Promises & Partnership

Gary Charness & Martin Dufwenberg*

October 11, 2004

Abstract: We examine, experimentally and theoretically, the impact of communication on trust and cooperation. We are especially interested in how problems of hidden action (as treated in contract theory) may be overcome. Our design admits observation of promises, lies, and beliefs. The evidence supports a theory according to which people strive to live up to others expectations in order to avoid guilt. The theory admits promises to enhance trustworthy behavior, which is what we observe in the experiment. Motivated by these results, we develop a notion of guilt-aversion equilibrium for extensive games. Besides explaining partnership interaction, the model sheds light on the role of language, discussions, agreements, and social norms more generally. The analysis also leads to some calls for more research in the field of psychological game theory.

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

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

We are grateful to Ninghua Du for excellent research assistance, Jon Baron, Paolo Battigalli, Jeanette Brosig, Steve Burks, Colin Camerer, Ernst Fehr, Ayelet Fishbach, Guillaume Frechette, Dan Friedman, Drew Fudenberg, Simon Gächter, Uri Gneezy, Brit Grosskopf, David Laibson, Rick Larrick, Dan Levin, David I. Levine, David K. Levine, Tanya Menon, Matt Parrett, Torsten Persson, David Reiley, Yuval Rottenstreich, David Strömberg, Richard Thaler, Dean Williamson, Bernd Wittenbrink, Bill Zame, seminar/conference participants at Antwerp, Berlin (Humboldt), Bonn, Chicago GSB, Cologne, Dortmund, Fuqua School of Business (Duke), Gothenburg, Harvard, Munich, Örebro, Oslo, Purdue, Santa Barbara, Stockholm, University of Arizona, 2002 GEW-Tagung (Wittenberg), 2002 Behavioral Research Council (Great Barrington), 2003 Public Choice (Tampa), the 2004 World Meetings of the Game Theory Society (Marseille), the SITE 2004 segment on Experimental Economics at Stanford, and ISNIE 2004 (Tucson) 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, and what the economic consequences will be. Considerable attention has been devoted to environments with *hidden action*, where a party's future choice is not contractible.[1] 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 examine, experimentally and theoretically, the impact of non-binding pre-play communication (*cheap talk*) on cooperation in a simple one-shot trust game designed to capture the essence of hidden action. The conventional contract theoretic approach implicitly assumes that such communication is ineffective in promoting partnership formation and cooperation.[2] Our study explores whether there are psychological aspects related to communication that conventional contract theory has not addressed.

In particular, we incorporate a new theoretical perspective, building on psychological game theory (see Geanakoplos, Pearce & Stacchetti 1989, henceforth GPS), which furnishes a reason why communication may foster trust and cooperation. The basic idea, which we refer to as *guilt aversion*, assumes that decision makers experience guilt if they believe that they let others down. That emotion leads to a non-standard conception of utility (from the viewpoint of classical game theory), whereby a decision-maker's preferences over strategies depend on his beliefs about the beliefs of others, even if there is no strategic uncertainty. Messages that may have seemed cheap talk gain cutting power; for example, guilt aversion permits a promise to feed a beneficial and self-fulfilling circle of beliefs about choices, beliefs about beliefs, trust, and trustworthiness.

---

[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).

[2] Our focus on *one*-shot interaction is worth emphasizing. We do not consider repeated games (cf., for example, MacLeod & Malcolmson 1989), in which communication may serve as an equilibrium-selection device.

In section 2, we consider the implications of guilt aversion for the specific games of our experiment. In the experiment, described in sections 3 and 4, we measure beliefs, record messages, and examine how 'statements of intent' correlate with subsequent choices. As it turns out, our theory does quite a good job explaining the data, even with respect to several alternatives that we consider. It turns out that a particular form of communication, namely promises to perform, can inspire an increased level of cooperation. We find, overall, that participants in both roles (principals and agents) make cooperative choices more than twice as often when the agent makes a promise to perform, with the corresponding rate of successful partnership formation more than quadrupling.

In light of the experimental support for guilt aversion, we find it natural to ask what this might imply in other games. In section 5 we consider what kind of solution concept we get if use the GPS framework to incorporate guilt aversion in general extensive games. We provide an answer, which we call *guilt-aversion equilibrium*. In order to demonstrate its usefulness, we consider several applications (our trust game, gift exchange, tipping, cartel, poker). We argue that the notion of guilt-aversion equilibrium provides the rudiments of a theory of why communication matters in many settings. However, we also discuss some inherent problems with the concept, which mainly have to do with restrictions built into GPS' underlying theoretical framework. This leads us to propose a few avenues for future research, mainly on the topic of generalizing psychological game theory.

The general idea that communication can affect strategic interaction in one-shot play is not new to our study (although other work seem to deal mainly with prisoners dilemmas, coordination games, or bargaining games, rather than trust games).[3] For example, a number of experiments (e.g., Dawes, McTavish & Shaklee 1977) provide evidence that face-to-face

---

[3] 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; Moreno & Wooders 1998; Charness 2000) and to matter in bargaining (Valley, Moag & Bazerman 1998; Bohnet & Frey 1999; Ellingsen & Johannesson 2002; Brandts & Charness 2003; Brosig, Ockenfels & Weimann 2003).

communication can greatly enhance cooperation, even when the dominant strategy (with selfish preferences) is ‘defection’. However, as Roth (1995) points out, there may be many confounding and uncontrolled effects in a face-to-face interaction. We do not allow face-to-face interaction, but instead permit anonymous, free-form messages to be transmitted from one party to the other. This permits a clean, controlled test of the role of communication.

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, taking into account social and psychological considerations in an attempt to understand partnerships and contracts. In this regard, we explore whether and why a given contract is acceptable to two parties, with and without communication. One objective is to gain insights about decision-making and motivation that are useful for further developing behavioral contract theory.

2. TRUST, COMMUNICATION & GUILT

This section sets the stage for the subsequent experiment. We introduce the trust game on which our design in based, giving it a contract-theoretic hidden-action interpretation (2.1). We incorporate communication (2.2), and introduce the key notion of guilt aversion from which our main hypotheses are derived (2.3). Finally, we discuss how some alternative theories may or may not shed light on the impact of communication (2.4).

2.1 A Trust Game

We consider ‘trust games’ like the one in Figure 1. The names of players and choices anticipate the experimental design. Payoffs are in dollars.

FIGURE 1

The unique subgame-perfect equilibrium for selfish risk-neutral players – the strategy profile (*Out*, *Don't Roll*) – is inefficient. The game thus has a 'dilemma' flavor, just like many previous trust games that have been studied.[4] Our game adds a twist: a chance move following (*In*, *Roll*) that determines whether A's payoff will be 12 (with probability 5/6) or 0 (with probability 1/6). This twist is essential to the following interpretation.

Think of A and B as a principal and an agent: The two 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 each get outside-option payoffs of 5. If the project is carried out, then the contract specifies a 'wage' that the principal pays the agent, and a (costly) 'effort' that the agent should exert. The project stochastically generates revenue for the principal, the success rate depending on the agent's effort. The strategy profile (*In*, *Roll*) corresponds to a Nash bargaining solution, assuming that effort and wage is contractible.[5] However, if effort *cannot* be regulated in the contract (so the agent is free to choose his effort) then the situation incorporates hidden action. There are two reasons why the Nash bargaining outcome may then not obtain. First, the agent may shirk, choosing lower effort. Second, the principal may foresee such a turn of events, dislike it, and not agree to form a partnership. The players' choices in the game of Figure 1 incorporate these two possibilities.[6]

Why have we bothered to include the chance move in our (otherwise rather spare) model, instead of replacing it with its expected outcome, (10,10)? The answer is that this is essential for making *conceptual* sense of our exercise as incorporating hidden action, in line with the scenario

---

[4] Cf. *e.g.* Güth, Ockenfels & Wendel (1994), Berg, Dickhaut & McCabe (1995), or Dufwenberg & Gneezy (2000).

[5] More precisely: The project can have two outcomes, poor or good. The poor outcome generates revenue 14, the good outcome involves an additional revenue of 12 (so the total revenue is 26). The probability of a good outcome is $5/6 \cdot e$, where $e \in [0,1]$ is the agent's effort (and 5/6 may be thought of as her talent). The agent's cost of effort is $4 \cdot e$. Given the outside options of 5 for each party, following Nash (1950) the bargaining solution for risk-neutral and selfish players is the wage-effort pair (w,e) that maximizes $[(14\text{-}w+5/6 \cdot e \cdot 12)\text{-}5] \cdot [(w\text{-}4 \cdot e)\text{-}5]$. The solution is $(w,e) = (14,1)$, with resulting payoffs as per (*In*, *Roll*) in the game of Figure 1.

[6] If the principal chooses not to join the partnership, each party earns the outside option of 5. If the principal chooses to join the partnership, and the agent chooses $e = 0$ the project fails ($5/6 \cdot 0 = 0$), so the principal gets revenue minus wage equals 14-14=0, while the agent gets wage minus effort cost equals $14\text{-}4 \cdot 0 = 14$. It is easy to verify that this corresponds to the various outcomes in Figure 1.

depicted in the preceding paragraph. Hidden action is a proposition of contractual limits, in this case that a contract cannot be conditioned on the agent's choice of effort. A typical justification, often stressed by contract theorists, is that an 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 in Figure 1 caters to this issue; with it, if the project fails due to low effort (*i.e.*, the choice *Don't Roll*), the agent (*i.e.,* player B) can claim that he exerted high effort but had bad luck. It cannot then be proven in court that this is a lie, since effort is not directly observable; the principal (*i.e.*, A) only knows she received a payoff of zero.[7]

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. We do, however, not examine which contract out of many feasible ones would be agreed upon given a particular motivation. Rather we stay with a given contract (as implicit in Figure 1; cf. footnotes 5 and 6), and consider the severity of the problems caused by hidden action in this case.

2.2 Incorporating Communication

We consider treatments that differ according to whether or not a pre-play communication opportunity is present. In the treatments that allow communication we let one player transmit a message to the other player. Figure 2 portrays a case in hand, where B sends a message to A.

FIGURE 2

[7] Independently of the contract-theoretic angle, we note that whether or not B's choice is observable to A may matter to the players' motivation (if they are not selfish; cf. sections 2.3 and 2.4). Perhaps B would feel worse choosing *Don't Roll* if she knew that A would know. We do not explore that interesting issue.

If the players are selfish, the communication stage obviously has no impact. Words alone can't change the subsequent payoffs, so (*Out, Don't Roll*) remains the unique subgame-perfect equilibrium for selfish players. On the other hand, if other concerns motivate the players, perhaps communication will matter. This is what we consider in the next two subsections.

2.3 Guilt aversion

In this paper we focus on guilt aversion, a motivation that provides a route by which communication may influence behavior. Before elucidating the connection with communication, we explain what guilt aversion is and how we test for it. This explanation has limited theoretical scope; it is specifically tailored to the game in Figures 1 and 2. In section 5 (after seeing the supportive experimental results) we develop a general theory of guilt aversion, and discuss how that framework relates to the 'pre-theoretic' notions here.

By guilt aversion we mean that a decision-maker suffers from guilt to the extent he believes he hurts others relative to what others believe he will do, and he tends to avoid such choices. Thus a guilt-averse decision maker is motivated by his *beliefs about others' beliefs*. Specifically, with reference to Figure 1, let τ denote the probability with which B chooses *Roll*; let τ' denote A's expectation of τ; let τ'' denote B's expectation of τ'. One may think of τ as a measure of trustworthiness, τ' as a measure of trust, and τ'' as a measure of expected trust. The higher τ'' is, the more B believes *Don't Roll* hurts A relative to what B believes A believes A will get. The higher τ'' is, the more guilt B hence experiences if he chooses *Don't Roll*. The game in Figure 3 models this.

FIGURE 3

$\gamma_B \geq 0$ is a constant measuring B's sensitivity to guilt. If B is rational he will choose to *Roll* if $14 - \gamma_B\tau'' < (5/6)\cdot 12 + (1/6)\cdot 0 = 10$. Note that the lower τ'' is, the higher γ_B must be for this inequality to hold, and vice versa.

