits of Multilateral Promising", *Ethics*, **100**, 616-27.
- Rabin, Matthew (1993), "Incorporating Fairness into Game Theory and Economics", *American Economic Review*, **83**, 1281-1302.
- Rousseau, Denise (1995), *Psychological Contracts in Organizations: Understanding Written and Unwritten Agreements*, Thousand Oaks, California: Sage Publications.
- Segal, Uzi & Joel Sobel (1999), "Tit for Tat: Foundations of Preferences for Reciprocity in Strategic Settings", mimeo.
- Selten, Reinhard (1967), "Die Strategiemethode zur Erforschung des Eingeschränkt Rationalen Verhaltens im Rahmen eines Oligopolexperiments", in *Beiträge zur Experimentellen Wirtschaftsforschung*, H. Saueremann, ed., 136-68.
- Siegel, Sidney & N. John Castellan (1988), *Nonparametric Statistics for the Behavioral Sciences*, Boston, McGraw-Hill.
- Sobel, Joel (1999), "Interdependent Preferences and Reciprocity", mimeo.
- Tangney, June (1995), "Recent Advances in the Empirical-Study of Shame and Guilt," *American Behavioral Science*, **38**, 1132-45.
- Valley, Kathleen, Joseph Moag & Max Bazerman (1998), "'A Matter of Trust': Effects of Communication on the Efficiency and Distribution of Outcome", *Journal of Economic Behavior & Organization*, **34**, 211-38.

APPENDIX A - INSTRUCTIONS

[text in the message treatment is shown in brackets]

Thank you for participating in this session. The purpose of this experiment is to study how people make decisions in a particular situation. Feel free to ask us questions as they arise, by raising your hand. Please do not speak to other participants during the experiment.

You will receive \$5 for participating in this session. You may also receive additional money, depending on the decisions made (as described below). Upon completion of the session, this additional amount will be paid to you individually and privately.

During the session, you will be paired with another person. However, no participant will ever know the identity of the person with whom he or she is paired.

Decision tasks

In each pair, one person will have the role of A, and the other will have the role of B. The amount of money you earn depends on the decisions made in your pair.

On the designated decision sheet, each person A will indicate whether he or she wishes to choose IN or OUT. If A chooses OUT, A and B each receives \$5. We will collect these sheets after the choices have been indicated. Next, each person B will indicate whether he or she wishes to choose ROLL or DON'T ROLL (a die). Note that B will not know whether A has chosen IN or OUT; however, since B's decision will only make a difference when A has chosen IN, we ask B's to presume (for the purpose of making this decision) that A has chosen IN.

If A has chosen IN and B chooses DON'T ROLL, then B receives \$14 and A receives \$0. If B chooses ROLL, B receives \$10 and rolls a six-sided die to determine A's payoff. If the die comes up 1, A receives \$0; if the die comes up 2-6, A receives \$12. (All of these amounts are in addition to the \$5 show-up fee.) This information is summarized in the chart below:

	A receives	B receives
A chooses OUT	\$5	\$5
A chooses IN, B chooses DON'T ROLL	\$0	\$14
A chooses IN, B chooses ROLL, die = 1	\$0	\$10
A chooses IN, B chooses ROLL, die = 2,3,4,5, or 6	\$12	\$10

[A Message

Prior to the decision by A and B concerning IN or OUT, B has an option to send a message to A. Each B receives a blank sheet, on which a message can be written, if desired. We will allow time as needed for people to write messages, then these will be collected. Please print clearly if you wish to send a message to A.

In these messages, no one is allowed to identify him or herself by name or number or gender or appearance. (The experimenter will monitor the messages. Violations (experimenter discretion) will result in B receiving only the \$5 show-up fee, and the paired A receiving the average amount received by other A's.) Other than these restrictions, B may say anything that he or she wishes in this message. If you wish to not send a message, simply circle the letter B at the top of the sheet.]

B

You may print a message to A below if you wish.

