Promises & Partnership

Gary Charness & Martin Dufwenberg*

February 13, 2003

Abstract: We examine, experimentally and theoretically, how communication within a partnership may mitigate the problem (highlighted in contract theory) of hidden action. What is the form and content of the communication? Which model of decision-making can capture the impact of communication? We consider free-form communication, measure beliefs (about actions and beliefs), and examine which motivational forces influence subjects. We find they harbor belief-dependent preferences that can be captured using psychological game theory. In particular, agents are influenced by guilt aversion, which suggests a theory of why and how communication influences behavior in which statements of intent and resulting expectations play a special role. This has bearing on how to understand partnerships and contracts.

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

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

Acknowledgments: We thank Jon Baron, Jeanette Brosig, Steve Burks, Ernst Fehr, Ayelet Fishbach, Guillaume Frechette, Dan Friedman, Drew Fudenberg, Simon Gächter, Uri Gneezy, Brit Grosskopf, David Laibson, Dan Levin, David I. Levine, Tanya Menon, Matt Parrett, Torsten Persson, Yuval Rottenstreich, David Strömberg, Richard Thaler, Bernd Wittenbrink, and participants at seminars in Antwerp, Berlin (Humboldt), Bonn, Chicago (GSB), Cologne, Dortmund, Gothenburg, Great Barrington (Behavioral Research Council), Harvard, Munich, Örebro, Oslo, Purdue, Santa Barbara, Stockholm (Econ. Dept. and IIES), Wittenberg (GEW-Tagung), and Tampa (Public Choice) for stimulating discussions and helpful suggestions. We are grateful to the Swedish Competition Authority for financial support.

* Contact: Gary Charness, University of California at Santa Barbara, charness@econ.ucsb.edu; Martin Dufwenberg, Stockholm University, md@ne.su.se.

1. INTRODUCTION

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

Most of contract theory makes scant reference to empirical observation. Perhaps one should not simply accept its logic without questioning its empirical relevance. We focus on one particular aspect where the traditional contract theoretic perspective may be too barren: *communication.* The conventional approach implicitly assumes that non-binding communication is ineffective in promoting partnership formation and cooperation; written contracts bind if supported by the law, but oral agreements (to quote Samuel Goldwyn) ”aren't worth the paper they're written on”. We feel that this view is at odds with reality, where promises, discussions, handshakes, threats, and other forms of communication are often used when agreements are made. These casual observations justify a suspicion that communication can foster trust and cooperation in settings with hidden action, in contrast to the prediction of conventional contract theory.

This paper contains an experimental and theoretical examination of this idea. We are interested in four basic questions:

---

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

Q1. *Are the predictions of classical contract theory borne out in the lab?*

Q2. *Does communication foster cooperation?*

Q3. *What is the form and content of the communication?*

Q4. *Which theories of decision-making and communication can explain the data?*

We find that the answer to Q1 is (a qualified) *No*, which justifies our interest in the other questions. Q2 concerns whether or not communication influences behavior, regardless of *why* this might be the case. Q3 sheds light on how people use communication strategically, to convey information that might foster trust and cooperation. Q4 is essential for understanding the motivational forces that drive behavior, and why communication may matter in this connection. If contract theory turns out to need revision, such insight is crucial for understanding what contracts are 'behaviorally optimal', and for deriving testable predictions that allow exploration of how far the results extend to other economic settings.

We approach Q4 by testing how well some recently proposed models of social preferences explain our data. Experimentalists have accumulated a wealth of evidence that decision-makers are not completely selfish. In response researchers have recently proposed several explanatory models, which make reference to various notions of fairness, reciprocity, etc.[2] To a person familiar with this literature, it will come as no surprise that these models can explain why problems of hidden action may be overcome. However, previous work has not focused on what these models imply regarding the impact of communication. This is a key feature of our analysis.

The mathematical structure of the social-preference models has important bearing both on how to set up experimental tests and on how to understand the role of communication. In some models decision-makers' preferences depend only the overall distribution of monetary earnings among the reference group. These *distributional models* do not represent a radical breach with conventional economic theory; they can be represented as conventional games, in which preferences are defined on strategy profiles only (because

[2] For descriptions of the experimental evidence and the social-preferences literature, see Fehr & Gächter (2000), Fehr & Schmidt (2001), and Sobel (2001).

strategic choices imply final distributions, which is a decision-maker's only concern). Another class of models, which may be labeled *belief-dependent*, is far more complicated as decision-makers' motivation depends directly on their perceptions of intentions or beliefs. Preferences must be defined on non-standard domains, as decision-makers care for more than strategy profiles. One is led outside the realm of standard game theory and into the framework, proposed by Geanakoplos, Pearce & Stacchetti (1989), of *psychological game theory*. To experimentally test psychological games based models with belief-dependent utility it is not sufficient to merely record strategic choices; one must also elicit certain beliefs (about actions and about beliefs). We develop a design that allows us to do so.[3]

The scope for communication to matter in distributional models is rather limited, and essentially concerns only equilibrium selection in games with multiple equilibria. In the games we consider, distributional models yield unique predictions (even without assuming equilibrium play) and therefore imply that communication should not matter. By contrast, subtle links surface in models with belief-dependent utilities, leading to communication affecting behavior. Messages influence beliefs, beliefs influence motivation, and motivation affects behavior. Since our experimental design involves belief elicitation, we can examine if and how such links matter, and thereby learn about the impact of communication.

The idea that communication can affect strategic interaction, and experiments related to the topic, are of course not new.[4] However, we propose that our tack is novel in many ways. First, the strategic setting we consider is motivated from a contract-theoretic perspective. The game we derive has a 'trust' structure, and is different from the games considered in other studies (which typically concern prisoner's dilemmas, coordination

[3] Measuring beliefs is a recent innovation in experimental work; Dufwenberg & Gneezy (2000) and Nyarko & Schotter (2002) have developed relevant methods.

[4] Some work is found in psychology. Dawes, McTavish & Shaklee (1977) find that 'relevant' face-to-face communication (about the situation at hand) leads to a substantially higher rate of cooperation in a commons dilemma. Orbell, Dawes & van de Kragt (1990) study multilateral communication, and find that promises enhance cooperation if *everyone* in a game makes them. Communication may also be relevant to the extent that contracts are social contracts (Blau, 1964), psychological contracts (Rousseau, 1995), or reflect social norms (Bicchieri, 2002). In experimental economics, communication has been found to improve outcomes in coordination games (e.g., Cooper, DeJong, Forsythe & Ross, 1990, 1992; Charness, 2000) and has been studied in bargaining environments (Valley, Moag & Bazerman, 1998; Bohnet & Frey, 1999; Ellingsen & Johannesson, 2002; Brandts & Charness, forthcoming; Brosig, Ockenfels & Weimann, forthcoming. A related strand of literature focuses on *deception* in games; we connect to this work in section 3.3 below.

games, or bargaining games). Second, and perhaps most importantly, by linking our work to psychological games we incorporate a new theoretical perspective. We measure beliefs and test models with belief-dependent utility; and eventually provide the rudiments of a new theory of why communication matters.

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, and would thus be sister to fields like behavioral finance and behavioral game theory. Behavioral contract theory takes into account social and psychological considerations in an attempt to understand partnerships and contracts.

Two approaches to behavioral contract theory may be distinguished. First, one may explore (experimentally or theoretically) which contract people choose when there are *many feasible ones*.[5] Second, one may consider *one specific contractual arrangement*, and attempt to understand that environment in some depth. The current paper belongs here, and the in-depth issues we explore are Q1-Q4. We explore whether and why a given contract is acceptable to two parties, with and without communication. We make no presumption that the contract we look at is 'optimal' - that would be impossible since we enter the analysis being open-minded with respect to what motivational forces are at work. The objective is to reveal insights about decision-making and motivation that are useful for developing behavioral contract theory further.

Theory, experiment, and conclusions appear in Sections 2, 3, and 4, respectively.

2. THEORY

In this section we develop a simple model of a partnership (2.1), from which we derive our experimental games (2.2). This exercise highlights the contract-theoretic backdrop to our design, and we argue that it also has some independent value. The model may provide

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

a useful framework for other experimental studies, since it is simple and yet flexible enough to admit examination of a variety of contract-theoretic issues. In this paper we elicit a simple mono-contractual setting incorporating hidden action, but other versions could be used for considering hidden information (adverse selection), multiple contracts, or richer strategic settings.

We also introduce several theories of social preferences (2.3), and derive testable predictions for the experimental analysis.