To derive a testable prediction, we shall assume that the guilt sensitivity differs among subjects and is independent of τ".[8] In this case, the higher τ" is, the greater the likelihood that B will choose *Roll*. This is the most important research hypothesis of this paper. A test requires us to somehow observe τ", and our design is set up to achieve this. (We explain how in section 3.)

The conclusions drawn here do not presume that both players have coordinated on some 'equilibrium', but refer only to the motivation of *one* particular player and properties of his utility function. In section 5 we develop a notion of guilt-aversion equilibrium, and critically examine the status of that concept relative to the arguments made here.

Guilt aversion provides a route by which communication may influence behavior. For example, by making a promise to choose *Roll*, B may strengthen A's belief that B will choose *Roll*. This may be plausible, because if B believes that A's belief that B will choose *Roll* is strengthened by the promise, then this strengthens the incentives for B to choose *Roll* (as the guilt associated with *Don't Roll* goes up). On balance, the promise may induce the parties to play (*In, Roll*) rather than (*Out, Don't Roll*).[9]

Our experimental design allows us to explore the empirical relevance of such matters. In the treatments with communication, we observe what messages people choose to transmit, and how these words may move the beliefs and motivation of guilt-averse players. We check, in particular, whether promises or other statements of intent move beliefs and behavior.

Guilt aversion meshes well with findings in social psychology. See, *e.g.*, Baumeister, Stillwell & Heatherton (1994, 1995) who (on the basis of autobiographical narratives) suggest that people suffer from guilt if they inflict harm on others. Although guilt could have a variety of sources, one prominent way to inflict harm is to let others down. Baumeister *et al* (1995, p. 173) write that "Feeling guilty [is] associated with ... recognizing how a relationship partner's

---

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

[9] 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.

standards and expectations differ from one's own". In economic theory, some applied theoretical work by Huang & Wu (1994) (on remorse in corruption) and by Dufwenberg (2002) (on guilt in marriage) considers related ideas for specific trust games. Original to us is the link to communication, and the extension to general games (in section 5).

Guilt aversion has not previously received much attention by experimentalists. Dufwenberg & Gneezy (2000) and Bacharach, Guerra & Zizzo (2002) collect data on second-order beliefs and test some related hypotheses, but do not examine the impact of communication. Hannan, Kagel & Moser (2002) and Charness & Rabin (forthcoming) consider the impact of requests or expressed hopes on responder behavior, but do not elicit beliefs or consider the idea of guilt aversion.

2.4 Possible confounds

We focus on guilt aversion, not because it is the only motivation that allows communication to play a role, but because it is what we had in mind from the outset and (in our reading of the data) it also provides the best explanation of our results. Here we discuss a few alternative theories, which *a priori* might have been expected to perform well. The insights sketched here will also be mentioned again in the experimental section.

Researchers have begun to develop models of 'social preferences' 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 hiring practices.[10] It is natural to ask if these models, where people care about others, can capture the impact of communication; a little reflection indicates that many can't. A striking example is the class of *distributional models*, which define preferences strictly according to the overall monetary distribution of payoffs.[11]

[10] For descriptions of the experimental evidence and the 'social-preferences' literature it has inspired, see Fehr & Gächter (2000), Fehr & Schmidt (2002), and Sobel (2001). Guilt aversion is not covered in these texts.

[11] See, *e.g.*, the Fehr & Schmidt (1999), Bolton & Ockenfels (2000) inequity-aversion models, as well as the basic Charness & Rabin (2002) model.

Such models have the virtue of being simple to apply and test, precisely because they abstract from motivational forces that are not distributional. Despite their simplicity, they have proven successful in organizing much experimental data; however, explaining the impact of communication may be beyond them. In our games words alone can't change the payoff distributions that result from various strategy profiles, so pre-play communication can't change whatever backward induction solution applies (much like in the case with selfish players).[12]

A trickier case is *(kindness-based) reciprocity*, the idea that decision-makers wish to be kind (unkind) to those they believe to be kind (unkind). The classic reference is Rabin (1993), and Dufwenberg & Kirchsteiger (2004) extend his normal-form analysis to extensive-form analysis, appropriate for games such as in Figure 1.[13] Their notion of *sequential reciprocity equilibrium* yields the same (unique) prediction in the games of Figures 1 and 2,[14] and so cannot explain the impact of communication for these cases.

This conclusion, however, is somewhat unfair to the reciprocity model. The prediction of positive correlation between τ'' and the likelihood of a *Roll* choice, which we derived in section 2.3 assuming guilt aversion, invoked no equilibrium supposition. Reasoning analogously for the reciprocity model (referring only to B's motivation) one can show that a *negative* correlation between τ'' and the likelihood of a *Roll* choice is predicted.[15] Note that the sign of

---

[12] Three caveats are warranted: (i) We do not claim that models of inequity aversion do not permit communication to matter in other games. They do, in game with simultaneous moves and multiple equilibria, where communication may facilitate coordination. (ii) Communication could 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, Klein & Schmidt, 2001), then A's behavior could be affected by a message that may signal B's 'type'. However, this could never affect B's behavior in the subgames we consider.

[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 & Gjerstad (2004).

[14] The sequential reciprocity equilibrium depends on B's 'sensitivity to reciprocity' Y_B (see Dufwenberg & Kirchsteiger): B chooses *Don't Roll* whenever $Y_B \leq 4/45$; B chooses *Roll* with probability $(45Y_B-4)/20Y_B$ whenever $Y_B \in [4/45, 4/25]$; B chooses *Roll* whenever $Y_B \geq 4/25$. Details are available on request.

[15] The reciprocity model is complicated enough that it is hard to briefly explain why, but we give an indication here: Kindness and believed kindness depend on what a person *believes* he accomplishes and what he *believes others believe* they accomplish. The kindness of A when he chooses *In* is positive, since *In* brings B a higher dollar payoff

this correlation is the *opposite* of that under guilt aversion. Moreover, the impact of communication is different: Suppose B promises to choose *In*; if B believes A believes B is honest ($\tau'' = 1$), then B views A as less kind than if $\tau'' < 1$. Thus B will have a stronger incentive to renege if he believes his promise is believed. Truth-telling is no longer self-enforcing!

A third approach to explaining how communication might matter is the idea that people do not like to mislead others, or have a *dislike for lying*. Gneezy (forthcoming) finds that people avoid lying, providing experimental support for this notion; Brandts & Charness (2003) find complementary support that people dislike being deceived. The idea can be modeled by allowing B to choose whether or not to make a promise, thereby selecting which of two subgames to enter. If the cost of lying is high enough, B issues a promise in the unique subgame-perfect equilibrium, creating a credible commitment to choose *Roll* so that A's best response is to choose *In*.

All in all, reciprocity and dislike for lying provide some scope for communication to have cutting power in our games. We shall return to these ideas towards the end of the experimental results section (in subsection 4.4), and compare their performance to that of guilt aversion.

3. EXPERIMENTAL DESIGN AND HYPOTHESES

In this section we describe and motivate our experimental design (3.1), and present several hypotheses to be tested (3.2).

3.1 Design

Several methodological concerns entered into our design. First, we chose one-shot interactions between participants, in order to eliminate any reputation or repeat-game effects. Second, as communication is at the heart of our study, we varied whether or not messages could be sent from one party to the other; in the interests of simplicity, we only allowed a single

---

than *Out*. However, the exact degree of kindness of *In* depends negatively on τ', since the lower is τ' the more A believes B will get. Now focus on B: The lower B's belief of τ', the kinder B believes A is. That is, the lower τ'' is, the kinder B believes A is and, *mutatis mutandis*, the more likely is B to choose *Roll* in return.

message from one party to the other. Third, as we are interested in whether some particular type of message is effective, we give the sender a blank piece of paper on which to write any (anonymous) message, instead of restricting the message space.

Participants were recruited at UCSB by sending out an e-mail message to the campus community. We conducted 15 sessions, three in each of five treatments. 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, for a total of 460 people; 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.

In each session, participants were referred to as 'A' or 'B' (as in the games of sect. 2). A coin was tossed to determine which side of the room was A and which side was B. 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 decisions.

In our first two treatments, we used exactly the game parameters displayed in Figures 1 and 2. In our first treatment, no messages were permitted; in the second treatment, each B had an option to send a non-binding message to A prior to A's decision concerning *In* or *Out*. B could decline to send a message by circling the letter B at the top of the otherwise-blank sheet. In both cases, the game was presented to each of the participants in the following table:

Table 1: Payoff Outcomes with a (5,5) outside option

	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**

In the treatment without messages, A chose *In* or *Out* in the first stage; in the message treatment, B's message was transmitted to A before the choice of *In* or *Out*. Next, B chose whether to *Roll* or *Don't Roll* a 6-sided die, without knowing A's actual choice. B's choice was made 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.[16] 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. After the decisions had been collected, a 6-sided die was rolled for each B; this was made clear to the participants in advance, to avoid the anticipated loss of public anonymity for B's who chose *Don't Roll*. This roll was determinative if and only if (*In, Roll*) had been chosen.

Our next two treatments were conducted after observing considerable effectiveness for communication. These treatments used exactly the game parameters displayed in Figures 1 and 2, *except* that the payoff vector was (7, 7) rather than (5, 5) in case A chose *Out*. These treatment may be seen as tests of robustness; in this case, the gap between A's expected payoff of 10 after (*In, Roll*) and A's reservation payoff is considerably smaller than before, making *In* presumably less attractive to A. Even though communication may be effective when large efficiency gains are available from a successful partnership, perhaps it is ineffective in this case. In each treatment, the game was presented to each of the participants in the following table:

Table 2: Payoff Outcomes with a (7,7) outside option

	A receives	**B receives**
A chooses *Out*	**$7**	**$7**
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**

Our fifth and final treatment was also conducted after observing the results in the initial treatments, and was designed to shed light on the explanatory power of the various models that permit communication to affect behavior. In this treatment, we use the (5,5) reservation payoffs of our first two treatments, but we change who gets to send the message so that A sends a

[16] 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. We are not aware of any case where a treatment effect found using the strategy method is not found when using the direct-response method.

message to B. Here dislike of lying (as discussed in the final paragraph of section 2.4) must be irrelevant for B's motivation, as he or she cannot meaningfully make a statement of intent.

A critical element of our design involves measuring beliefs, as these are crucial to the guilt-aversion story (and also to the reciprocity model; cf. section 2.4). 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*.[17] 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% that 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.

3.2 Hypotheses

Given the wealth of experimental data on social preferences, we presume that behavior will differ from the selfish paradigm. However, it is an open question whether communication will affect behavior, and if it does so, whether this relationship is consistent with guilt aversion. Our first null hypothesis is that communication is irrelevant:

[17] 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.

H1: *The possibility of communication will increase neither the proportion of A's who choose* In *nor the proportion of B's who choose* Roll. *(*In, Roll*) outcomes are not more common when messages are feasible.*

Second, whether or not communication is relevant, there may be no relationship between beliefs and behavior, despite the prediction of guilt aversion. Our null hypothesis in this case is:

H2: *There is no correlation between the B's expectation of A's expectation of* Roll *and the frequency of* Roll *choices.*

H2 will be tested against the alternative of positive correlation, in line with the discussion in section 2.3. In that section we also discussed the possibility that promises may affect beliefs, thereby providing a self-fulfilling expectation that these statements of intent will not be reneged upon. We test the following pair of null hypotheses against the just-described alternative:

H3a: *Neither the proportion of A's choosing* In *nor the proportion of B's choosing* Roll *will be affected if a sent message contains a statement of intent (a* promise*).*

H3b: *Neither A's beliefs about the likelihood that B's will* Roll *nor B's beliefs about A's beliefs that B's will* Roll *will be affected by statements of intent (*promises).