A

MAKE A GUESS

We now ask you to guess the percentage of **B's who chose ROLL.**

I guess that _____% of all B's chose ROLL.

Payment for the guess

If your guess differs by no more than 5 percentage points from the actual percentages, you will receive \$5.00.

If your guess differs by more than 5 percentage points from the actual percentages, you will receive \$0.

B

MAKE A GUESS

We have asked A's to make guesses about the percentages of B's who chose ROLL. We now ask you to guess some of the average guesses made by those A's who chose IN.

For A's who chose IN, I guess that the average guess about the percentage of B's who chose ROLL is _____%.

Payment for guess:

If your guess differs by no more than 5 percentage points from the actual percentages, you will receive \$5.00.

If your guess differs by more than 5 percentage points from the actual percentages, you will receive \$0.

(If there are no A's who chose IN, you will be paid \$5.00 for your guess, regardless of your answer.)

APPENDIX B - MESSAGES

In this table: P = Promise, E = Empty Talk, N = No Message, R = ROLL, DR = DON'T ROLL

Sess.	ID	Message	Class	Action	Principal
1	1	Please choose In so we can get paid more.	E	DR	OUT
1	2	Choose <u>in</u> , I will roll dice, you are 5/6 likely to get 2,3,4,5, or 6 → \$12. This way both of us will win something.	P	DR	IN
1	3	If you stay in, the chances of the die coming up other than 1 are 5 in 6 – pretty good. Otherwise, we'd both be stuck at \$5. (If you opt out)	E	DR	IN
1	4	I have to do laundry tonight and I really don't want to do it! But I don't have any clean underwear left and I don't want to go commando tomorrow. We'll see what I decide tonight. This man acts funny doesn't he? But he seems cool, he's quite a character. All this mystery is kinda cool.	E	R	OUT
1	5	If you will choose "In", I will choose to roll. This way, we both have an opportunity to make more than \$5! ☺	P	R	IN
1	6		N	R	OUT
1	7	If I roll a 2-6 (you'll know when you receive the \$, you will give \$5.00 to a stranger. [[[then there is a line, under which is written "Sign here if you are so kind]]] Thanks. You'll still be gaining more than if I had chosen Don't roll.	P	R	IN
1	8	The fairest thing to do is if you opt "IN". Then I will proceed to choose "roll." That way you and I have 5/6 chances to make money for the both of us. That's much better than just making \$5 each. Increases both our chances. Thanks.	P	R	IN
1	9	Choose In and I will Roll You have my word	P	DR	IN
1	10	Good luck I do not know what I'm going to do, so I have no hints on how to advise you on choosing "in" or "out." Though it would be beneficial for me to pick don't roll and hope you pick "in", I also like to give you a chance to gain some cash. <u>Who knows?</u>	E	R	IN
1	11	What's up? Good luck on your decision. Choose whatever. If you choose "out," you get only \$10 total. If you choose "In," you can get \$17 total instead of only \$10. 7 bucks is a lot of money!	E	DR	IN
1	12	Hey. OK I think that the best way for both of us to make a profit is for you to choose <u>IN</u> and for me to roll. That way we both make some money. There's no point in me not rolling because that would give you <u>and</u> me less profit. So I'm a roller if you're in ☺.	P	DR	IN
1	13	take a risk	E	R	IN
1	14	If you choose IN the first round and then I will choose Don't Roll at first. I will get \$14 but then after that I will choose roll each time after the 1 st role. Chances are most likely you will get \$12 and I will get only \$10. I will the only take 7 rolls for you to get even with me. That way we both leave with a good amount of money. Hope you have a great evening and that this works out for both of us. ☺	E	R	OUT