2.1 A benchmark model

A principal and an agent consider forming a partnership in which a project is carried out. If no partnership is formed, then no contract is signed, no project is carried out, and the parties get their outside-option payoffs x for the principal and y for the agent (measured in dollars). If the project is carried out, then the contract specifies a 'wage' w that the principal must pay to the agent. The project generates revenue for the principal. There can be two outcomes: failure or success. A failure generates revenue $r > 0$, while a success involves an additional bonus of $b > 0$, so that total revenue is $r+b$. The probability of these outcomes depends on the choice and characteristic of the agent; we assume the agent chooses 'effort', $e \in [0,1]$, and has a given 'talent', $t \in [0,1]$, and that the probability of success is et.[6] The agent experiences increasing 'effort cost', measured in dollars and equal to ce, where $c > 0$.

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

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

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

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

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

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

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

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

2.2 The experimental games

The gist of modern contract theory derives from assumptions about asymmetric information between the principal and the agent. A major issue is the choice of contract when the agent's effort is not contractible (hidden action) or when the principal does not know the agent's talent (hidden information). The assumption is typically maintained that the principal and the agent are perfectly selfish.

We shall consider circumstances characterized by hidden action. However, we will *not* examine which contract out of many feasible ones will be agreed upon, given a particular motivation. Rather we stay open-minded with respect to the nature of the motivation and examine, for a given contract, how serious the problems caused by hidden action are in the first place. That contract corresponds to the benchmark solution of the previous section, for the following specific set of parameters:

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

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

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

FIGURE 1

First, the principal decides whether to say *Yes* or *No* to the contract according which he must pay $w^* = 14$ to the agent. If he says *Yes*, then a partnership is formed and the project is carried out, but it remains for the agent to choose his effort level $e = 0$ or $e = 1$. The payoffs are derived using (4) and the assumptions spelt out in section 2.1.[7]

If the players are selfish the game has an obvious (backward induction) solution. If called upon to play, the agent would exert low effort. The principal's best response is to choose *No*, not agreeing to form a partnership. This outcome is inefficient, since both parties receive a higher expected payoff with the (*Yes*, $e = 1$) outcome then when no partnership is formed; this is a simple illustration of how conditions of hidden action may undermine efficient contracting.

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

FIGURE 2

If the players are selfish and risk-neutral, the presence of the communication stage has no impact on the analysis. The above argument, leading to the (*Yes*, $e = 1$) outcome in the game of Figure 1, goes through unchanged for the post-communication subgame in Figure 2. Hence, with selfish players, again the theoretical prediction is that no partnerships are formed and that the outcomes will be inefficient.

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

2.3 Social preferences and the impact of communication

In order to apply the non-self-interested behavior observed in experiments to economic settings such as consumer response to price changes or employee response to changes in wages and employment practices, researchers have begun to develop formal models of social preferences that assume people are not solely selfish, but also care in some way about others. In this section, we examine how such ideas bear on our setting.

The development of the social-preference literature has to a large extent been inspired by experimental results. Fehr & Gächter (2002) summarize the development. They suggest that a diverse set of economic phenomena can be understood as involving some form of 'reciprocal behavior', but note that there are several (competing or complementary) models that can explain various facets of such behavior, and that more work is needed to understand the motivational forces at work. Here we consider how three approaches might explain cooperation and the impact of communication in our context.

First we have the purely distributional approach, in which only the final monetary distribution is seen as relevant to the decision. The process leading to this distribution is either seen as irrelevant or as captured implicitly by distributional proxies. We focus on *inequity aversion*, as presented in Fehr & Schmidt (1999) and Bolton & Ockenfels (2000).[8] Second, we consider *kindness-based reciprocity*, as described in Rabin (1993) and Dufwenberg & Kirchsteiger (1998). Under this motivation, people consider the degree of kindness embodied in choices made by other people. Third, we discuss a form of expectations-based motivation that we call *guilt aversion*. People prefer to avoid the guilt that results from disappointing the expectations of a person who has acted decently. These three approaches differ in terms of how communication may influence behavior, and we derive testable predictions that are subsequently addressed in the experiment.

We shall test these theories in a very basic form, referring only to properties of the utility functions of individual decision-makers. We shall not derive and test elaborate notions

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

of ‘equilibrium’. Most of economic theory *assumes* that subjects coordinate on some kind of equilibrium, and the entailed predictions may fail *either* because the underlying utilities are erroneously specified *or* because the equilibrium coordination is not present. We shall concentrate solely on the accuracy of the preference description, without worrying also about coordination. We shall furthermore focus mainly on the behavior of people in the role of agents.

Inequity aversion

In models of inequity aversion, people are presumed to like money but also (possibly with heterogeneity) dislike disparities in payoffs. Such models are quite successful in explaining a variety of experimental data; applied to our games, they allow the agent to prefer $e = 1$ to $e = 0$, if he is sufficiently averse to getting a higher payoff than the principal. Thus, inequity-aversion models are capable of explaining why problems of hidden action are overcome.

These models refer only to the monetary payoffs that result from the agent's choices. Such distributional models have the great virtue of being simple to apply and to test. It is therefore important to assess their range of applicability. We check one particular dimension of this issue: Can purely distributional models capture how communication influences strategic interaction in a setting with hidden action?

In games where distributional models admit multiple equilibria, communication may play a role by facilitating coordination on some particular equilibrium. However, in our game distributional models generate *unique* predictions concerning the behavior of agents (barring unlikely cases of indifference). Therefore, a stark prediction emerges concerning the impact of communication. Note that the $e = 1$ prediction for an inequity-averse agent was made without reference to whether or not the player's choices were preceded by a communication stage. The upshot is that the inequity-aversion models predict that *communication does not matter*.[9]

9 Two comments are in order: (i) Communication may matter if the *degree* of inequity aversion is made a function of whether there is communication. We do not think much of this idea; a key point of distributional

Kindness-based Reciprocity

The classic reference for a model of kindness-based reciprocity is Rabin (1993). In his model, decision-makers wish to be kind to those they believe to be kind, and to be unkind to those they believe to be unkind. The notion of kindness in his theory is rather complex, taking into account what a person *believes* he accomplishes with his behavior. Intentions matter. Similarly, the definition of believed kindness takes into account what a person believes about the kindness of others. Since kindness is belief-dependent, believed kindness depends on what the person *believes others believe* they accomplish with their behavior. To exemplify all of this, if someone flies into a building he is not unkind unless he did it on purpose, and if others believe this was non-purposeful behavior, the event would be no cause for engaging in war. Rabin's theory captures such ideas, and requires the toolbox of psychological game theory because of the belief-dependency of kindness, and believed kindness.

Rabin's model is developed for games represented in normal form, and is only meant to highlight the key qualitative features of reciprocity. As Rabin notes (p. 1296), it is not well-suited for applied work involving games that have (as here) a non-trivial dynamic structure. Dufwenberg & Kirchsteiger (1998) develop a model of reciprocity for extensive-form games, which is more amenable to applied work.[10] We shall rely on that model to derive a prediction concerning the agent in our game. We refer to the original paper for a full presentation of the theory, and here just introduce enough notation to handle our specific example: Let τ denote the probability with which the agent chooses $e = 1$; let τ' denote the principal's expectation of τ; let τ'' denote the agent's expectation of τ' (everything measured at the agent's decision node).

---

models is that one need *only* make reference to distributions. Assuming communication-sensitive inequity aversion destroys that virtue. (ii) If there is incomplete information about the players' degree of inequity aversion (which seems reasonable, and is in fact assumed in Fehr et al., 2002), then a principal's behavior could be affected by a message that may signal the agent's 'type'. However, this could not affect the agent's behavior, which is our main focus.

[10] Segal & Sobel (1999) and Cox & Friedman (2002) present other approaches to modeling reciprocity. There are also models that combine distributional preferences and reciprocity; see Falk & Fischbacher (1998) and Charness & Rabin (2002).

Consider first the kindness of the principal when he chooses *Yes*. The choice of *Yes* brings about a higher payoff to the agent than the choice of *No*, so *Yes* is kinder than *No*. However, *the exact degree* of kindness of *Yes* depends on the principal's beliefs. Namely, the principal's kindness when choosing *Yes* depends negatively on τ'. The logic is that the lower is τ', the more money the principal believes she believes she gives to the agent, so the kinder she is. Now focus on the agent. To determine his preferred choice after the principal chooses *Yes*, the agent weighs his monetary incentives against his desire to be kind in return (by choosing $e = 1$). Since the principal's kindness depends negatively on τ', the lower the agent's belief of τ', the kinder the agent believes the principal is. In other words, the lower τ'' is, the kinder the agent believes the principal is. Hence, the lower τ'' is, the more 'inclined' the agent will be to be kind in return and choose $e = 1$.