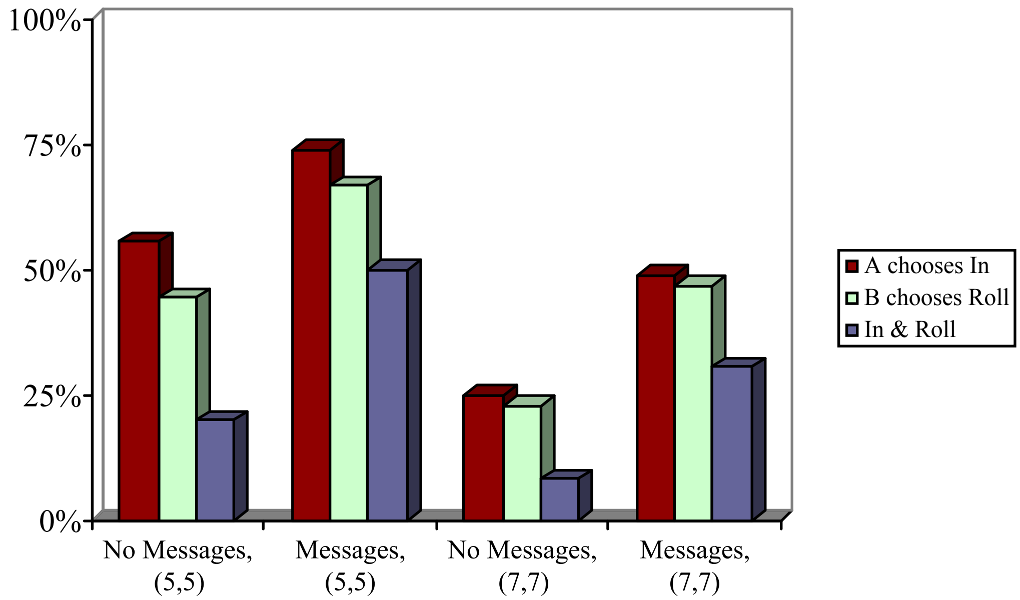
4. EXPERIMENTAL RESULTS

We find two main effects in our results. First, communication from B to A is quite useful in enhancing successful partnership formation and achieving efficient social outcomes; this is true for both of our payoff calibrations. Second, we observe a very strong pattern between B's choice and his belief about the expectation of A's expectation of his choice – the more B feels that A's who choose *In* expect 'trustworthy' behavior, the more likely it is that he will perform. We consider the effect of communication (H1) in section 4.1, beliefs and behavior (H2) in section 4.2, the effect of promises on beliefs and behavior (H3a and H3b) in section 4.3, and consider other possible explanations in relation to our data in section 4.4.

4.1 The effect of communication

Figure 1 summarizes choices with and without B messages in our payoff calibrations:

Figure 1 - The Effect of Messages from B



In the (5,5) treatment without B messages, 20 of 45 (44%) B's chose *Roll* and 25 of 45 (56%) A's chose *In*. When B could send a message to A, we observe considerably more cooperation: 28 of 42 (67%) B's chose *Roll* and 31 of 42 (74%) A's chose *In*. The (*In,Roll*) choice occurred 20% of the time (9 of 45 pairs) without communication, compared to 50% (21 of 42 pairs) with messages possible from B's.

We observe similar effects in the (7,7) treatment. Without B messages, 12 of 48 (25%) B's chose *Roll* and 11 of 48 (23%) A's chose *In*. When B could send a message to A, we once again observe considerably more cooperation: 24 of 49 (49%) B's chose *Roll* and 23 of 49 (47%) A's chose *In*. The (*In,Roll*) choice occurred 31% of the time (15 of 49 pairs) without communication, compared to 8% (4 of 48 pairs) with messages possible from the B's.

Thus, in both cases, A and B behavior was substantially different when communication from B to A was allowed and we see a much larger proportion of successful partnerships when B messages are feasible than when they are not. In all cases, behavior differs substantially from the predictions made by the selfish-preferences model.[18]

We can perform formal tests of our first hypothesis using the aggregate data provided in this subsection.[19] This hypothesis states that the possibility of communication will not affect behavior; our alternative hypothesis is that communication will improve rates of cooperative behavior. Table 3 summarizes the effect of communication on behavior for each of our three message treatments:

Table 3: Tests for the Effect of Communication

	A's *In* rate			B's *Roll* rate			(*In, Roll*)		
Treatment	M	NM	*Z*-stat	M	NM	*Z*-stat	M	NM	*Z*-stat
(5,5) B Messages	31/42 (74%)	25/45 (56%)	1.78**	28/42 (67%)	20/45 (44%)	2.08**	21/42 (50%)	9/45 (20%)	2.94***
(7,7) B Messages	23/49 (47%)	11/48 (23%)	2.48***	24/49 (49%)	12/48 (25%)	2.44***	15/49 (31%)	4/48 (8%)	2.76***
(5,5) A Messages	31/46 (67%)	25/45 (56%)	1.16	18/46 (39%)	20/45 (44%)	-0.51	12/46 (26%)	9/45 (20%)	0.69

M/NM mean that messages/no messages were feasible. The *Z*-stat reflects the test of proportions for the two populations (see Glasnapp and Poggio 1985). **and *** indicate $p < 0.05$ and 0.01, respectively, one-tailed tests.

[18] Figure 1 also illustrates that 'cooperative' choices are more frequent when the available outside option is (5,5), rather than (7,7). While this is not at all surprising for A, who is taking a bigger chance by choosing *In* when the outside option is (7,7), the significant ($Z = 3.23$ and $Z = 2.60$ for the respective no-communication and B-message comparisons, both significant at $p = 0.01$) differences in A behavior show that the A's indeed appear to be responding to incentives. B's are also substantially less likely to choose *Roll* in both comparisons, although the effects are less pronounced ($Z = 1.97$ and $Z = 1.70$ for the two comparisons, still significant at $p = 0.05$ and $p = 0.09$, respectively, two-tailed tests). It seems that the smaller gain to B resulting from A choosing *In* negatively impacts the urge to *Roll*.

[19] Nearly all of our tests are conducted using nonparametric statistics. However, we also run probit regressions, which produce essentially the same conclusions as the nonparametric tests. These regressions are shown in Appendix D.

We can reject H1 in favor of the alternative hypothesis for both A's and B's whenever the communication takes the form of messages from B to A. We may conclude that B messages have a major influence on behavior and outcomes in this case.[20]

We mentioned earlier that we also conducted a (5,5) treatment in which A could send a message to B, in order to distinguish the effects of a dislike for lying from other behavioral motivations. In this case, communication was ineffective in improving the rate of cooperative behavior: 31/46 (67%) A's chose *In*, while 18/46 (39%) B's chose *Roll*; the (*In, Roll*) choice occurred 26% of the time (12/46 pairs). None of these rates differ substantially or significantly from the rates found in the (5,5) no-communication treatment, although A's are slightly more likely to choose *In* when A messages are permitted.[21] However, whether these non-differences are consistent with a dislike of lying is not settled from these data alone, and we shall return to this issue in section 4.4.

4.2 Beliefs and behavior

We have seen that communication affects behavior, in line with the results of some previous experimental studies involving cheap talk. While this is interesting, a key issue motivating our experimental design lies in the relationship between beliefs and choices. Specifically, guilt-aversion predicts a positive relationship between B's second-order beliefs and the likelihood that B will choose ROLL, contrary to our null hypothesis H2. If A's are responding to incentives, we would also expect that A's who expect B's to be more likely to *Roll* will be more likely to choose *In*. Table 4 details the observed relationship between beliefs and behavior in each of our treatments:

---

[20] We also note that agent messages affected A's guesses: In the (5,5) treatment, the average A guess was 59.4 when messages were feasible, compared to 41.0 in the no-message treatment. Similarly, in the (7,7) treatment, the average A guess was 56.9 when messages were feasible, compared to 32.7 in the no-message treatment.

[21] In fact, 29 of the 37 (78%) A's who sent messages chose *In*.

Table 4: Beliefs and Behavior

	A's average guess			B's average guess		
Treatment	*In*	*Out*	*Z*-statistic	*Roll*	*Don't*	*Z*-statistic
(5,5) No Messages	51.3	28.2	2.55***	54.2	39.6	1.99**
(5,5) B Messages	65.4	42.5	2.02**	73.2	45.1	3.20***
(5,5) A Messages	56.7	35.4	2.65***	69.6	50.0	2.80***
(7,7) No Messages	35.7	31.8	1.06	69.4	41.7	3.08***
(7,7) B Messages	70.0	45.3	3.00***	66.9	36.9	3.52***

The *Z*-statistic reflects the Wilcoxon-Mann-Whitney ranksum test for the two populations compared (see Siegel & Castellan 1988). *, **, and *** indicate $p < 0.10$, 0.05 and 0.01, respectively, one-tailed tests.

We observe a strong correlation between beliefs and behavior, both for A's and B's. In each of the five treatments, A's who chose *In* made higher average guesses about the likelihood of *Roll*; in four of these cases, the difference is statistically significant in the conservative nonparametric test. Results for B behavior are even stronger: In *all* five treatments, B's who chose *Roll* made significantly higher guesses about A's guesses than did B's who chose *Don't Roll*. Thus, H2 is strongly rejected, as we find that a B who chooses *Roll* makes a substantially and significantly higher guess about A's guess than a B who chooses *Don't Roll*. We conclude that the support for guilt aversion is considerable in all of our treatments.[22]

4.3 Promises and beliefs

We have seen that beliefs differ substantially for people that choose different actions. Is there some particular aspect to messages that causes these beliefs to be so affected by communication? We focus on whether or not a message contains a 'statement of intent', or promise. Since messages can have nearly any form, this requires a classification of the

---

[22] It has been suggested to us (mainly by psychologists) that a "false consensus effect" (cf. Ross, Greene & House 1977) might also produce a positive correlation, if choices shape beliefs about beliefs rather than vice versa. This is intriguing, although we note that false consensus usually means that a person believes others would act similarly rather than that a person believes others believe he or she would make a certain choice. Cf. also Croson & Miller (2004), who test whether beliefs cause behavior or vice versa and conclude in favor of the former view.

messages. We use three rough categories: promises, empty talk, and no message; our classification is given in Appendix B, along with the raw data on individual choices.[23] Promises are only meaningful when coming from B (although one A nevertheless promised to choose In!), so we need only consider the two B-message treatments. The promises category is broad, including any 'statement of intent' that we found. To be sure, some messages were on the boundary between promise and empty talk and could arguably be placed in either category; nevertheless, the overall pattern is quite clear and is robust to alternative classifications.

Our null hypothesis H3a states that statements of intent will not affect behavior, while our alternative hypothesis is that such statements will make cooperative behavior more likely. Table 5 shows A and B behavior, according to whether a promise was sent or received:

Table 5: Promises and Behavior

	A's *In* rate			B's *Roll* rate			(*In, Roll*)		
Treatment	P	NP	*Z*-stat	P	NP	*Z*-stat	P	NP	*Z*-stat
(5,5) B Messages	22/24 (92%)	9/18 (50%)	3.04***	18/24 (75%)	10/18 (56%)	1.32*	16/24 (67%)	5/18 (27%)	2.49***
(7,7) B Messages	16/24 (67%)	7/25 (28%)	2.71***	20/24 (83%)	4/25 (16%)	4.71***	14/24 (58%)	1/25 (4%)	4.13***
Pooled	38/48 (79%)	16/43 (37%)	4.07***	38/48 (79%)	14/43 (33%)	4.49***	30/48 (62%)	6/43 (14%)	4.73***

P /NP mean that a promise/no promise was sent or received. The Z-stat reflects the test of proportions for the two populations compared. * ,**, and *** indicate $p < 0.10$, 0.05, and 0.01, respectively, one-tailed tests.

In all cases but one, the *In* rate, the *Roll* rate, and the *ex post* (*In, Roll*) realizations were much higher following a promise than otherwise. Note that N is fairly small here, as we split the

[23] 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 of the '(5, 5) Messages from B' treatment, which contains a poem by Samuel Francis Smith and fictitious references to desires and advice from some famous persons....

observations in each treatment into two categories; if we compensate for this by pooling the data from the two treatments, the differences in behavior are even more significant. Thus, we can strongly reject H3a.