1	15	If you choose in I'll roll. Why? If you choose out, we walk out with \$10 each. If you choose IN & I choose IN then both of us coin. So it's a compromise. By agreeing to this I guarantee myself more \$ than risking you choose out. So if you choose out I get \$10 (\$5 diff.) if you choose in I get \$15 vs. \$19 (\$4 diff.). that's why	P	R	IN
1	16		N	DR	OUT
1	17		N	R	OUT
1	18	Choose "In" so we can both make some \$\$ What are the chances me rolling a 1? I'll try my best.	P	R	IN
2	1	I'm going to roll.	P	R	OUT
2	2	I'll choose roll.	P	R	IN
2	3	I will choose roll.	P	DR	IN
2	4	I'm going to choose roll	P	R	IN
2	5	choose in, & I'll roll.	P	R	IN
2	6	You can have the 2 extra dollars. I'll be nice and choose to roll. ☺	P	R	IN
2	7		N	R	IN
2	8	Hey, choose in and I will roll. You have to like your odds that I will roll a 2,3,4,5, or 6. 5/6 odds ain't bad.	P	DR	IN
2	9	If you choose "In", I'll choose Roll and you've got a 5/6 chance of getting \$12.	P	R	IN
2	10	Stay IN, I really need the money.	E	R	IN
2	11	If you choose IN, and I roll, the chances of our getting the most \$ are very high. The likelihood of my rolling a 1 is small compared to the chances of rolling a 2-6. So we both get cash.	E	DR	OUT
2	12	Hi, well I'm going to Roll so you have at least a shot for more money. I hope it works out.	P	R	IN
3	1	Hopefully I'll make a lucky role.	E	DR	OUT
3	2	It's much more likely that I'll roll a 2-6 and thus get more money then if we don't roll or choose out. I promise that I won't cheat you and that I'll choose to roll. ☺	P	R	IN
3	3	Tee hee, this is kinda Twilight Zone – ism; Why not "go for it", eh? I hope you have a lovely evening as well.	E	R	OUT
3	4	Hello fair stranger, anonymous partner ... Choose whatever you want. Far be it from me to influence your decision, but I think you should choose "in" and I should choose "roll" and we should take the chance at both earning as much as we can. 5 chances out of 6 say it'll work, and I'm totally broke, looking to rake in stray cash however I can. I feel the luck in the air. I don't really have much else to say. Hope you're doing well, whoever you are. Yes. That's all. Random note from random human	E	R	IN
3	5	Both of 'us' can earn.	E	DR	IN

3	6	<p>Ok. You're probably thinking, lets chose out, and I'll at least get 5 bucks. But... ...Chose 'IN', and I WILL chose to roll.</p> <p>The probability that I will roll a 2,3,4,5, or 6 is pretty high, and I think worthy of trying for.</p> <p>(I have no way of assuring you that I will roll ... but, its probably worth going for, you'll get \$12 for finding out, where I could get \$10.)</p> <p>x. I WILL ROLL</p>	P	R	IN
3	7	<p>I <u>will roll</u>, so if you stay in, you've got a 5/6 chance of getting \$12.</p> <p>If you don't mind the risk, if you stay in we'll both probably get more than \$5 ... Pretty cool to get money, eh? I'm kinda bored. Hope you've had a great day so far!</p> <p style="text-align: center;">My country Tis of Thee Sweet Land of Liberty Of Thee I sing. Land where my fathers died Land of the Pilgrim's Pride On every mountainside Let freedom ring.</p> <p>George W. Bush wants you to go in! Bin Laden says "out"! ☺</p>	P	R	IN
3	8	<p>Lets together get the most \$ out of this that we can. ⇒</p> <p style="text-align: center;">you 12 0 0 5 me 10 10 14 5</p> <p>I promise not to do this one. ↑</p> <p>I promise I will choose to roll. You can have the extra \$2 bucks. It's good karma.</p> <p style="text-align: center;">Thanks.</p> <p>I will choose ROLL in any case considering I will get the same amount no matter what you choose, as long as you choose IN.</p> <p style="text-align: center;">please excuse the awful handwriting. I'm trying</p>	P	DR	IN
3	9	<p>I'm choosing ROLL, which gives you a chance to get \$12 instead of \$5, so stay. It's a risk, but you could end up getting a lot more.</p>	P	R	IN
3	10	<p>If you choose in then I'm going to choose roll. This gives you a 5/6 chance of getting 12 dollars. That is 7 more than if you choose out. Since the money is free anyway – why not believe me. I'm don't lie – I promise I will choose roll.</p>	P	R	IN
3	11	<p>If you choose <u>IN</u> you have the best opportunity to make the most money. You have a 5/7 chance of making more money! So <u>IN</u> would be your best bet. Cheers. ☺</p>	E	DR	IN
3	12	<p>Choose IN. I promise I'll ROLL.</p>	P	R	OUT