More precisely, the model includes a parameter capturing the decision-maker's sensitivity to reciprocity. The lower is τ'', the lower the reciprocity-sensitivity need the agent have to prefer $e = 1$. Dufwenberg & Kirchsteiger develop an equilibrium solution concept, which applied to our game here would pin down behavior and beliefs, dependent on the reciprocity-sensitivity parameter. In this paper, however, we do not wish to invoke any equilibrium assumption (cf. our remark above). Rather, we wish to derive a prediction concerning the agent's behavior, *referring only to his utility function.*

We proceed as follows: First, we assume that the reciprocity-sensitivity differs among people. This assumption seems reasonable, because a wealth of research in personality psychology suggests that individuals differ in their sensitivity to emotional concerns; see Krone (2002) for a general discussion. Moreover, much experimental evidence indicates that different people have different motivations in similar situations. Second, we assume that *the reciprocity-sensitivity is independent of* τ''. While this assumption is admittedly a bit stronger than one might like, we cannot observe the reciprocity-sensitivity parameter, and we invoke the assumption for lack of an obvious better alternative. It is important to note that what follows hinges on this assumption. Namely, *the lower is* τ'' *the more likely it is that a given person will prefer* $e = 1$. From the viewpoint of an outside observer (who does not know an individual's sensitivity to reciprocity), a *negative correlation* between the likelihood of the *e*

= 1 choice and τ'' is predicted. Our experimental design allows us to measure τ'' and to test this prediction.

Communication may play a role when decision-makers are motivated by kindness-based reciprocity. Since motivation depends on beliefs, if communication can affect beliefs there will be a link from words to actions. However, the link is subtle: Suppose the agent says "I promise to choose $e = 1$." If the agent believes that the principal believes he is honest (so that $\tau'' = 1$), the previous arguments suggest that the agent must view the principal as less kind than if the agent said "I promise to choose $e = 0$" (so that $\tau'' = 0$ if the agent believes the principal believes the agent is honest). Thus, if the agent believes the principal believes him, he will have an incentive to renege! Under kindness-based reciprocity, truth-telling is not self-enforcing.

This is intriguing and somewhat mind-boggling, and a detailed theoretical examination of communication among agents motivated by kindness-based reciprocity may prove to be rewarding research. Here we avoid getting more deeply involved in the issue, and simply record choices and beliefs in our experiment and perform statistical tests regarding the correlation described above.

Guilt Aversion

The third social preference we consider is that decision-makers suffer from guilt if they believe they let others down, acting at variance with the expectations of people who have acted with decency. Decision-makers therefore shy away from making such choices. We label such decision-makers ‘guilt averse’.

There is a literature in social psychology that discusses closely-related ideas, although no mathematical modeling is done. See, *e.g.*, Baumeister, Stillwell & Heatherton (1994, 1995) who (on the basis of autobiographical narratives) suggests that people suffer from 'guilt' if they inflict harm on others. One prominent way to inflict harm is to let others down. In the words of Baumeister *et al.* (1995, p. 173): “Feeling guilty [is] associated with...recognizing how a relationship partner's standards and expectations differ from one's own”. It is such a sentiment we now propose to model and test.

This motivation makes sense in many settings, but simple as it may seem it calls for non-standard modeling. Consider the following example, chosen to illustrate these points in the simplest possible way, rather than because of its immediate relevance to contract theory:

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

Perhaps surprisingly, conventional game theory cannot model Björn's motivation and behavior. Consider a standard game where Björn (player 1) chooses a tip, and the waitress (player 2) has no choice (her strategy set is modeled as a singleton). Björn's choice of tip determines a full strategy profile. In game theory, payoffs are defined only on strategy profiles, so Björn's set of best choices must be independent of his belief of the waitress's belief. This contradicts the example. Hence, Björn's preferences cannot be described using conventional game theory; as in the case of kindness-based reciprocity, psychological game theory must be used.

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

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

where $\gamma > 0$ is a constant measuring Björn's sensitivity to guilt. Björn's utility thus depends negatively on how much money he gives away, and on the extent to which he believes he does not live up to the waitress' expectations. We make several observations: (i) This leads to a psychological game rather than a standard game, because although t determines a strategy profile, the payoff $-t-\gamma \cdot \max\{t''-t, 0\}$ is not a number but rather a function of t''; (ii) If $\gamma > 1$,

Björn's optimal choice would be $t = t''$; (iii) Incorporating guilt this way implies a *positive correlation* between t and t''.

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

FIGURE 3

In Figure 3, the parameter $\gamma > 0$ again scales the agent's guilt sensitivity. If γ is large enough, and if the agent believes sufficiently strongly that the principal believes he will choose $e = 1$ (i.e., if τ'' is large enough), then the inequality $14-\gamma\tau'' < (5/6)\cdot 12+(1/6)\cdot 0 = 10.$ holds, and the agent will indeed choose $e = 1$. Alternatively put, the lower is τ'', the higher γ must be in order for the agent to prefer $e = 1$. If γ differs among individuals and is independent of τ'', there will be positive correlation between the likelihood of an $e = 1$ choice and τ''.[11] Note that the sign of the predicted correlation is the *opposite* of that we derived for kindness-based reciprocity.

What is the role of communication in this connection? As it turns out, it is quite remarkable. Again, communication may influence beliefs, beliefs influence motivation, and motivation influences behavior. With the assumed guilt effect, this link is more straightforward than in the case of reciprocity, however. Suppose the agent says "I promise to choose $e = 1$." If the agent believes that the principal believes him, this will make the agent *more* inclined to choose $e = 1$. This in turn gives the principal a reason to believe the agent's statement. For an agent capable of feeling guilty, truth-telling is therefore self-enforcing. By issuing a promise the agent can gain commitment power regarding the exercise of his choice $e = 1$.[12]

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

[12] We focus on communication by the agent, but note that with an agent prone to feeling guilty the principal also has an incentive to indicate that he has high expectations. Similarly, in the tipping example, the waitress or

Both the principal and the agent gain relative to the (*No*, $e = 1$) outcome. A happy ending is brought about which reflects on Leith & Baumeister's (1998, p. 1) assertion that "guilt serves many adaptive, beneficial, and prosocial functions", and that "guilt helps strengthen and maintain close relationships" (p. 2).

Guilt aversion has not received much attention by scholars working on social preferences, but related ideas appear in some applied theoretical work by Huang & Wu (1994) (on remorse in corruption) and by Dufwenberg (2002) (on guilt in marriage). Experiments by Dufwenberg & Gneezy (2000) and Bacharach, Guerra & Zizzo (2002) involve second-order belief-elicitation and findings in line with guilt aversion. None of these studies deal with communication.[13]

3. THE EXPERIMENT

In this section we present the experimental design (3.1), state the formal hypotheses we shall test (3.2), and give the experimental results (3.3).

3.1 Design

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

---

the restaurant owners may have analogous incentives. For example, if Björn visited the *Crab House* restaurant at Pier 39 in San Francisco, his waitress would give him a plastic card which reads (in six languages): "Thank you for dining with us. Many guests ask us about tipping. We want you to know that no additional tip or service charge has been added to your bill. In the United States, quality service is rewarded with a tip, or gratuity, of at least 15%."

[13] Charness & Rabin (2001) consider a form of communications, asking first movers to express preferences between the moves available to responders. They find strong evidence that responder behavior is sensitive to these expressed preferences (but only when the first mover has not 'misbehaved'). To the extent that expressed preferences are believed to be meaningful, they should affect responder behavior under an expectations-based model. However, beliefs are not measured in this study.

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

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

The treatment without messages began with each A choosing IN or OUT. Next, each B indicated (without knowing A's actual choice) whether he or she wished to choose ROLL or DON'T ROLL, *contingent upon A having chosen IN*, as B's choice is immaterial if A has chosen OUT. We thus obtain an observation for every B.[14] The outcome corresponding to a successful project occurred if and only if the die came up 2, 3, 4, 5, or 6 after a ROLL choice.

In the message treatment, each B had an option to send a non-binding message to A prior to A's decision concerning IN or OUT. Each B received a sheet, on which any (non-identifying) message could be written, if desired. B could also decline to send a message by circling the letter B at the top of the otherwise-blank sheet.[15]

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

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

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

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

As explained earlier, the motivation for our experiment made it crucial to measure some of the subjects' beliefs. After we collected the 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.[16] B's were 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. This form of belief elicitation sharpens the incentives in comparison with quadratic-scoring rules, at the cost of the exclusion of (rational) guesses of less than 5% or greater than 95%. As our game is one-shot and we don't mention guesses until after actions have been chosen, the fact that we elicit beliefs should not affect participants' prior choices.