Our null hypothesis H3b states that statements of intent will not affect beliefs, while the alternative hypothesis is that guesses will be higher with promises. Table 6 shows average A and B guesses in the B-message treatments, according to whether a promise was sent or received:

Table 6: Promises and Beliefs

	Average A Guess			Average B Guess		
Treatment	P	NP	*Z*-stat	P	NP	*Z*-stat
(5,5) B Messages	65.8 (24)	50.0 (18)	1.63*	66.2 (24)	59.9 (18)	1.10
(7,7) B Messages	63.1 (24)	50.9 (25)	1.44*	59.6 (24)	51.0 (25)	1.17
Pooled B Messages	64.4 (48)	50.5 (43)	2.24**	63.1 (48)	54.7 (43)	1.74**

P/NP means that a promise was sent or received; N is in parentheses. The Z-stat reflects the Wilcoxon ranksum test for the two populations. * and ** indicate $p < 0.10$ and 0.05, respectively, one-tailed tests.

In every case, guesses are highest when a promise is made, but no within-treatment test is more than marginally significant. Once again, N is fairly small for these tests, and if we pool the data from the two treatments to increase the sample size, we do see results that are significant at $p = 0.05$ on the indicated one-tailed test. Thus, the evidence tends to go against H3b, with promises affecting beliefs, but the effects are modest: In the pooled B-message treatments, A guesses after promises are 27.5% higher than after non-promises, while B guesses after promises are 15.4% higher than after non-promises.

4.4 Tests of possible confounds

In section 2.4, we mention some other behavioral models that predict deviations from pure own-money maximization. Recall that the distributional models all predict that communication will have no effect on behavior, our first null hypothesis. Since our data from treatments with B messages readily reject this hypothesis, they also reject these models as an explanation for the effect observed.

Regarding (kindness-based) positive reciprocity, in section 2.4 we derived a testable prediction of negative correlation between the likelihood that B performs and B's beliefs about A's beliefs about the likelihood of B performing. But we have just seen that in fact this correlation is significantly *positive*; thus, we can reject this alternative reciprocity hypothesis to H2 with one-tailed tests of even greater statistical significance.[24]

Finally, we consider the notion that people have a *per se* dislike for lying. Our treatment in which A can send B a message is an attempt to bring a tighter focus on the influence of a dislike of lying on behavior in our setting since, when A's send messages to B's, there is no obvious sense in which a dislike of lying could be a factor. If we observed that these messages facilitated successful partnerships, we could conclude that a dislike of lying is not a factor. However, we have seen that messages from A's lead to nearly the same behavior as in the no-communication treatment, so that we cannot clearly rule out a dislike of lying as a factor in our treatments with B messages.

As this does not settle the matter, we must delve more deeply into the data. If a dislike of lying is present, we might expect a difference in *Roll* behavior across the A-message and B-message treatments in the (5,5) outside-option case, taking into account B's guess. To test this, we perform a probit regression using only the data from these two treatments, with *Roll* as the

[24] The *p*-values for the five rows in Table 5 are 0.023, 0.001, 0.003, 0.001, and 0.0002, respectively. Recall also (from section 2.4) that the differently-construed testable prediction of sequential reciprocity equilibrium implied that communication has no impact. We rejected that conclusion in section 4.1.

dependent variable and a dummy variable for A messages and another dummy variable for the interaction between A messages and B guesses. Standard errors are in parentheses:

$$\underset{}{\textit{Roll}} = \underset{(.752)}{-1.869} \; \underset{(.011)}{+\, 0.037^{*}\text{Guess}} \; \underset{(.991)}{-\, 0.54^{*}\text{A\_message}} \; \underset{(.015)}{-\, 0.010^{*}\,\text{A\_message}^{*}\text{Guess}} \qquad (6)$$

We see that B's guess is highly relevant for B's decision whether to Roll, but that there is no difference across treatments, reflected in the insignificance of the coefficient of the terms with an A_message dummy, in both the constant term and in the 'slope' term (marginal effect) for a change in B's guess. As this indicates that, holding beliefs constant, B's in the B-message treatment are no more likely than B's in the A-message treatment to Roll, it is indirect evidence that a simple dislike of lying is not a major factor in our data. Nevertheless, it seems plausible that guilt aversion and a dislike of lying are both present to some degree, and perhaps complement each other in some manner.

5. GUILT IN GAMES IN GENERAL

Experiments may not only test preconceived ideas, but also inspire the development of new theory. We had guilt aversion in mind from the outset when we embarked on this study, but our conjectures were particular to the trust games of our design. Having established some empirical support for guilt aversion in that context, it is natural to wonder about the implication of guilt aversion generally. In this section, building on the GPS framework we first define a notion of guilt-aversion equilibrium, which applies to general extensive game forms (5.1). In order to indicate the usefulness of this concept, we then apply it to our trust games, a dictator/tipping game, a gift exchange game, a Cournot game, and poker (5.2). Finally, we point out some limitations of the notion of guilt-aversion equilibrium (5.3). This last part highlights some needs for future research. In this connection, we also explain why we elicited beliefs as we did and why our hypotheses tests referred to properties of utility functions rather than the concept of guilt-aversion equilibrium.

5.1 Guilt-aversion equilibrium

While in standard game theory players' payoffs (or utilities) depend only on actions, in psychological game theory they also depend on beliefs. Modeling guilt aversion requires the latter framework. Our discussion in section 2.3 illustrates this: When B makes a choice in Figure 3 he knows which strategy profile is played, and yet his preference depends on something more, namely the second-order belief τ''.

GPS provide a general framework, which they call *psychological game theory*, which can describe belief-dependent preferences. A few studies have used psychological game theory to model motivation (more or less) reminiscent of guilt aversion in specific trust games,[25] but no theory exists for general games. Our objective in this section is to answer the following question: What kind of equilibrium notion do we get if use the GPS framework to incorporate guilt aversion in general extensive games?

We first present GPS' class of extensive psychological games, trusting the reader to be familiar with standard mathematical machinery of extensive games (else, refer to section 3 of GPS, or to a game theory textbook). We henceforth drop the qualifier "extensive", which should be understood whenever we talk of game forms, games, or psychological games.

Let F be a finite game form (*i.e.*, F consists of all components that make up a game *except* a payoff function and there are finitely many players each with finitely many strategies). Let N be the associated player set, M_i the set of behavior strategies for player i, $M = \times_{i\in N} M_i$, and $M_{-i} = \times_{j\neq i} M_j$. Next, define a vector of utilities, $(u_i)_{i\in N}$, such that

(1) $u_i: M \times B_i \rightarrow \mathrm{I\!R}$,

where B_i is the set of i's possible beliefs, before the start of play, about other players' strategies and beliefs, and where u_i is derived from a utility function with domain $Z \times B_i$, where Z is the set

[25] See Huang & Wu (1994), Dufwenberg & Gneezy (2000), Dufwenberg (2002), Bacharach *et al* (2002).

of endnodes of F, by taking expectations. A *psychological game* is a pair $\Gamma = (F, (u_i)_{i\in N})$. Its distinguishing feature, relative to a standard game, is that the domain of each u_i includes i's beliefs, not only strategy profiles. The structure of B_i is analogous to a belief hierarchy in the literature on incomplete information games (e.g. Mertens & Zamir, 1985, with $B = \times_{i\in N} B_i$ being a so-called "universal type space". A psychological game may seem complicated because B_i is a complicated object.

However, we need not concern ourselves with details here because our focus on guilt aversion permits us to zoom in on a simple special case where only a very particular belief (derived from the relevant element of B_i) influences i's utility. Take as primitives a game form F and a vector $(v_i)_{i\in N}$ of functions $v_i: M\rightarrow \mathbb{R}$, where v_i is interpreted as a 'monetary' payoff function which records i's dollar earnings (v stands for "*v*alue"). The pair $(F, (v_i)_{i\in N})$, is formally a game, but we refrain from analyzing it using traditional game-theoretic tools because in our model v_i does not represent i's preferences. $(F, (v_i)_{i\in N})$ should merely be thought of as a partial description of a situation, including its strategic possibilities and the associated monetary payoffs; i's preferences, however, depend on guilt as well as monetary earnings. Specifically, in words, if i chooses m_i, then i suffers from guilt to the extent that i believes that j gets a lower monetary payoff than j would get if i chose what *i believes that j believes that i chooses*.

To formulate this mathematically, and to thus derive the relevant psychological game, let $c_{iji}\in M_i$ (where c_{iji} is derived from b_i) represents i's belief about j's belief about which mixed strategy i is choosing.[26] (The notation is inspired by Dufwenberg & Kirchsteiger, 2004.) Now define a specific functional form of (1), as follows:

(2) $$u_i(m, b_i) = v_i(m) + \gamma_i \bullet \Sigma_{j\neq i} \min\{0, v_j(m) - v_j(c_{iji}, m_{-i})\},$$

[26] In GPS' theory, $b_i \in B_i$ specifies beliefs of all orders, and the entailed beliefs about beliefs about choices outnumber those describable via c_{iji} because c_{iji} is derived only from the first-order moments and neglects higher-order moments. Thus c_{iji} embodies a hypothesis regarding which elements of the second-order beliefs are motivationally relevant. Rabin (1993) and Dufwenberg & Kirchsteiger (2004) make analogous assumptions.

where $\gamma_i \geq 0$ is a parameter which measures *i*'s 'sensitivity' to guilt. Equation (2) decomposes *i*'s utility into a monetary and a psychological part. The latter sums up all the guilt *i* incurs by making choices that, according to *i*'s beliefs, fail to meet the expectations of his co-players.

We will for the most part limit attention to multi-stage game forms with observed actions,[27] a class that is easy to describe and yet large enough to cover most experimental or applied work. For such structures, GPS' notion of subgame-perfect psychological equilibrium is natural, and we shall draw on it to develop our own solution concept. A definition requires some notation: For any $m \in M$, let $\beta(m) = (\beta_i(m))_{i \in N} \in B$ be the belief hierarchy where everyone assigns probability 1 to *m* being played, everyone assigns probability 1 to everyone assigning probability 1 to *m* being played, …etc, *ad infinitum*. In other words, $\beta(m)$ describes the players' beliefs when *m* is 'common knowledge'. Given a game form *F* and a strategy profile $m \in M$, let $\Gamma(m) = (F, (w_i)_{i \in N})$ be the game that has payoff function w_i: $M \to \mathrm{IR}$ defined by $w_i(m') = u_i(m', \beta_i(m))$ for all $m' \in M$. (Note: $\Gamma(m)$ is a standard game, not a psychological game.)

DEFINITION 1: The strategy profile $m \in M$ of the psychological game Γ is a *subgame-perfect psychological equilibrium* if *m* is a subgame-perfect equilibrium of $\Gamma(m)$.[28]

Our solution concept for guilt-averse players draws on Definition 1:

DEFINITION 2: Consider a pair $(F, (v_i)_{i \in N})$, where *F* is a finite game form and $(v_i)_{i \in N}$ is a vector of monetary payoff functions, as well as the associated utilities $(u_i)_{i \in N}$ as defined in (2). The strategy profile $m \in M$ is a *guilt-aversion equilibrium* of the psychological game $\Gamma = (F, (u_i)_{i \in N})$ if *m* is a subgame-perfect psychological equilibrium of $\Gamma(m)$

[27] These are games in which play proceeds in 'stages', and in which all instances of imperfect information are due to simultaneous moves. See Fudenberg & Tirole (1991, chapter 3).

[28] In GPS' original definition, equilibria are belief-strategy pair profiles, while Definition 1 concerns strategy profiles. We find it simpler to keep beliefs implicit, but the difference is otherwise immaterial.

EXISTENCE*:* Drawing on the GPS results, we can easily derive an existence result for guilt-aversion equilibrium: GPS prove that if a game form F is finite, and the utility u_i, as specified in (1), is continuous, then the psychological game (F, $(u_i)_{i\in N}$) must have a subgame-perfect psychological equilibrium. In Definition 2, F is finite, u_i is a special case of (1), and u_i is continuous. Since a guilt-aversion equilibrium is a subgame-perfect psychological equilibrium, existence follows.