APPENDIX C – SEQUENTIAL RECIPROCITY EQUILIBRIUM

The following calculations underlie the SRE derivations referred to in section 4.1. The technique resembles that of the proof of Observation 2 in DK's section 4(i).

Let the principal (agent) be player 1 (2). Using (6) and the other notation from section 4.1, and applying DK's theory, we get 2's utility, $m_2 + Y_2 \cdot \kappa_{21} \cdot \lambda_{212}$, from choosing $e=1$ with probability τ as

$$[14 - 4\tau] + Y_2 \cdot [10\tau - 5] \cdot [(14 - 4\tau'') - (9.5 - 2\tau'')]$$

Given Y_2 and τ'' , 2 will maximize this utility by choice of τ . In equilibrium, beliefs must furthermore be correct so that $\tau'' = \tau' = \tau$. With $Y_2 < 4/45$, this can only happen for $\tau'' = \tau' = \tau = 0$. With $Y_2 < 2/25$, this can only happen for $\tau'' = \tau' = \tau = 1$. With $Y_2 \in (2/45, 4/45)$ this can only happen for $\tau'' = \tau' = \tau = (45 Y_2 - 4)/20 Y_2$. The analysis so far pins down 2's choice uniquely for each possible value of Y_2 .

Now look at player 1 (the principal). In equilibrium she has correct expectations, so that $\tau' = \tau$. If she were selfish in the sense that $Y_1 = 0$, she would choose *No* whenever $\tau' = \tau < 1/2$, and *Yes* whenever $\tau' = \tau > 1/2$. That conclusion will not change when $Y_1 > 0$; the condition $\tau = 1/2$ determines the cutoff point between 2 being kind and unkind, and since 1 is unkind by choosing *No* and kind by choosing *Yes* any inclination to reciprocate must go in the same direction as the material incentive that would be relevant were $Y_1 = 0$.

It remains to look at the cases where in fact $\tau = 1/2$ (which requires $(45 Y_2 - 4)/20 Y_2 = 1/2$, or equivalently $Y_2 = 4/35$). In this case 1 may make any choice (including mixes) in equilibrium.