3.2 Hypotheses

[16] 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.

The numbering 1-4 of the hypotheses below reflects the connection to the questions Q1 – Q4 listed in the introduction.

If an agent has purely selfish preferences, it is a dominant strategy to to choose zero effort when effort is not observable. Anticipating this, the principal will reject the contract. With respect to Q1, this leads to:

> **H1:** *No agent chooses high effort with hidden action. No principal ever accepts the contract.*

In accordance with findings in many experimental studies, where many people seem not to maximize own money, it is not unreasonable to expect that H1 might be rejected. In that case, the remaining hypotheses presented below come into play. With regard to Q2, we have:

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

Regarding Q3, the form and content of communication, we examine this issue by classifying the messages into different categories, and by considering how the agents' behavior and the outcomes correlate with these categories. Of course, it is in principle possible to cut the classification boundaries in a large number of ways.[17] Our categorization is tailored to fit the specific social preferences models we test (cf. section 2.3 above and hypothesis H4(a) and H4(b) below). Recall, in particular, the idea, discussed in section 2.3, that with belief-dependent utility (kindness-based reciprocity or guilt aversion) the agent may wish to convey a message about his intentions to the principal. Against this background, our classification keeps track of messages that include a 'statement of intent'.

More specifically, the classification takes the following perspective: An agent may or may not send a message; if he does send a message, it may or may not contain an affirmative

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

statement concerning his intention to behave in a manner favorable to the principal. We consider whether we can discern patterns of agent and principal behavior according to these categories. Our null hypothesis on this point is:

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

Our final question (Q4) concerns the types of motivational forces that are able to organize the data well. We focus on the three approaches discussed in Section 3 and the testable predictions derived there.

The distributional models of Bolton & Ockenfels and Fehr & Schmidt predict that communication does not matter. This implication again suggests testing H2, but in order to make clear the conceptual motivation and connection to Q4 we state a separate hypothesis:

> **H4(a):** *Same statement as H2.*

The other two approaches discussed in Section 3 were kindness-based reciprocity and guilt aversion. We derived correlations of opposite sign between the frequency of high-effort choices and agents' beliefs concerning principals' beliefs concerning the agents' high-effort choices. Our experimental design allows us to measure the agents' second-order beliefs, so we can test these predictions. The following hypothesis is relevant in both cases:

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

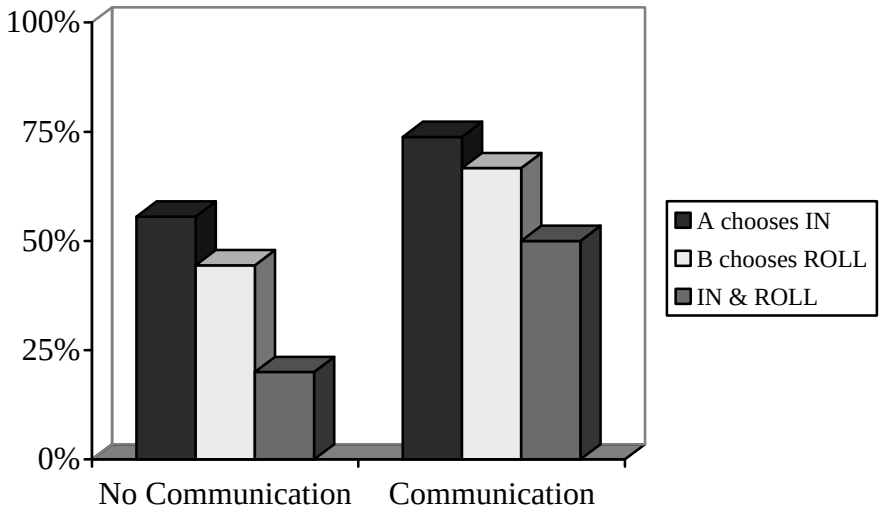
If kindness-based reciprocity were important for cooperation, we would expect to reject H4(b) in favor of the alternative hypothesis with *negative* correlation; if guilt aversion were important for cooperation, we would expect to reject H4(b) in favor of the alternative hypothesis with *positive* correlation.

We now turn to an examination of our results in light of these hypotheses.

3.3 Results

Figure 4 summarizes principal and agent choices. These data are needed to test hypotheses H1, H2, and H4(a). To test the other hypotheses requires some additional data (on message classification, and elicited beliefs), which we introduce later. Complete disaggregated results (by individual) are available at http://www.econ.ucsb.edu/~charness/papers.

Figure 4 - Observed Choices



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

H1: This is the classical hypothesis; we can readily reject it for both the no-communication and communication treatments, for both principals and agents. For the principals, the test of the difference of proportions (see Glasnapp & Poggio, 1985) gives Z = 5.88 and 7.01 for the respective treatments, with $p < 0.000001$; for the agents, this test gives Z

= 5.07 and 6.48, with similarly high statistical significance. Own monetary reward is not the only motivation present.

H2: This hypothesis states that the possibility of communication will not affect behavior. We can reject it for both principals and agents. In the no-communication treatment, 55.6% A's chose IN, while in the communication treatment this proportion increased to 73.8%. The test of the difference of proportions gives a test statistic of $Z = 1.78$, with significance at $p = 0.038$.[18] Without communication 44.4% B's chose ROLL, while with communication this proportion increased by half to 66.7%. The test of the difference of proportions gives a test statistic of $Z = 2.08$, with significance at $p = 0.019$. The likelihood of a successful partnership more than doubles, and the test of proportions gives $Z = 2.94$, $p = 0.002$. We may conclude that messages have a major influence on behavior.

H3: This hypothesis concerns whether we can find *ex post* patterns of behavior that depend on the particular content of a message. Specifically, does a statement of intent affect behavior and the resulting outcomes? One of the features of our design is that messages can have nearly any form. Nevertheless, we can group the agents' message choices into categories and see if certain types of messages are more effective than others. Appendix B presents the precise messages sent by the agents, and shows how we have classified the choices into three categories: *promises*, *empty talk*, and *no message*.[19] 24 of the 42 agents (57%) made promises to ROLL; 14 agents (33%) wrote messages not containing a statement of intent, and four agents (10%) sent no message.

We can readily reject H3 for principal behavior with communication, and reject the hypothesis more marginally for agent behavior. 22 of the 24 principals (92%) who received promises chose IN. This compares to eight of the 14 principals (57%) who received an empty-talk message, or nine of the 18 principals (50%) who did not receive a promise. The

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

[19] We distinguish between promises and empty talk by characterizing a promise as an affirmative statement on one's intention or plan to behave in a manner favorable to the principal.

differences in principal behavior are quite significant: $Z = 2.52$, $p = 0.006$ and $Z = 3.04$, $p = 0.001$ for the respective promise/empty talk and promise/no promise comparisons. With respect to agents' behavior, eighteen of the 24 promises (75%) were actually kept (the agent chose ROLL). (Note that the remaining 25% communicated a more or less outspoken lie.) This compares with only seven of the 14 empty-talk agents (50%) choosing ROLL, or 10 of the 18 non-promising agents overall. Comparing the ROLL rate with a promise to the rate with empty talk or non-promises overall, we see that the difference is weakly significant ($Z = 1.57$, $p = 0.06$, or $Z = 1.32$, $p = 0.10$, respectively).

We can also compare across treatments to see if not promising in the communication treatment (where promising is possible) yields the same outcomes as in the no-communication treatment (where it is not). Regarding agent behavior, neither the 7/14 ROLL rate nor the 10/18 ROLL rate is significantly different than the 20/45 ROLL rate in the no-communication treatment ($Z = 0.36$ and 0.80, for the respective comparisons). Similarly, for principal behavior, the 8/14 or 9/18 IN rates with empty talk or non-promises are nearly the same as the 25/45 IN rate in the no-communication treatment ($Z = 0.10$ and 0.40, for the respective comparisons).

It is instructive to consider the outcomes achieved with and without promises by agents in the communication treatment. When a promise was made, the successful partnership outcome (IN, ROLL) resulted in 67% of the cases (16/24). This compares to 28% successful outcomes (5/18) for non-promises, and an expected rate of 25% in the no-communication treatment.[20] It seems clear that the observed difference between communication and no-communication treatments is driven not by messages *per se*, but rather by the promises made by agents.