To determine whether $m'\in M$ is a guilt-aversion equilibrium of $\Gamma = (F, (u_i)_{i\in N})$ one must check two things: (i) c_{iji} (which is the only belief appearing in the right-hand-side of (2)) must be derived from $\beta(m')$, so that c_{iji}=m_i' for all i,j; (ii) m' must be a subgame-perfect equilibrium in $\Gamma(m')$, *i.e.* the (standard) game that is induced once m_i' is substituted for c_{iji} in the right-hand-side of (2) for all i,j.

5.2 Applications

We next visit several examples that illustrate how guilt-aversion equilibrium works, and that we hope indicate the general usefulness of the concept. We give particular emphasis to the role played by communication.

EXAMPLE 1. Consider first the game in Figure 1. For simplicity, we assume that $\gamma_A = 0$. Using (2), we can derive the associated guilt-aversion equilibria. Recall from section 2.3 that τ denotes the probability with which B chooses *Roll*; let τ' and τ'' denotes the corresponding initial first- and second-order expectations. The players' payoffs are just as depicted in Figure 1, *except* that B's utility given the profile (*In*, *Don't Roll*) depends on τ'', as follows (cf. (2)):[29]

(3) $u_B((In, Don't\ Roll), b_B) = 14 + \gamma_B\bullet\min\{0, 0-[(1-\tau'')\bullet 0 - \tau''\bullet(1/6\bullet 0+5/6\bullet 12)]\}$

$= 14 - \gamma_B\bullet 10\bullet\tau''$

[29] To see that payoffs remain as given by v_A and v_B elsewhere, note first that A's case follows from $\gamma_A = 0$. As regards B, refer to (2) and note that if m is such that B chooses *Roll* then $\min\{0, v_A(m) - v_A(c_{iji}, m_{-i})\}=0$.

Thus, the resulting psychological game has the same depiction as the game in Figure 3, except that it should say $14 - \gamma_B \cdot 10 \cdot \tau''$ rather than $14 - \gamma_B \cdot \tau''$. To describe the guilt-aversion equilibria, note first that if $\gamma_B < 0.4$, then $14 - \gamma_B \cdot 10 \cdot \tau'' > 10$, so a rational B must choose *Don't Roll* as the guilt feelings incurred cannot outweigh the material gain. In this case (*Out*, *Don't Roll*) is the unique equilibrium (entailing $\tau'' = \tau' = \tau = 0$). However, there are multiple guilt-aversion equilibria if $\gamma_B \geq 0.4$, in particular:

1. (*Out*, *Don't Roll*)

2. (*In*, *Roll*)

The first equilibrium entails that $\tau'' = \tau' = \tau = 0$; the second that $\tau'' = \tau' = \tau = 1$.[30] Note that *both* A and B gain in the second equilibrium relative to the first. In the second equilibrium, B rationally chooses *Roll* in light of the guilt that would plague him otherwise. Along the equilibrium path, no guilt is felt. A happy ending is the result, reflecting on Leith & Baumeister's (1998) assertions that "guilt serves many adaptive, beneficial, and prosocial functions" (p. 1), and that "guilt helps strengthen and maintain close relationships" (p. 2).

It is time to stress the key point we wish to make: *communication may help bring about the second equilibrium.* For example, suppose B says "I promise to choose *Roll*". If B believes that A believes him, this will make B *more* inclined to choose *Roll*. This in turn gives A a reason to believe B's statement. For a guilt-averse B, truth-telling can thus be self-enforcing. By issuing a promise, B may gain commitment power regarding the exercise of his *Roll* choice.

Note the structure of our approach here: In section 5.1 we defined guilt aversion equilibrium for multistage games with observable actions. The game in Example 1 belongs to that class, as do augmented versions that add pre-play messages. We compare the equilibria of games with messages to the equilibria of the game without messages, and focus on behavior in

---

[30] There are also equilibria in mixed strategies. Let σ denote the probability with which B chooses *In*. The additional equilibria are: **3.** $\gamma_B \in [0.4, 0.8]$ and $(\sigma,\tau) = (1, 0.4/\gamma_B)$ with $\tau''=\tau'= 0.4/\gamma_B$; **4.** $\gamma= 0.8$ and $(\sigma,\tau) = (\sigma, 0.4/\gamma_B)$ with $\sigma \in [0,1]$ and $\tau''=\tau'=0.4/\gamma_B$; **5.** $\gamma_B \geq 0.8$ and $(\sigma,\tau)=(0,4/\gamma_B)$ with $\tau''=\tau'=0.4/\gamma_B$.

the post-communication subgames. There is a multiplicity of equilibria even without messages, which implies a vast multiplicity of equilibria with messages added. Some of these have the form that if a 'sufficiently strong' promise is made then players coordinate on something 'good', and otherwise they coordinate on something 'bad', so in equilibrum a sufficiently strong promise is made. Adding messages to the game in this way does not eliminate the 'bad' equilibria, but we believe it helps to highlight those 'good' equilibria that communication may bring about. Our approach for analyzing the examples that follow is analogous.

EXAMPLE 2. Our second example concerns tipping. Consider the following story:

> Björn feels guilty if he lets others down. In restaurants, this influences his tipping. The more he believes that his waitress believes she will receive as a tip, the more he tips. More precisely, he 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'.)

Besides depicting something arguably realistic, the story illustrates in the starkest possible way that classical game theory cannot adequately model guilt aversion. 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 $\{\omega\}$). Björn's choice of tip determines a full strategy profile. In classical game theory, payoffs are defined 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.

By contrast, guilt-aversion equilibrium handles the story nicely. To make this clear, let $t \in \{0, 1, ..., x\}$ denote Björn's tip, with x being the number of dollars in his wallet. Let t' denote the waitress's expectation of t; let t'' denote Björn's expectation of t'. In any pure-strategy guilt-aversion equilibrium Björn chooses t to maximize (4) (cf (2)):

(4) $u_{\text{Björn}}((t, \omega), b_{\text{Björn}}) = (x - t) + \gamma_{\text{Björn}} \cdot \min\{0, t - t''\}$,

If $\gamma_{\text{Björn}} > 1$, Björn's optimal choice is $t = t''$, which corresponds to a guilt-aversion equilibrium. Note further that communication may matter to the outcome. For example, the waitress may say something before Björn tips to influence his beliefs about what she expects.[31]

This example also illustrates a way in which guilt aversion is different from reciprocity. In the example, Björn gives away money even though the only other player has no way of influencing the strategy profile played. In most models of reciprocity (see Rabin 1993, Dufwenberg & Kirchsteiger 2004) that is impossible; a player may choose to be kind only if others are actively kind in return.

EXAMPLE 3. Guilt-aversion equilibrium may shed light on results in gift-exchange games, sometimes called wage-effort games since they may depict an employer who pays a wage followed by an effort choice by a worker in return. In experimental tests, a positive wage-effort relationship is often found, which is often taken to illustrate reciprocal forces at work (see Fehr & Gächter, 2000, for references and a discussion).

Guilt aversion provides an alternative route to explaining such behavior. To make this precise, consider a game (as in Brandts & Charness 2004) where first the employer choose a wage $w \in \{0, 1, \ldots, 10\}$, and then the worker observes the wage choice and chooses an effort $e \in \{0, 1, \ldots, 10\}$. The resulting monetary payoffs are:

$2 \cdot e - w$ for the employer, and

$2 \cdot w - e$ for the worker

Suppose the worker is very guilt averse. Consider the strategy profile where the employer offers a wage of 10, and the worker responds to any w with the choice $e = w$. This strategy profile is a guilt-aversion equilibrium: In equilibrium the beliefs are correct. Guilt

[31] 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%."

aversion supports each of the worker's effort choices, including in particular e = 10 in response to w = 10. On the other hand, there are guilt-aversion equilibria where low effort choices are met in response to any wage, and low wages are then offered in the first place. No one expects much, no one delivers, no one feels guilt.

Which of these equilibria should one expect to prevail? We wish to make the point that communication may improve the prospect of a good equilibrium. There is actually some experimental evidence that we take to indicate the empirical relevance of this claim. In some gift-exchange studies, a form of one-sided communication by 'firms' appears: Firms offer contracts consisting of wage and a *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, then statements of desired effort may become self-fulfilling for guilt-averse decision-makers.

EXAMPLE 4. Analyses of collusion in oligopoly usually draw on repeated game theory. Guilt aversion furnishes a complementary reason why cooperative pricing may be sustained, even in one-shot market games. We illustrate with a simple Cournot example; the game has infinitely many strategies, but the extension from Definition 2 should be clear.

Each of two competitors has the strategy set [0, 1] and a monetary payoff given by

$$x \bullet (1 - x - y),$$

where x is the own choice and y is the competitor's choice. The unique Nash equilibrium for selfish players is (1/3, 1/3). However, there are multiple guilt-aversion equilibria including (1/4, 1/4), which corresponds to the monopoly outcome. One can show that the lower γ_1 and γ_2 are, the closer to the usual Cournot equilibrium (1/3,1/3) the outcome must be: the strategy profile (z, z) with $0 < z \leq 1/3$ is a guilt-aversion equilibrium only if $\gamma_i \geq (1-3z)/z$ for i = 1,2.

This analysis may suggest a reason for competitors to meet and discuss ways to collude, even if repeated play is not an issue. At first glance such meetings may seem pointless, since cartel agreements are illegal and therefore not enforceable in court. However, an 'oral contract' may help guilt-averse cartelists coordinate on a good guilt-aversion equilibrium. This may in turn provide a rationale for anti-trust clauses that deem it illegal for company representatives to meet and discuss pricing (see *e.g.* Hovenkamp, 2000, for a discussion of such laws).

EXAMPLE 5: Our final example is poker.[32] It is hard to imagine poker players feeling guilty; we argue that this is nevertheless consistent with guilt-aversion equilibrium. In poker, players are (clearly!) expected to maximize their own earnings. This is close enough to minimizing other players' earnings (it is equivalent in two-player poker). As it is impossible to let others down, the second term in the right-hand-side of (2) must equal 0. Hence if each player maximizes his own earnings given the others' strategies, this is a guilt-aversion equilibrium.

A striking and intriguing feature of this example is that it supports guilt aversion vis-à-vis dislike of lying as a motivating factor (cf. our earlier discussion in section 4.4). Dislike of lying is *not* a factor in poker. In fact, leading poker texts actively encourage lies, or at least very deceptive use of language and demeanor. For some colorful testimony, we refer to several examples in Brunson (2002); see *e.g.* pp. 427-8, 80-1, 88-9, 105-6.

5.3 Limitations

Guilt-aversion equilibrium is our answer to the question: What kind of equilibrium notion do we get if use the GPS framework to incorporate guilt aversion in general extensive games? Although the examples in section 5.2 were all meant to suggest that guilt-aversion equilibrium is a useful concept, in this subsection we will discuss three important limitations: (i)

[32] Any table-stakes version of poker is a finite game which can be modeled using Definition 2. Because of the initial chance move ("the deal"), the game form is not multi-stage with observed actions, but we never formally invoked that restriction and Definition 2 can still be applied although there are no subgames.

the restriction to equilibrium analysis may be unreasonable, (ii) the GPS framework is too restrictive, and (iii) the focus on guilt-aversion is too narrow. Each of the limitations leads to a proposal for further research. We discuss these issues in turn.

Assuming equilibrium is assuming too much

Guilt-aversion equilibrium embodies assumptions about motivation *and* assumptions that guarantee that people hold correct expectation about one another. Such equilibrium suppositions are commonplace in economic theory, and maintaining that parallel may help facilitate comparisons of different models. However, it is important to note that people may be rational even if they fail to coordinate on an equilibrium. In conventional game theory, such reasoning has inspired work on, *e.g.*, rationalizability (see Bernheim 1984, Pearce 1984).