Figure 1

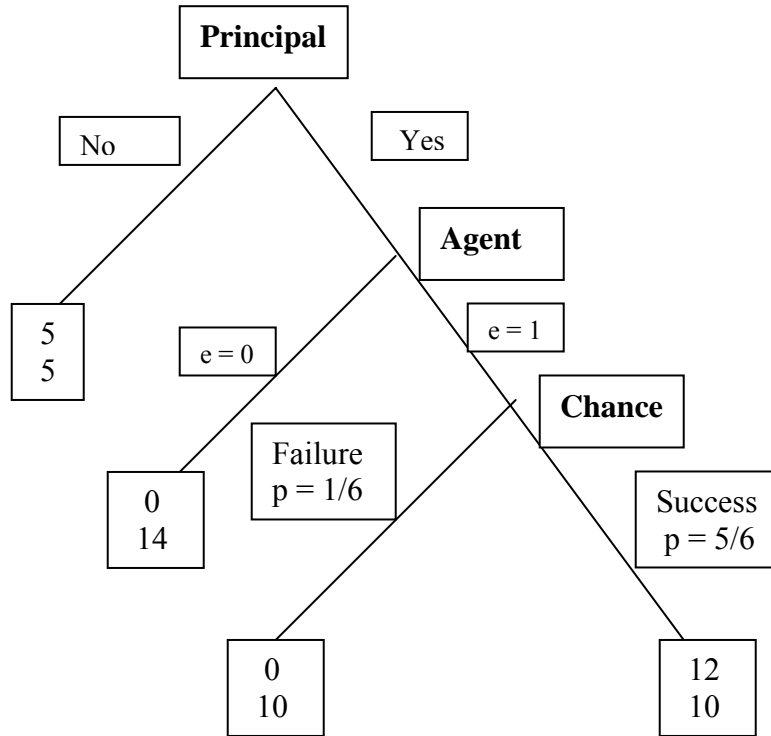


Figure 2

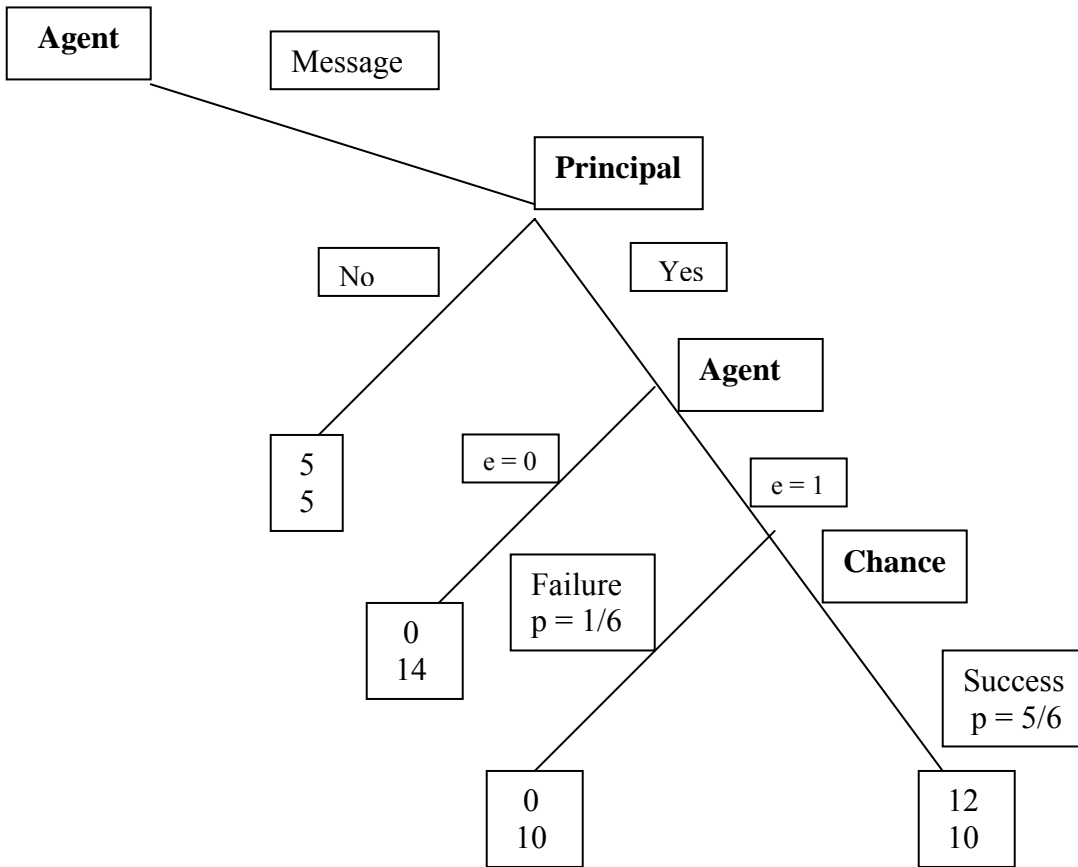


Figure 3