There is a literature on self-serving deception in (signaling or bargaining) games with asymmetric information. These studies typically differ crucially from ours in that the monetary payoff structure of the game is not common information. It is nevertheless

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

interesting to note that certain aspects of the results compare well. The tendency of agents to not break promises in our setting meshes well with Gneezy's (2002) finding that people experience a cost of lying and the Brandts & Charness (2003) result that people dislike deception *per se*. Nevertheless, these and several other studies show that many people engage in deception, much like those six participants in our design who made a promise that they did not keep.[21]

H4(a): Distributional models predict that communication does not influence behavior (or at least an agent's behavior; cf. footnote 9). We already know that this prediction does not stand, because the hypothesis H4(a) has the same content as the rejected hypothesis H2. Distributional models appear to be inadequate for addressing the observed impact of communication.

H4(b): How well do kindness-based reciprocity and guilt do in explaining our belief-action data? Recall that if reciprocity matters we expect to reject the hypothesis in favor of the alternative with negative correlation, while if guilt aversion is the motivation, we expect a positive correlation.

In the no-communication case, agents who chose DON'T ROLL guessed that principals who chose IN guessed on average that 40.4% of agents would ROLL, compared to the 53.2% guessed by the B's who chose ROLL. A Wilcoxon-Mann-Whitney rank-sum test (see Siegel & Castellan, 1988) finds that the guesses of the ROLL group are significantly higher ($Z = 1.99$, $p = 0.046$, two-tailed test). This gap across agents widens considerably in the communication treatment, where agents who chose DON'T ROLL guessed 45.1% and agents who chose ROLL guessed 73.2% ($Z = 3.20$, $p = 0.001$, two-tailed test). Thus H4(b) is rejected in favor of the alternative hypothesis with positive correlation, consistent with guilt aversion.

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

An examination of agent choices as a function of guesses about the guess of those principals who chose IN confirms the positive relationship between agent choices and second-order beliefs. This is shown in Figures 5 and 6:

FIGURE 5 - B play in guess ranges

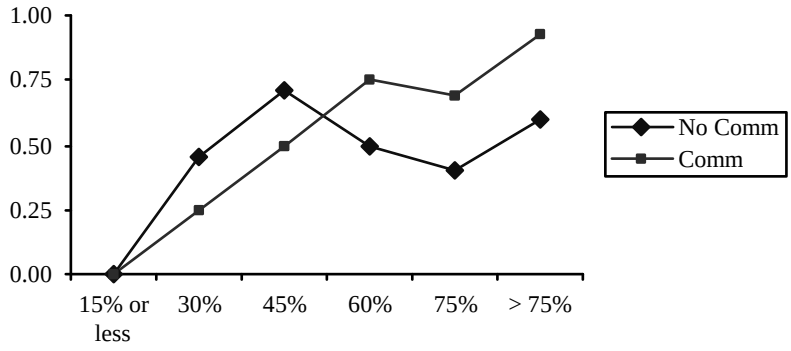
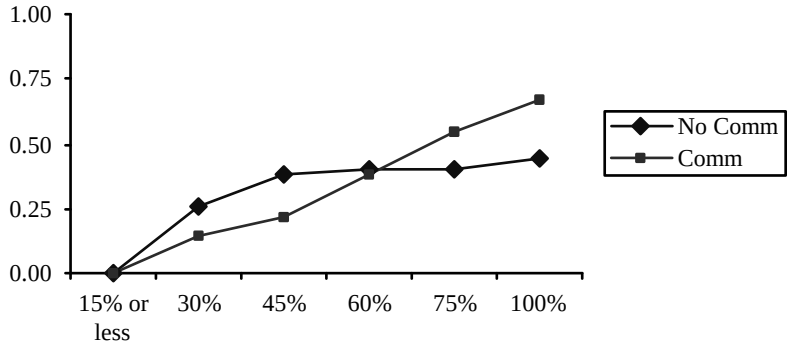


FIGURE 6- Cumulative B play in guess ranges



There is a definite positive trend in both cases; the more likely that an agent who believes that the principal expects him to ROLL is generally more likely to choose ROLL. This tendency is much more pronounced when communication is possible.

There is some interaction between our results on guesses about guesses and the earlier results on promises (cf. H3). In particular, agents who make a promise and choose ROLL exhibit higher guesses about principals' guesses than do other agents (Z = 1.70, p = 0.045,

one-tailed Wilcoxon test). A possible interpretation is that promises foster high expectations, which coupled with guilt aversion creates a commitment device for agents to ROLL.

4. DISCUSSION

We present experimental evidence concerning behavior and motivation in a game designed to capture the essence of hidden action, as treated in much of contract theory. Hidden action turns out to be less of a stumbling block in the lab than classical contract theory predicts, and communication within the partnership mitigates the problem further. 'Statements of intent' by agents seem particularly responsible for the benefits observed. Such *promises* dramatically affect the behavior of both principals and agents, resulting in a much higher proportion of successful partnerships.

Why do these results hold? What model of decision-making captures the relevant effects? The questions should be probed, since the answers inform subsequent research. To know which motivational forces are at work and why and how communication influences the interaction is essential for deriving insights regarding optimal contracting, and for formulating testable predictions in other settings to which the results may extend. We examine whether several approaches to modeling social preferences in games can explain our findings.

We first look at models in which decision-makers care only about the overall final distribution of monetary payoffs in a game, focusing in particular on the Bolton & Ockenfels and Fehr & Schmidt models of inequity aversion. These models are relatively simple to define, to test empirically, and to apply theoretically. One may therefore wish to use them even if they only get predictions *approximately* right (cf. Fehr, Klein & Schmidt, 2001, p. 9). Nevertheless, it is important to assess the range of situations for which such an approach makes sense. We check whether the models come close to capturing how communication influences strategic interaction in a setting with hidden action. The answer is negative; in the game we consider, distributional models predict that communication does not matter, a proposition that our data clearly refute. Distributional models can explain the behavior seen

in *either* the no-communication or the communication treatment, but cannot explain the behavioral differentials *across* treatments.

This finding leads us to address more complicated models, based on psychological game theory in which decision-makers have belief-dependent utilities. We consider two models in particular: kindness-based reciprocity and guilt aversion. These models imply that the likelihood of the agent's high-effort choice is correlated with his belief about the principal's belief of that choice. That correlation is predicted to be negative with reciprocity and positive with guilt aversion; thus, our data speak in favor of the guilt-aversion explanation. The stronger an agent believes that his principal believes that the agent will make the high-effort choice, the *more* likely the agent is to do this (not *less* likely, as kindness-based reciprocity would have it).

We have three comments. First, our case for rejecting kindness-based reciprocity is weaker than meets the eye, because it relies on an auxiliary assumption that may be unjustified. As explained in section 3.2, we *assume* independence between the agent's reciprocity-sensitivity (as expressed by a certain parameter in Dufwenberg & Kirchsteiger's theory) and his degree of belief that his principal believes that the agent will make the high-effort choice (τ''). We have little justification for this assumption, beyond lack of obvious alternative, so our rejection of kindness-based reciprocity should be taken with a grain of salt. More research, aimed at revealing *why* beliefs (such as τ'') come about, is clearly needed.

Second, our result, insofar it is valid, only concerns *positive* kindness-based reciprocity and guilt aversion when the principal made a choice favorable to the agent. Reciprocity has two sides, positive reciprocity, where a player is kind in return to another's kind choice, and negative reciprocity, where a player is unkind in return to another's unkind choice. In our game, the only way the principal can be unkind is by not agreeing to the partnership. In that case the agent has no subsequent choice, so we cannot observe negative reciprocity. In fact, negative reciprocity seems an important motivational force in other games.[22] In addition, the fact that preference expression in Charness & Rabin (2001) is

[22] A proper test of negative reciprocity should involve belief measurement, since the relevant theory involves belief-dependent utilities. It is nevertheless worth noting that numerous experimental studies suggest that

completely ineffective after unfavorable play by the first mover strongly suggests that expectations-based models have little or no bite in negative contexts.