There is no reason to assume that equilibrium coordination is easier in psychological games than in standard games. In constructing experimental tests of guilt aversion, one may wish to isolate tests of basic psychological assumptions from tests regarding whether or not people have correct beliefs. Otherwise, one runs the risk of incorrectly rejecting a valid insight about motivation, just because people did not coordinate on an equilibrium. The tests we devised in section 2.3 are based on this outlook, which explains why we referred to guilt-averse players' utilities rather than the notion of guilt-aversion equilibrium.[33]

The remarks made here entail a suggestion for future research, concerning psychological game theory generally. The solution concepts developed by GPS, as well as extant applications of psychological game theory, almost exclusively presume equilibrium behavior. The time may be ripe to escape this straightjacket, and Battigalli & Dufwenberg (2004) take early steps.

[33] Another difference is that the theory of section 5.1 presumes that the players' guilt sensitivities are commonly known, while the hypothesis derived in section 2.3 implicitly presumed that only B's knows his or her own guilt sensitivity. Incomplete information about guilt sensitivity seems empirically plausible but theoretically less tractable (since there may be signaling issues; we avoid this in the experiment by merely measuring (rather than explaining) τ''.

The GPS framework is too restrictive

Guilt aversion equilibrium is based on GPS' notion of subgame-perfect psychological equilibrium, which presumes that only *initial* beliefs may enter the domain of a player's utility. This is too restrictive for many useful purposes (as GPS themselves note on p. 78). For example, in our trust games, one could argue that the decision by A to choose *In* 'signals' something about A's beliefs, and that a guilt averse B should take that into account and let his *updated* belief, not the initial pre-play belief, influence the guilt feelings caused by a *Don't Roll* choice. Indeed, our experimental design actually caters to this; our belief elicitation protocol ask B to guess the guess only of A's that chose *In*, which amounts to measuring a updated, or conditional, belief. Dufwenberg (2002) explores related themes, and his analysis of "psychological forward induction" shows how this may have a dramatic effect on the analysis. Another example is reciprocity theory, in which updated views of whether or not a co-player is kind calls for updated beliefs to enter the utility functions domains. Such concerns lead Dufwenberg & Kirchsteiger (2004) to deviate from the framework offered by GPS.

There is a need for basic research regarding more general classes of psychological games that allow belief-dependent motivation for other than initial beliefs. To construct such a general theory is somewhat complicated, because one has to deal with complicated hierarchies of conditional beliefs. These structures were not well-understood in the late 80's when GPS wrote their article, which may explain their focus on hierarchies of initial beliefs. Recently, however, Battigalli & Siniscalchi (1999, 2002) have made progress on how to describe hierarchies of conditional beliefs. Battigalli & Dufwenberg (2004) attempt to draw on these insights to prove general results for a larger class of extensive psychological games than that considered by GPS.

The focus on guilt-aversion is too narrow

Guilt-aversion equilibrium embodies two motivational forces: concern for own monetary reward, and aversion to guilt. We believe this simplification may deliver useful insights, but we do not pretend that other motivational forces are irrelevant. We will not make a list of every conceivable motivational force we have neglected, but we will mention the one we feel may be the most important: *negative* (kindness-based) reciprocity.

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. To the extent that we dismissed reciprocity in section 4.3, that evidence only concerns the positive side; in our trust games, the only way A can be unkind is by choosing *Out*, in which case B gets no shot at taking revenge. However, negative reciprocity may be important in other games; rejection in ultimatum games is a convincing example. Guilt aversion cannot, however, explain such rejections.

6. CONCLUSION

Conventional contract theory, as applied to many one-shot games, implicitly assumes that written contracts bind if supported by the law, while oral agreements (to quote Samuel Goldwyn) "aren't worth the paper they're written on". One may 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.

We examine the impact of communication on cooperation in a one-shot game designed to capture the essence of hidden action, as treated in much of contract theory. Our game has the form of a 'trust' game, but has a stochastic twist that preserves the flavor of hidden action. Our design allows us to preserve the richness of the message space, while maintaining the control of an anonymous (not face-to-face) interaction.

While we observe some degree of cooperation even without communication, successful partnership formation is greatly enhanced (for two different payoff calibrations) when we allow an agent to send a free-form message to a principal. In our hidden-action context, we find that the improvement in social outcomes is largely driven by promises (statements of intent) made by the would-be performer. There is also a strong positive relationship between performance and the performer's beliefs about the beliefs of his counterpart about anticipated performance.

These results mesh well with the notion of guilt aversion, according to which the agent lives up to his promise because he believes the principal believes it. We define a solution concept (*guilt-aversion equilibrium*) that illustrates how communication may help select a 'cooperative' equilibrium when the agent is guilt averse. Empirically, it turns out that all forms of communication are not equally efficient in moving beliefs and motivation, however. Beliefs are higher when promises are made. It seems that statements of intent are instrumental in changing perceptions and facts about what people do.

We suggest that there are a variety of partnerships where guilt-aversion may be relevant. Examples include husband & wife, lawyer & client, procurement agency & contracted firm, inventor & producer, talented young golfer & rich sponsor, co-owners of firms, employer & employee, cartelists, etc. If such relationships are modeled as games, guilt-aversion equilibrium may be the right solution concept, and communication (and in particular promises) may help the partners coordinate on a 'good' equilibrium.

We believe there is ample scope for theoretical work based on guilt aversion. Contract theory has a history of basking in the light of great intellectual achievement, and incorporating communication and guilt aversion into the analysis would extend this tradition. One might, for example, attempt to characterize optimal contractual arrangements when agents are affected by guilt aversion. This could be done for contexts with hidden information (adverse selection) as well as contexts with hidden action (moral hazard). To answer such questions seems to us an exciting challenge in behavioral contract theory.

We close this paper with some remarks concerning how guilt aversion, in connection with communication, may matter in ways the importance of which is not limited to the more traditional topics of contract theory. Our list may be viewed as an attempt to inspire future research. First, we propose that the ideas that go into the notion of guilt-aversion equilibrium may help explain subtle aspects regarding how people use language. Why do people discuss, argue, and debate so much? Perhaps they are bargaining on what they should all agree is the right thing to do. Perhaps guilt aversion makes people adhere to agreements, once they are made. Perhaps guilt aversion can explain respect for democratic decision making, from voters who have accepted the legitimacy of the rules of some political process.

Second, an interesting issue for future research is whether people manipulate the guilt aversion of others in self-serving ways. For example, do authors of research papers attempt to convey, between the lines, the impression that they expect their paper to be accepted in a good journal? That would make sense if their referees were guilt averse; facing a marginal decision a referee may be swayed toward acceptance in order to avoid the guilt he would experience if he rejected the paper and let the authors down.

Finally, a further issue for future research concerns the relationship between guilt-aversion equilibrium and social norms. The literature on social norms is vast (see Elster, 1989, for a discussion and a classification). One central idea is to view a social norm as a social moral expectation, which people are inclined to live up to.[34] We suggest that in many situations living up to someone else's expectation may be just what people feel is the relevant social moral expectation, and that the cost of breaking that norm is related to the cost induced on others. Guilt-aversion equilibria may then capture a desire to obey social norms.

REFERENCES

Bacharach, Michael, Gerardo Guerra & Daniel Zizzo (2001), "Is Trust Self-Fulfilling? An Experimental Study", mimeo.

Battigalli, Pierpaolo & Martin Dufwenberg (2004), "Dynamic Psychological Games", mimeo.

Battigalli, Pierpaolo & Marciano Siniscalchi (1999), "Hierarchies of Beliefs and Interactive Epistemology in Dynamic Games", *Journal of Economic Theory*, **88**, 188-230.

Battigalli, Pierpaolo & Marciano Siniscalchi (2002), "Strong Belief and Forward Inductive Reasoning", *Journal of Economic Theory*, **106**, 356-391.

Baumeister, Roy, Arlene Stillwell & Todd Heatherton (1994), "Guilt: An Interpersonal Approach", *Psychological Bulletin*, **115**, 243-67.

[34] See Bernheim (1994) and Dufwenberg & Lundholm (2000; especially footnote 5) for models in this spirit.

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.

Bernheim, Douglas (1984), "Rationalizable Strategic Behavior", *Econometrica*, **52**, 1007-28.

Bernheim, Douglas (1994), "A Theory of Conformity", *J of Political Economy* **102**, 841-77.

Bicchieri, Cristina (2002), "Covenants without Swords", *Rationality & Society*, **14**,187-222.

Blau, Peter (1964), *Exchange and Power in Social Life*, New York: John Wiley.

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.

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.

Brandts, Jordi & Gary Charness (2004), "Do Labour Market Conditions Affect Gift Exchange? Some Experimental Evidence," *Economic Journal*, **114**, 684-708.

Brosig, Jeannette, Axel Ockenfels & Joachim Weimann (2003), "The Effect of Communication Media on Cooperation", *German Economic Review*, **4**, 217-41.

Brunson, Doyle (2002), *Super System: A Course in Power Poker*, 3rd edition [1st edition, 1978], Cardoza Publishing, New York.

Charness, Gary (2000), "Self-serving Cheap Talk and Credibility: A Test of Aumann's Conjecture", *Games & Economic Behavior*, **33**, 177-94.

Charness, Gary & Matthew Rabin (2002), "Understanding Social Preferences with Simple Tests", *Quarterly Journal of Economics*, **117**, 817-69.

Charness, Gary & Matthew Rabin (forthcoming), "Expressed Preferences and Behavior in Experimental Games", *Games and Economic Behavior*.

Cooper, Russell, Douglas DeJong, Robert Forsythe & Thomas Ross (1990), "Selection Criteria in Coordination Games: 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, Daniel Friedman & Steven Gjerstad (2004), "A Tractable Model of Reciprocity & Fairness", Working Paper 04-15, Department of Economics, University of Arizona.

Croson, Rachel & Mark Miller (2004), "Explaining the Relationship Between Action and Beliefs: Projection vs. Reaction", mimeo.

Crawford, Vincent (1998), "A Survey of Experiments on Communication via Cheap Talk", *Journal of Economic Theory*, **78**, 286-98.

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 (2004), "A Theory of Sequential Reciprocity", *Games & Economic Behavior*, **47**, 268-98.

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, Amsterdam, North-Holland.

Ellingsen, Tore & Magnus Johannesson (2002), "Promises, Threats, and Fairness", *Economic Journal*, **114**, 397-420, 2004.

Elster, Jon (1989) Elster, Jon (1989), "Social Norms and Economic Theory", *Journal of Economic Perspectives*, **3**, 99-117.

Falk, Armin & Urs Fischbacher (1998) "A Theory of Reciprocity," mimeo.

Farrell, Joseph & Matthew Rabin (1996), "Cheap Talk", *J 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.

Fudenberg, Drew & Jean Tirole (1991), *Game Theory*, MIT Press.

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 (forthc.), "Deception: The Role of Consequences", *American Economic Review*.

Güth, Werner, Peter Ockenfels & Marcus Wendel (1994), Efficiency by Trust in Fairness? Multiperiod Ultimatum Bargaining Experiments with an Increasing Cake", *International Journal of Game Theory* **22**, 51-73.

Hannan, Lynn, John Kagel & Donald Moser (2002), "Partial Gift Exchange in an Experimental Labor Market: Impact of Subject Population Differences, Productivity Differences, and Effort Requests on Behavior", *Journal of Labor Economics*, **20**, 923-951.

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.

Hovenkamp, Herbert (2000), *Federal Antitrust Policy: The Law of Competition and Its Practice*, 2nd ed., West Wadsworth.

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.

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.

Mertens, Jean-François & Shmuel Zamir (1985), "Formulation of Bayesian Analysis for Games of Incomplete information", *International Journal of Game Theory* 14, 1-29.

Moreno, Diego & John Wooders (1998), "An Experimental Study of Communication and Coordination in Noncooperative Games", *Games & Economic Behavior* 24, 47-76.

Nash, John (1950), "The Bargaining Problem", *Econometrica*, **18**, 155-62.

Orbell, John, Robyn Dawes & Alphons van de Kragt (1990), "The Limits of Multilateral Promising", *Ethics*, **100**, 616-27.

Pearce, David (1984), "Rationalizable Strategic Behavior and the problem of Perfection", *Econometrica*, **52**, 1029-50.

Rabin, Matthew (1993), "Incorporating Fairness into Game Theory and Economics", *American Economic Review*, **83**, 1281-1302.

Ross, Lee, David Greene & Pamela House (1977), "The False Consensus Effect: An Egocentric Bias in Social Perception and Attribution Processes", *Journal of Experimental Social Psychology*, **13**, 279-301.

Roth, Alvin (1995) "Bargaining Experiments," In the *Handbook of Experimental Economics*, John Kagel & Alvin Roth eds., 253-348, Princeton University Press.

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. Sauermann, ed., 136-68.

Siegel, Sidney & N. John Castellan (1988), *Nonparametric Statistics for the Behavioral Sciences*, Boston, McGraw-Hill.

Sobel, Joel (2001), "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 receive $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 **<u>B’s who chose ROLL.</u>**

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

(5,5) Messages from B

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

Sess.	ID	Message	Class	A	B
1	1	Please choose In so we can get paid more.	E	OUT	DR
1	2	Choose in, 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	IN	DR
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	IN	DR
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 tommorrow. 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	OUT	R
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	IN	R
1	6		N	OUT	R
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	IN	R
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	IN	R
1	9	Choose In and I will Roll You have my word	P	IN	DR
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. Who knows?	E	IN	R
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	IN	DR
1	12	Hey. OK I think that the best way for both of us to make a profit is for you to choose IN 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 and me less profit. So I'm a roller if you're in ☺.	P	IN	DR
1	13	take a risk	E	IN	R

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 1st 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	OUT	R
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	IN	R
1	16		N	OUT	DR
1	17		N	OUT	R
1	18	Choose "In" so we can both make some $$ What are the chances me rolling a 1? I'll try my best.	P	IN	R
2	1	I'm going to roll.	P	OUT	R
2	2	I'll choose roll.	P	IN	R
2	3	I will choose roll.	P	IN	DR
2	4	I'm going to choose roll	P	IN	R
2	5	choose in, & I'll roll.	P	IN	R
2	6	You can have the 2 extra dollars. I'll be nice and choose to roll. ☺	P	IN	R
2	7		N	IN	R
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	IN	DR
2	9	If you choose "In", I'll choose Roll and you've got a 5/6 chance of getting $12.	P	IN	R
2	10	Stay IN, I really need the money.	E	IN	R
2	11	If you choose IN, and I roll, the chances of our getting the most $ are very high. The likelyhood of my rolling a 1 is small compared to the chances of rolling a 2-6. So we both get cash.	E	OUT	DR
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	IN	R
3	1	Hopefully I'll make a lucky role.	E	OUT	DR
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	IN	R
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	OUT	R
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	IN	R
3	5	Both of 'us' can earn.	E	IN	DR

3	6	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. The probability that I will roll a 2,3,4,5, or 6 is pretty high, and I think worthy of trying for. (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.) x. I WILL ROLL	P	IN	R
3	7	I will roll, so if you stay in, you've got a 5/6 chance of getting $12. 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! 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. George W. Bush wants you to go in! Bin Laden says "out"! ☺	P	IN	R
3	8	Lets together get the most $ out of this that we can. ⇒ you 12 0 0 5 me 10 10 14 5 I promise not to do this one. ⇑ I promise I will choose to roll. You can have the extra $2 bucks. It's good karma. Thanks. 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. please excuse the awful handwriting. I'm trying	P	IN	DR
3	9	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	IN	R
3	10	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	IN	R
3	11	If you choose IN you have the best opportunity to make the most money. You have a 5/7 chance of making more money! So IN would be your best bet. Cheers. ☺	E	IN	DR
3	12	Choose IN. I promise I'll ROLL.	P	OUT	R

(7,7) Messages from B

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

Sess.	ID	Message	Class	A	B
1	1	If you choose to be in this, then I'll choose to roll. That way we both receive more money, because it's likely that he will roll 2,3,4,5 or 6 rather than 1. I'm willing to bet this if you will. ☺ In return you will probably receive $12!	P	IN	R
1	2	Choose IN and you get 12–7=$5 more (most probably) and I can get 10–7=$3 more. Please think about it!	E	IN	DR
1	3	Hi. I hope you're having a great day! And I hope that we will both be able to make some money!	E	OUT	DR
1	4	I'm going to choose "roll." I'd rather have $10 than $7. I promise I won't screw you over and pick "don't roll!"	P	OUT	R
1	5	I solemnly promise that I WILL choose Roll. (If I don't —you can hit me.) (That means you might want to choose in.) (This is Pareto optimal.)	P	IN	R
1	6	CHOOSE IN, SO WE CAN ROLL AND GET $12 AND $10.	E	OUT	DR
1	7	Hello Buddy, I would rather have $10 than $7…there is such a small chance he'll roll a 1…it's totally worth the risk. Do you like Pina Coladas? And getting caught in the rain? Me too.	E	OUT	DR
1	8		N	OUT	R
1	9		N	OUT	DR
1	10	Hi A. Hope you are having a nice day! I think I shall choose roll so you can hope for a number 2,3,4,5 or 6 if you choose IN. ……6x.	E	OUT	R
1	11	If you choose IN, I'll roll the die (it maximizes profit. 12+10=22 vs. any other option)	P	IN	DR
1	12		N	OUT	DR
1	13		N	OUT	DR
1	14	Hi. I'm choosing Roll. If we were doing this many times, on average you would get (5/6)*$12=$10>$7, but it's riskier. Anyways, it's your choice, but I don't like to screw people, and I believe in the Golden Rule.	P	IN	R
1	15	I'm still trying to decide what I want to do…Not sure.	E	OUT	DR
1	16	Hello!	E	OUT	DR
1	17	Have a happy day ☺	E	OUT	DR
2	1	You can trust me…I will Roll. I figure you have a 5/6 chance to get $12—that's a high probability. Let's try to get the 12 and 10 dollars.	P	IN	R
2	2	Don't choose "out" & we'll make more $	E	OUT	DR
2	3	I am going to roll. I promise. ☺	P	OUT	R
2	4	Choose "In" & I'll choose "Roll"; we'll have a 5 out of 6 chance of both getting big payouts.	P	OUT	DR
2	5	Hope you have a nice weekend! ☺	E	IN	DR

2	6	I will definitely choose ROLL.	P	IN	R
2	7	Hi! ☺ If you choose "IN," I will choose roll (I'll choose "roll" no matter what assuming you chose "roll") and then you'll have a 5/6 chance of getting $12 (so we'll each get about the same).	P	OUT	R
2	8	Choose IN because then I will choose Roll and there is a 5/6 chance that the die will roll a 2,3,4,5 or 6. That way you will receive $5 more than if you were to choose out.	P	OUT	R
2	9		N	IN	DR
2	10		N	OUT	DR
2	11	I'm going to choose "roll." Choose in.	P	IN	R
2	12	I'm choosing Roll and hoping that you choose in. I doubt it will be a 1 and if you choose in we will both get more money. I'm not concerned about you getting $2 more than me, just that we both can maximize our earnings in the fairest way. Trust me I won't screw you. I have morals and I'd feel bad.	P	IN	R
2	13	Take a chance. Good Luck. ☺	E	OUT	DR
2	14	I'll be picking "Roll," so if you want you should pick "In."	P	OUT	R
2	15	If nothing is better than life And cheese is better than nothing Is cheese better than life? Interesting isn't it…OK I will win money no matter what, so I am not about to screw you over. On my word, I guarantee I will roll. That's it, I am rolling. MY WORD is said!	P	IN	R
2	16	Hey, this is your B pair. I am going to roll (though I cannot promise that I am not going to roll 1☺). So, just ROLL…	P	IN	R
2	17	I don't have anything to say, except that I will roll the die. Good Luck!	P	IN	DR
2	18	Gonna <u>roll</u>, high odds of both getting good payoff, I'm all for trying to get the best outcome for both of us.	P	IN	R
3	1	Writing messages is the best. Still, I'm confused. Can't wait till Summer. I'm bored. To the center this goes.	E	IN	DR
3	2	I will make a decision where we both get money.	E	IN	DR
3	3	I will choose roll for two reasons. 1) you need an incentive to not pick Out, I would much rather have $10 than $7. 2) I assume you are here because you are poor, like me, and I derive no pleasure from getting 4 more dollars at the expense of you getting nothing. I swear on my mother's grave I am choosing ROLL.	P	IN	R
3	4	Good afternoon. I will choose ROLL.	P	OUT	DR
3	5	Quote the wisdom of Fred Durst: Keep on <u>rollin</u>' baby, You know what time it is! Do it, you, and buy yourself a Jamba Juice later. ☺	E	OUT	R
3	6	If you choose IN, I'll choose roll. Let's take this guy for as much as possible. ☺	P	IN	R
3	7	Hopefully we both walk out of here with more than $5.	E	IN	R
3	8	Hello! It's such a beautiful day, and I hope you are having a wonderful day. ☺	E	OUT	DR

3	9	I will honestly choose "Roll" if you choose "IN." We both have the best chances at making money since the odds of rolling a one are only 1 in 6 and it doesn't matter if you make 2 more dollars than me as long as I get $10 for sure.	P	IN	R
3	10	I will choose to ROLL. If you choose IN, and rely that it won't be a 1, you will receive 12 dollars. I swear that I chose to ROLL. Please choose IN. I am willing to only get 10 dollars so you can get 12. I promise that I have chosen to ROLL. 83% chance you will get 12 dollars.	P	IN	R
3	11	Choose in and I will roll. That way, we'll both get extra money.	P	OUT	R
3	12	☺ Hi	E	OUT	DR
3	13	This is free money! Let's go for it!	E	IN	DR
3	14	CHOOSE "IN."	E	OUT	DR

(5,5) Messages from A

In this table: R = ROLL, DR = DON'T ROLL

Sess.	ID	Message	A's choice	B's choice
1	1		OUT	R
1	2	Let's share & Roll. Yey money wahoo cool ☺	IN	R
1	3	Good chance I'll be getting screwed over here cuz I'll be going IN … But taking the chance, and interested in whether I'll lose out big here or not … Care to help me out? If not, no bigs. I'd probably go for $14 if I was B too… So either way, I understand … go for 14 if you'd like … I'm for sure going IN.	IN	DR
1	4	Mathematically, you're ahead no matter what I do, so let's gamble!	OUT	R
1	5	Choose roll!	IN	DR
1	6		OUT	DR
1	7	You can't lose! ☺ but you can take a chance and sacrifice four dollars for both of us to go home happy. The decision is yours. ☺ Best of luck.	IN	DR
1	8	Hello B! I'm going to choose IN, hoping that you'd choose ROLL. I'm assuming that you're not a selfish type to choose don't roll to only benefit yourself. So … hopefully, by me choosing IN and you choosing ROLL, we'll both get at least $10. (I'm willing to take the risk of getting $0) Bye! ☺	IN	R
1	9	I like the computer-based experiments better – this game has a low pay-off and takes TOO long!	OUT	R
1	10	Let's roll! Chances will be the results will be good for at least one of us. Right?	IN	DR
1	11	Your most rational choice is to choose don't roll. My most rational decision is out. I've given up gambling so I'm going with out. Sorry. Enjoy the $10.	OUT	DR