Third, using guilt aversion instead of kindness-based reciprocity as the underlying motivation for cooperative behavior does not imply disregard of previous experimental findings. Rather, it is a reinterpretation of these findings that is suggested. For example, results in gift-exchange games as well as in trust games are usually taken to illustrate reciprocal forces at work (see Fehr & Gächter, 2000). We propose that guilt aversion be considered a serious candidate for understanding the results. This would mean that the more effort a 'worker' expects that his 'firm' expects him to exert, the more effort he will choose if the firm has offered a decent wage, because if he did not he would feel guilty letting the firm down. If firms believe that higher wages trigger higher effort, and agents believe this, the oft-observed positive wage-effort relationship will result. Another example could be the investment game studied by Berg, Dickhaut & McCabe (1995). Given the evidence consistent with guilt aversion in other 'trust' games (this study; Dufwenberg & Gneezy, 2000; Bacharach et al., 2001), it would seem quite plausible that guilt aversion also underlies the Berg *et al.* (1995) findings.

An intriguing aspect of guilt aversion is its potential as *a theory of why communication matters*. Our design permits us to elicit certain beliefs (about actions and beliefs), and leads us to conclude that trustworthy behavior by agents is consistent with their desire to live up to the expectations of the principal with whom they are matched.[23] This suggests a plausible channel through which *non-binding communication may influence behavior*. By making a promise to behave in a trustworthy fashion, the agent strengthens the

---

negative reciprocity is important (Kahneman, Knetsch & Thaler, 1986; Blount, 1995; Charness, 1996; Offerman, 2002; Brandts & Charness, 1999; Andreoni, Brown & Vesterlund, 2002; Kagel & Wolfe, 2001; Charness & Rabin, 2002) and a handful of studies that suggest that positive reciprocity is not (Charness, 1996; Offerman, 2002; Cox, 2000; Charness & Rabin, 2002).

[23] We note that promises may matter more when the party making a promise knows that the other party will find out whether the promise is kept (which may still be consistent with a hidden-action environment, if these facts are not provable in court). In this respect, there is an issue of whether internal norms or external norms drive the behavioral effect. Is there a difference between conforming to internal norms (one's view of what is appropriate) or external norms (the expectations of one's counterpart)? Charness & Grosskopf (2000) consider the interplay between cheap talk and such information provision, and find that providing information substantially enhances the degree of coordination on the payoff-dominant equilibrium. This seems to be a worthwhile topic for future research.

principal's degree of belief that the agent can be relied upon, and the agent's own belief that the principal reacts this way creates an incentive for him to live up to his promise.[24] A promise feeds a beneficial, self-fulfilling circle of beliefs, beliefs about beliefs, and trustworthy behavior. Truth-telling becomes 'self-enforcing'; in a trust situation there will be no incentive to lie or renege on a promise. (As noted in Section 2.3, kindness-based reciprocity does not possess such 'honesty-sustaining' features.)

Guilt aversion thus suggests a reason why people may experience a cost of lying, which make them keep promises, or avoid stating falsehoods. Gneezy (2002) models lying as having an individual-specific (but otherwise constant across circumstances) internal positive cost, and shows that this has explanatory value in a deception experiment. Our findings suggest that these costs may depend on one's beliefs about whether a promise, or a lie, was believed by the person to whom it was made. These belief-dependent costs of guilt may, however, extend beyond promises and lies. Why and how do people discuss, argue, and debate? How do such exchanges influence group decisions, formation of and adherence to social norms, partnership formation, and contracting? Perhaps a key aspect is that these exchanges lead up to commonly-expected standards of behavior or judgment which, once in place, are shared and not violated by guilt-averse people.

This suggests a promising lode to explore in future research. Communication and guilt aversion may be important not only in simple mono-contractual settings with hidden action, but also in other partnership situations involving hidden information, multiple contracts, or richer strategic possibilities. First, one might explore various variations close to our chosen design, to explore the robustness of our findings. What is the impact on behavior and beliefs of, for example, varying details of the communication protocol, say, having messages from the principal or having interactive responses between principals and agents?[25]

---

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

[25] Comparison with related work on *other* games is then possible. For example, a form of one-sided communication by principals appears in some recent gift-exchange studies: firms offer contracts consisting of wage and *desired effort*; see Fehr, Gächter & Kirchsteiger (1997), Fehr & Gächter (2002), and Fehr, Klein & Schmidt (2001). The last two studies report positive correlations between desired and actual effort (though not

Second, one may consider settings plagued by hidden information. One may reasonably suspect that again communication fosters cooperation. Agents might say, "I promise I am a person with high talent", and perhaps guilt-averse people with low talent find it unbearable to lie about such matters. Examining this proposition could shed light on whether or not hidden information is more or less problematic than hidden action.

Third, there is ample scope for theoretical work. Contract theory has a history of basking in the light of tremendous intellectual achievement. That tradition should be continued. As experimentalists accumulate insights regarding which behavioral ideas that seem to have bearing on understanding partnerships and contract, it makes sense to develop new theory based on these ideas. How might one, for example, characterize optimal contractual arrangements when agents are affected by guilt aversion? To answer this question seems to us an exciting challenge in behavioral contract theory.

Although more work is needed to delineate more exactly the range of situations in which guilt aversion plays a role,[26] we propose that the idea be seriously considered as fundamental for understanding partnerships, contracts, and human interaction quite generally. Examples range from everyday experiences, like tipping in a café, to many kinds of partnerships, including husband & wife, lawyer & client, procurement agency & contracted firm, inventor & producer, talented young golfer & rich sponsor, co-owners of firms, employer & employee, etc. To underscore the potential scope of the idea, we close our paper by sketching out in somewhat more detail one specific application, drawn from industrial organization rather than classical contract theory.

A particular kind of partnership where promises may play a role is a *cartel*. We first note a paradox: Against the backdrop of conventional economic theory, selfish decision-makers and irrelevant promises, the modern competition law that prohibits certain 'concerted practices' among firm appears superfluous. In order to fine companies, a common

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.

26 Clearly that is not *always* the case; it is hard, *e.g.*, to imagine poker players feeling guilty, no matter how they deceive others. On the other hand, if a social norm (or expectation) is perceived to be more pervasive ('tighter'), perhaps a greater degree of adherence will result from guilt aversion. Another issue concerns expectations that may be deemed 'unreasonable' or 'obnoxious'. How would this be determined, and what might be the effect?

prerequisite is that contacts between company representatives have been documented (e.g., meetings prior to price changes; see Hovenkamp, 1996). Agreements made during such contacts can hardly build on much else but promises, since cartel agreements are illegal and will not be enforced in court.

If an oral contract is not worth the paper it's written on, one may wonder if the rules against concerted practices are necessary![27] The skeptical view of oral contracts stands in some contrast to findings in experimental industrial organization, which suggest that communication may foster collusion; see the survey by Holt (1995, pp. 409-11). We have shown that communication may matter in a different context, and point to a specific motivational force that may be responsible. In a competition context, it is conceivable that an 'oral contract', under which guilt-averse cartelists promise one another to stick to monopoly pricing, serves as the glue that makes cartelists stick together. The experiment in this paper is not concerned with a competition game, so we shall not push the example further, but merely propose the matter for further research. Experiments could investigate how promises, discussions, agreements, and other forms of information exchange influence beliefs, beliefs about beliefs, competition in markets, and the ability of cartelists to collude.

[27] Some economists have made claims that go even further. McCutcheon (1997) claims that these rules actually foster collusion. It is the combination of the game-theoretic notion of renegotiation-proof equilibrium, the idea that the competitors are completely selfish, and the assumption that competition law makes discussions among firms somewhat more costly that produce this effect. In McCutcheon's model, firms lie without remorse, and a victim of such a ploy would enthusiastically and perpetually re-negotiate with the competitor who fooled him.

REFERENCES

Anderhub, Vital, Simon Gächter & Manfred Königstein (2002), "Efficient Contracting and Fair Play in a Simple Principal-Agent Experiment", *Experimental Economics* **5**, 5-27.

Andreoni, James, Paul Brown & Lise Vesterlund (2002), "What Produces Fairness? Some Experimental Evidence", *Games & Economic Behavior*, **40**, 1-24.

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

Baumeister, Roy, Arlene Stillwell & Todd Heatherton (1994), "Guilt: An interpersonal approach", *Psychological Bulletin*, **115**, 243-267.

Baumeister, Roy, Arlene Stillwell & Todd Heatherton (1995), "Personal narratives about guilt: role in action control and interpersonal relationships", *Basic & Applied Social Psychology*, **17**, 173-98.

Berg, Joyce, John Dickhaut & Kevin McCabe (1995), "Trust, Reciprocity, and Social History," *Games & Economic Behavior*, **10**, 122-42.

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

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

Blount, Sally (1995), "When Social Outcomes Aren't Fair: The Effect of Causal Attributions on Preferences", *Organizational Behavior & Human Decision Processes*, **LXIII**, 131-144.