1	12	Please choose to roll! That way we'll both have a chance to get some extra money instead of just $5. ☺	IN	R
1	13	Hey, this is beer money. Be cool and choose to roll. You're the difference between Natty Ice and something better.	IN	R
1	14	Hi. I am choosing "IN", I would like you to roll. If you roll I have a chance of making more than $5. You will make more than $5 in any case, as I will select roll. Thanks!	IN	DR
1	15	Dear B, I came to participate in this exp because I am low on cash, really low. Have you ever been in that position? Please be nice and we'll both make $. ☺	IN	DR
1	16		OUT	DR
1	17	PLEASE! Choose to Roll, that way we can both have a chance of winning any money. No matter what happens if you Roll you will get $10. Please don't be selfish and choose not to Roll, Feel sorry for me!	IN	R
1	18		OUT	DR
2	1	Let's make this beneficial for both of us … (a condition where we would both win something). ☺	IN	DR
2	2	I will choose to opt in, please choose to roll, it will be good for both of us. I'm trusting that you will do the right thing and not leave me with anything, but opting in I risk getting nothing if you roll a 1, please trust me, I am trusting you!!	IN	DR
2	3	Please roll ☺	IN	R
2	4	Let's do it! Let's roll!!! I am in!	IN	R
2	5	Cut me some slack. I'm really broke. Please roll the dice.	IN	R
2	6		IN	DR
2	7	Let's make this fun, please pick roll.	IN	R
2	8	Better be Safe than Sorry.	OUT	R
2	9	Please roll, & then I will choose in & we can both have a greater chance of winning 7 if it happens that you roll 2,3,4,5 or 6 I can give you $4 so that you won't lose out @ all. Thanks	IN	DR
2	10	I need to go to Albertsons so lets aim for money.	IN	R
2	11	Roll … you are guaranteed at least a ten.	IN	DR
2	12	If I were to choose out we would both only get $5 but if I choose in you are guaranteed to receive more than 5 dollars. I am going to choose in so that you have a higher chance at getting money. Please choose roll so that we will both get more money hopefully. I think we could split the money so we both have a really good chance of getting $11 or a guarantee of $5 each.	IN	DR
2	13	I'm deciding if you're greedy or not. If you are, you're gonna choose don't roll and screw me over where I get nothing & you get 14. If you're not greedy, you're not greedy & want to help me, I'll help you too. I want us both to win. Be kind.	IN	DR
2	14	If you choose Roll and it is not a 1 result, I will give you $4.	IN	R
2	15	Choose roll so we can both get paid.	IN	DR
3	1		IN	DR

3	2	It would be nice of you to choose ROLL so we both get money & can have a happy Memorial day weekend yey! ☺ Go 2,3,4,5,6 you can do it	IN	R
3	3	An elephant never forgets	OUT	DR
3	4	Do you want to go in or out? It'd be nice if we both got $.	OUT	R
3	5	Probably going to choose "out" just to make sure I get something substantial out of this session. Hope ya don't mind too much!	OUT	DR
3	6	Please choose roll	IN	DR
3	7		OUT	DR
3	8	I am going to play conservatively and choose out.	OUT	DR
3	9	I promise to choose in, This way we both get more money You're guaranteed $10 as opposed to $5 & I have a 5/6 chance of getting $12 & a 1/6 chance of getting $0. I'm ok with these odds though! PLEASE choose roll! I'm putting myself on the line here, don't let me leave here with $0 earnings. Thanks a lot	IN	DR
3	10	GOOD LUCK! ☺	IN	DR
3	11		OUT	R
3	12		OUT	DR
3	13	I am choosing "IN" in order to ensure you will make $15. I would ask that you show the same consideration and choose to "ROLL" so I am not left with only $5. Together we can both walk away with a fair amount of money. Thanks in advance.	IN	DR

APPENDIX D - REGRESSIONS

Probit Regressions for Determinants of Behavior

Independent variables	Dependent variable (1) ROLL	(2) ROLL	(3) IN	(4) IN
Constant	-1.520*** (.235)	-1.743*** (.264)	-0.884*** (.187)	-1.270*** (.220)
Guess	0.023*** (.004)	0.023*** (.004)	0.016*** (.003)	0.015*** (.003)
B Message	-0.103 (.262)	-0.054 (.267)	-0.237 (.249)	-0.176 (.257)
A Message	-0.140 (.247)	-0.347 (.267)	0.599** (.243)	0.216 (.265)
Promise sent	0.831*** (.300)	0.797*** (.300)	-	-
Promise received	-	-	0.944*** (.292)	0.944*** (.300)
(5,5) treatment	-	0.439** (.209)	-	0.303** (.077)
N	230	230	230	230
Pseudo R^2	.204	.218	.163	.209

Standard errors are in parentheses; *** indicates significance at p = 0.01, ****** indicates significance at p = 0.05, and ***** indicates significance at p = 0.10 (all two-tailed tests). 100 values = 1 if the cards had 100 values and is 0 otherwise; 100 values first = 1 if the cards had 100 values initially in the session and is 0 otherwise.

One's guess is highly significant for both A and B behavior, as is whether a promise is sent or received. B's message *per se* has no significant influence on the *Roll* or *In* rate, while a promise sent or received makes a big difference in behavior. We see that there is significantly more cooperative behavior in the (5,5) treatment, where there is more to gain from cooperation. Finally, A's are more likely to choose *In* when they can send a message, although this is not significant in specification (4).

Figure 1

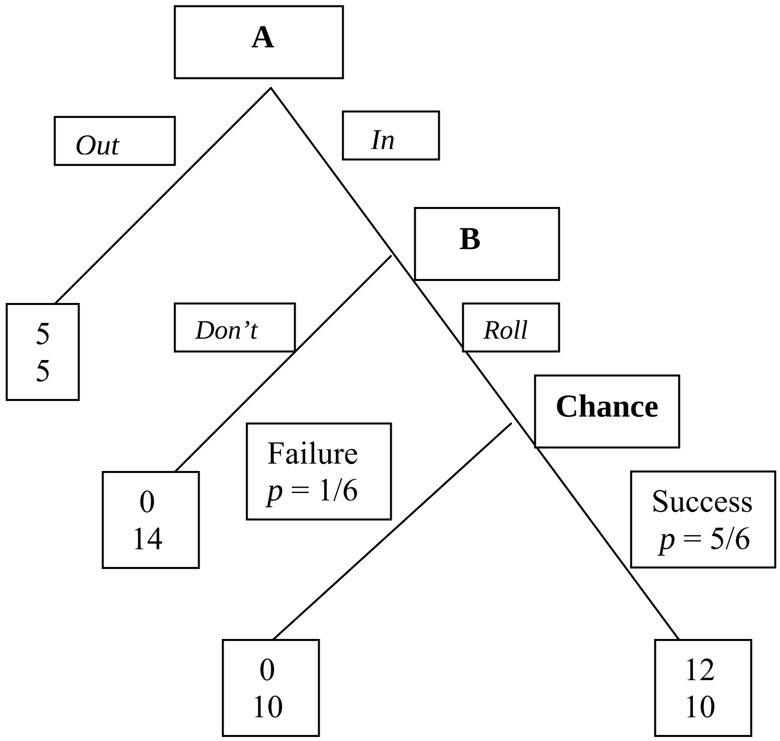

A
Out
In
B
5
5
Don't
Roll
Chance
Failure
p = 1/6
0
14
Success
p = 5/6
0
10
12
10

Figure 2

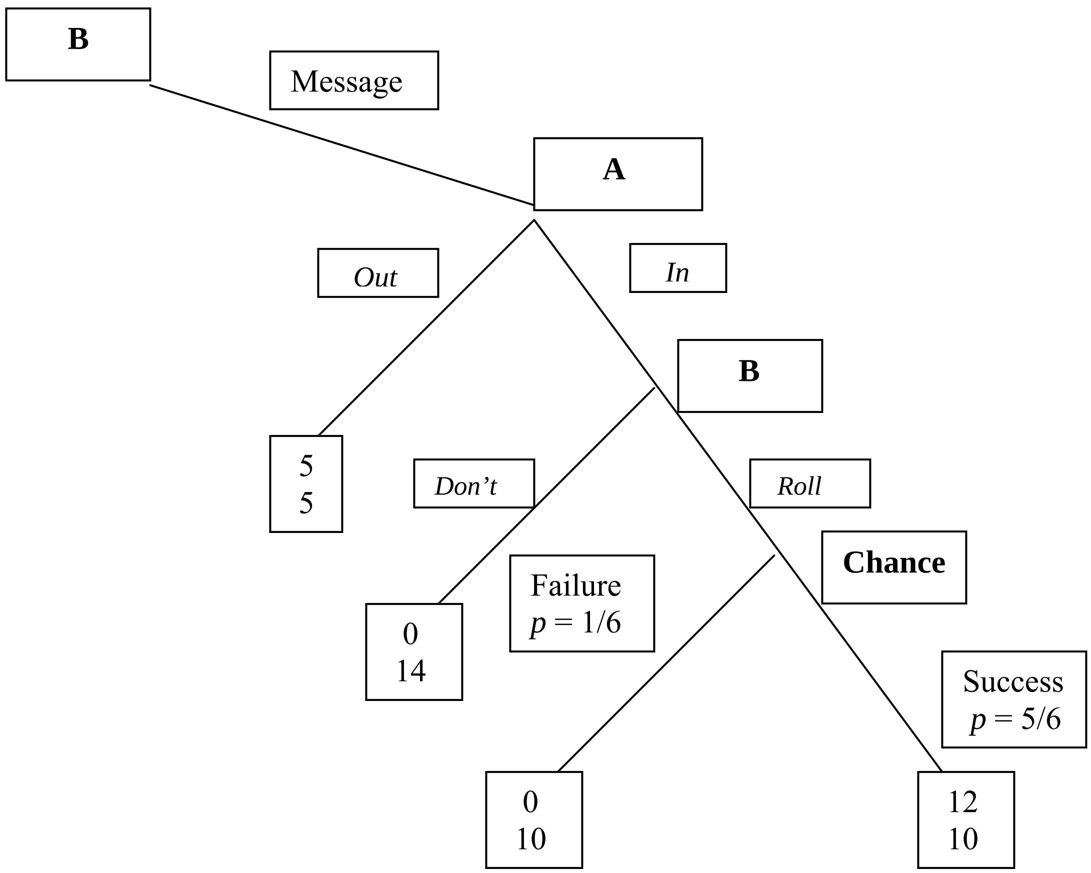
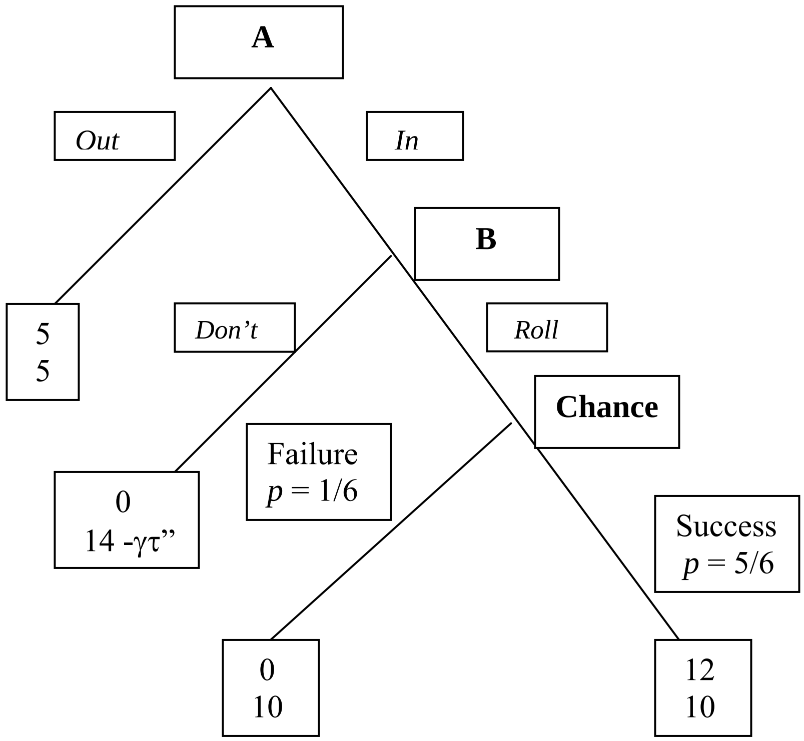

B
Message
A
Out
In
B
5
5
Don't
Roll
Chance
Failure
p = 1/6
0
14
Success
p = 5/6
0
10
12
10

Figure 3


A
Out
In
B
5
5
Don't
Roll
Chance
Failure
p = 1/6
0
14 -γτ"
Success
p = 5/6
0
10
12
10