Blume, Andreas, Douglas DeJong, Yong-Gwan Kim & Geoffrey Sprinkle (2001), "Evolution of Communication with Partial Common Interest", *Games & Economic Behavior*, **37**, 79-120.

Bohnet, Iris and Bruno Frey (1999), "The sound of silence in prisoner's dilemma and dictator games", *Journal of Economic Behavior & Organization*, **38**, 43-57.

Boles, Terry, Rachel Croson, Keith Murnighan (2000), "Deception and retribution in repeated ultimatum bargaining", *Organizational Behavior & Human Decision Processes*, **83**, 235-259.

Bolton, Gary & Axel Ockenfels (2000), "ERC: A Theory of Equity, Reciprocity, and Competition", *American Economic Review*, **90**, 166-193.

Brandts, Jordi & Gary Charness (2003), "Truth or Consequences: An Experiment", *Management Science*, **49**, 116-130.

Brosig, Jeannette, Axel Ockenfels & Joachim Weimann (forthcoming), "The Effect of Communication Media on Cooperation", *German Economic Review*.

Cabrales, Antonio & Gary Charness (2000), "Optimal contracts, adverse selection, and social preferences: An experiment," mimeo.

Charness, Gary (forthcoming), "Attribution and Reciprocity in an Experimental Labor Market", *Journal of Labor Economics*.

Charness, Gary (2000) “Self-serving Cheap Talk and Credibility: A Test of Aumann’s Conjecture”, *Games & Economic Behavior*, **33**, 177-194.

Charness, Gary & Brit Grosskopf (2000), “Cheap Talk, Information, and Coordination: Experimental Evidence, mimeo.

Charness, Gary & Matthew Rabin (2001), “Expressed Preferences and Behavior in Experimental Games”, mimeo.

Charness, Gary & Matthew Rabin (2002), “Understanding Social Preferences with Simple Tests”, *Quarterly Journal of Economics*, **117**, 817-869.

Cooper, Russell, Douglas DeJong, Robert Forsythe & Thomas Ross (1990), “Selection Criteria in Coordination Games: Some Experimental Results”, *American Economic Review*, **53**, 218-233.

Cooper, Russell, Douglas DeJong, Robert Forsythe & Thomas Ross (1992), “Communication in Coordination Games”, *Quarterly Journal of Economics*, **53**, 739-771.

Cox, James (2000), "Trust and Reciprocity: Implications of Game Triads and Social Contexts", mimeo.

Cox, James & Daniel Friedman (2002), "A tractable model of reciprocity and fairness", mimeo.

Crawford, Vincent (1998), “A survey of experiments on communication via cheap talk”, *Journal of Economic Theory*, **78**, 286-298.

Croson, Rachel (2002), “Game-Theoretic and Experimental Perceptions of Deception”, mimeo.

Croson, Rachel, Terry Boles & J. Keith Murnighan (forthcoming), “Cheap talk in bargaining experiments: Lying and threats in ultimatum games”, *Journal of Economic Behavior & Organization*.

Dawes, Robyn, Jeanne McTavish & Harriet Shaklee (1977), “Behavior, communication, and assumptions about other people’s behavior in a commons dilemma situation,” *Journal of Personality & Social Psychology*, **35**, 1-11.

Dufwenberg, Martin (2002), "Marital investment, time consistency, and emotions", *Journal of Economic Behavior & Organization*, **48**, 57-69.

Dufwenberg, Martin & Uri Gneezy (2000), “Measuring beliefs in an experimental lost wallet game”, *Games & Economic Behavior*, **30**, 163-82.

Dufwenberg, Martin & Georg Kirchsteiger (1998), “A theory of sequential reciprocity”, Research paper 1998:1, Department of Economics, Stockholm University.

Dufwenberg, Martin & Michael Lundholm (2000), “Social norms and moral hazard”, *Economic Journal* , **111**, 506-25.

Dutta, Prajit K. & Roy Radner (1993), "Moral Hazard", in Robert Aumann & Sergiu Hart (eds.), *Handbook of Game Theory*, vol. 2.

Ellingsen, Tore & Magnus Johannesson (2002), "Promises, Threats, and Fairness", mimeo.

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

Farrell, Joseph & Matthew Rabin (1996), "Cheap Talk", *Journal of Economic Perspectives*, **10**, 103-118.

Fehr, Ernst & Simon Gächter (2000), "Fairness and Retaliation: The Economics of Reciprocity", *Journal of Economic Perspectives*, **14**, 159-181.

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**, 833-860.

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**, 817-868.

Fehr, Ernst & Klaus Schmidt (2001), "Theories of Fairness and Reciprocity - Evidence and Economic Applications," forthcoming in: M. Dewatripont, L. Hansen & St. Turnovsky (Eds.), *Advances in Economics & Econometrics - 8th World Congress, Econometric Society Monographs*.

Forsythe, Robert, Russel Lundholm, Thomas Rietz (1999), "Cheap Talk, Fraud, and Adverse Selection in Financial Markets: Some Experimental Evidence", *Review of Financial Studies*, **12**, 481-518.

Geanakoplos, John, David Pearce & Ennio Stacchetti (1989), "Psychological games and sequential rationality", *Games & Economic Behavior*, **1**, 60–79.

Glasnapp, Douglas & John Poggio (1985), *Essentials of Statistical Analysis for the Behavioral Sciences*, Columbus, Merrill.

Gneezy, Uri (2002), "Deception: The Role of Consequences", mimeo.

Güth, Werner, Wolfgang Klose, Manfred Königstein & Joachim Schwalbach (1998), "An experimental study of a dynamic principal-agent relationship", *Managerial & Decision Economics*, **19**, 327-341.

Hart, Oliver D. & Bengt Holmström (1986), "The Theory of Contracts", in Truman Bewley (ed.), *Advances in Economic Theory*, Cambridge University Press.

Holmström, Bengt (1979), "Moral Hazard and Observability", *Bell Journal of Economics*, **10**, 74-91.

Holt, Charles (1995), "Industrial Organization: A Survey of Laboratory Research", in *Handbook of Experimental Economics*, eds. J. Kagel and A. Roth, Princeton, NJ: Princeton University Press.

Hovenkamp, Herbert (1994), *Federal Antitrust Policy: The Law of Competition and Its Practice*, West Publishing Company, St. Paul Minnesota.

Huang, Peter & Ho-Mou Wu (1994), "More Order without More Law: A Theory of Social Norms and Organizational Cultures", *Journal of Law, Economics & Organization*, **10**, 390-406.

Jamison, Julian (2000), "Valuable cheap talk and equilibrium selection", mimeo.

Kagel, John & Katherine Wolfe (2001), "Tests of Fairness Models Based on Equity Considerations in a Three-Person Ultimatum Game", *Experimental Economics*, **4**, 203-219.

Kahneman, Daniel, Jack Knetsch & Richard Thaler (1986), "Fairness and the Assumptions of Economics", *Journal of Business*, **59**, S285-S300.

Krone, Heinz Walter (2002), "Individual differences in emotional reactions and coping", forthcoming in *Handbook of Affective Sciences* (eds. R.J. Davidson, K.R Scherer, H.H. Goldsmith), New York: Oxford University Press.

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 behavioural economics", *The Economic Journal*, **109**, F25-F34.

McCutcheon, Barbara (1997), "Do meetings in smoke-filled rooms facilitate collusion", *Journal of Political Economy*, **105**, 330-50.

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

Nyarko, Yaw & Andrew Schotter (2002), "An experimental study of belief learning using real beliefs", *Econometrica*, **70,** 971-1005.

Offerman, Theo (1998), "Hurting Hurts More than Helping Helps: The Role of the Self-serving Bias", forthcoming in *European Economic Review*.

Orbell, John, Robyn Dawes & Alphons van de Kragt (1990), "The Limits of Multilateral Promising", *Ethics*, **100**, 616-627.

Rabin, Matthew (1993), "Incorporating fairness into game theory and economics", *American Economic Review*, **83**, 1281-1302.

Rousseau, Denise (1995), *Psychological Contracts in Organizations: Understanding Written and Unwritten Agreements*, Thousand Oaks, California: Sage Publications.

Segal, Uzi & Joel Sobel (1999), "Tit for Tat: Foundations of Preferences for Reciprocity in Strategic Settings", mimeo.

Selten, Reinhard (1967), "Die Strategiemethode zur Erforschung des Eingeschränkt Rationalen Verhaltens im Rahmen eines Oligopolexperiments", in *Beiträge zur Experimentellen Wirtschaftsforschung*, H. Sauermann, ed., 136-168.

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

Sobel, Joel (1999), "Interdependent Preferences and Reciprocity", mimeo.

Tangney, June (1995), "Recent Advances in the Empirical-Study of Shame and Guilt," *American Behavioral Science*, **38**, 1132-1145.

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**, 211-238.

APPENDIX A - INSTRUCTIONS

[*text in the message treatment is shown in brackets*]

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

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

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

Decision tasks

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

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

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

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

[**A Message**

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

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

B

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

A

MAKE A GUESS

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

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

Payment for the guess

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

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

B

MAKE A GUESS

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

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

Payment for guess:

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

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

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

APPENDIX B - MESSAGES

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

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

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

3	6	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	R	IN
3	7	I <u>will roll</u>, 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	R	IN
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	DR	IN
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	R	IN
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	R	IN
3	11	If you choose <u>IN</u> you have the best opportunity to make the most money. You have a 5/7 chance of making more money! So <u>IN</u> would be your best bet. Cheers. ☺	E	DR	IN
3	12	Choose IN. I promise I'll ROLL.	P	R	OUT

Figure 1

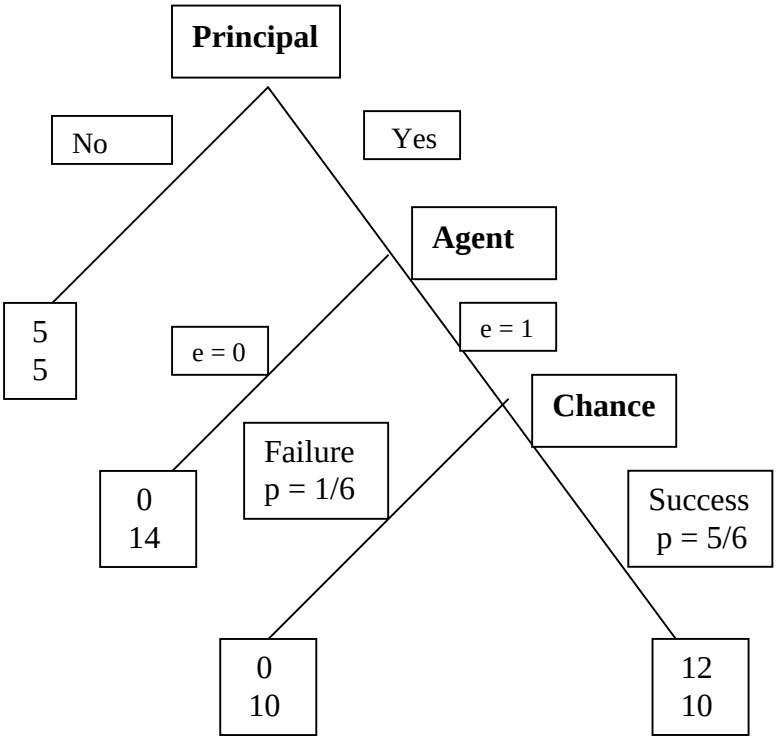

Principal
No
Yes
5
5
Agent
e = 0
e = 1
0
14
Failure
p = 1/6
Chance
Success
p = 5/6
0
10
12
10

Figure 2

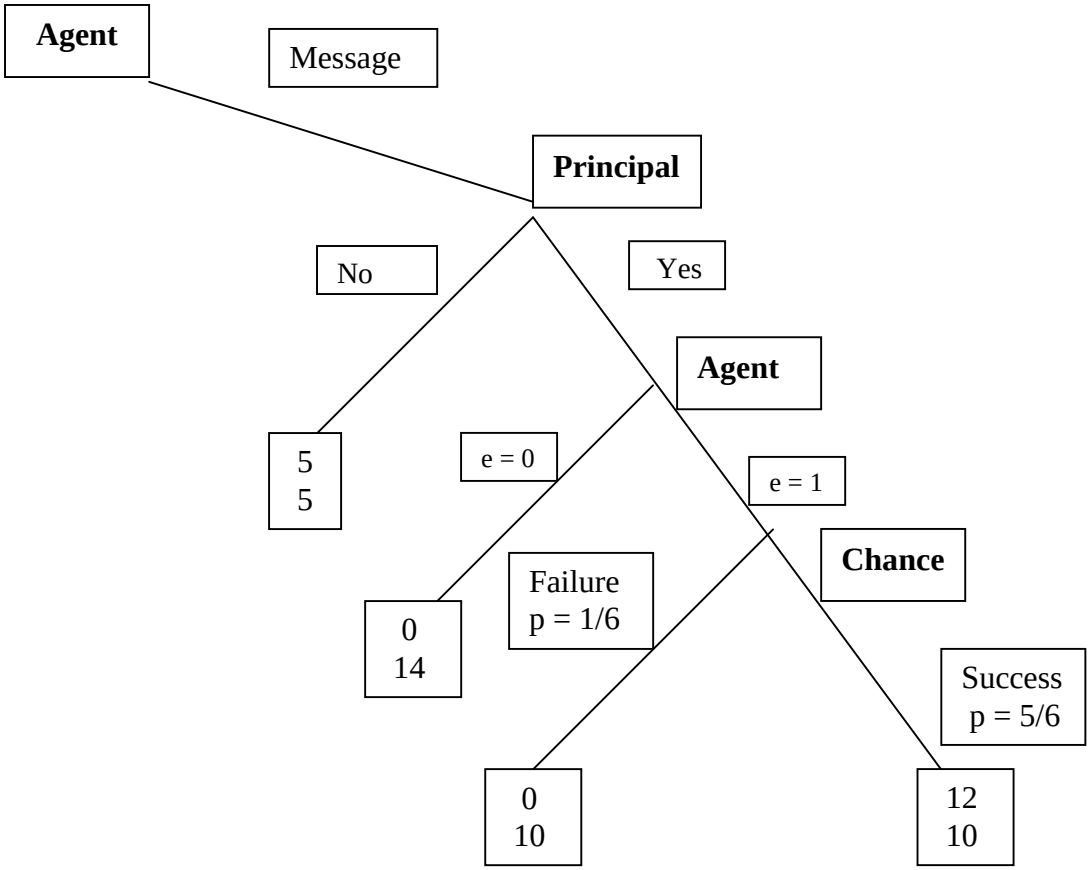
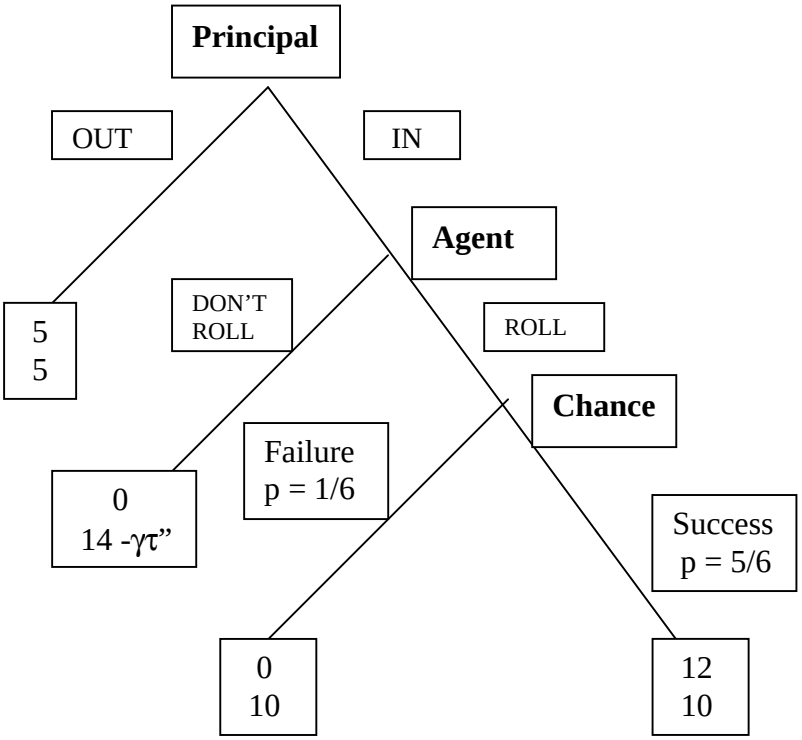

Agent
Message
Principal
No
Yes
5
5
Agent
e = 0
e = 1
0
14
Failure
p = 1/6
Chance
Success
p = 5/6
0
10
12
10

Figure 3


Principal
OUT
IN
5
5
Agent
DON'T ROLL
ROLL
Chance
0
14 -γτ"
Failure
p = 1/6
Success
p = 5/6
0
10
12
10