

# Contracts & Communication

Gary Charness & Martin Dufwenberg\*

November 4, 2008

**ABSTRACT:** Feelings of guilt may influence trust & cooperation. This can affect partnership interaction so that market failure due to hidden information is avoided. Communication may help further as words move beliefs to make guilt matter more. We analyze this in a framework that merges contract theory with psychological games. Predictions are derived regarding the form that communication will take. We design experiments to test for empirical relevance and find that promises ameliorate the hidden-information problem if *all* involved parties can reap some gain, but not if efficiency calls for players of certain types to step aside and forego any gain.

**KEYWORDS:** Adverse selection, hidden information, guilt-from-blame, psychological games, communication, trust, lies, deception, social preferences, behavioral economics

**JEL CODES:** A13, B49, C72, C91, D63, D64, J41

**ACKNOWLEDGMENTS:** We have benefited from helpful discussions with Pierpaolo Battigalli and Tore Ellingsen, and we also thank participants at the Arne Ryde Symposium on Communication in Games and Experiments at Lund University, the IWEBE conference in Lyon, the St. Andrews Applied Microeconomics Workshop, the ESA conference in Lyon, and seminar participants at the University of Gothenburg and Brown University for helpful comments. The research was conceived with the support of the Russell Sage Foundation and completed with the support of the National Science Foundation (grant # SES-0617923). Dufwenberg did part of the research while visiting the University of Gothenburg and he is grateful for the hospitality.

\* **CONTACT:** Gary Charness, Department of Economics, University of California at Santa Barbara, [charness@econ.ucsb.edu](mailto:charness@econ.ucsb.edu); Martin Dufwenberg, Department of Economics and Institute of Behavioral Economics, University of Arizona, [martind@eller.arizona.edu](mailto:martind@eller.arizona.edu).

# 1. INTRODUCTION

Human collaboration has produced much in the world. Research in contract theory (often collaborative efforts!) explores which partnerships form, what contracts are signed, and what the consequences will be. Considerable attention has been given to settings with hidden action (where a party's future choice is not contractible) or hidden information (where a contract cannot be conditioned on a party's private information). When parties act opportunistically, these are hurdles that may preempt fruitful collaboration.<sup>1</sup>

A less gloomy picture, not stressed by contract theorists, emerges if a contracting party feels guilty if he hurts another. Psychologists have documented that guilt reduces well-being and that behavior changes in its anticipation (making the associated pangs counterfactual). According to Baumeister, Heatherton & Stillwell (1994; cf. Baumeister *et al* 1995, Tangney 1995), "the prototypical cause of guilt would be the infliction of harm, loss, or distress on a relationship partner. ... [I]f people feel guilt for hurting their partners ... and for failing to live up to their expectations, they will alter their behavior (to avoid guilt) in ways that seem likely to maintain and strengthen the relationship". We believe modern businessmen qualify as 'relationship partners' (perhaps especially if they took MBA courses in business ethics).

We incorporate guilt feelings to a contract-theoretic model with hidden information, show that guilt may foster trust & cooperation, argue that communication among the contracting parties may strengthen these effects further and in predictable ways, and finally examine the predictions in an experiment. In theorizing, we employ the tools for modeling guilt in games

---

<sup>1</sup> For an entry to the literature, see Bolton & Dewatripont (2005). The gloomy outlook can be exemplified with reference Akerlof's (1970) classic work on hidden information: The seller of a used car knows its quality while the buyer does not. This creates an obstacle to reaching socially-attractive agreements, and market failure results. The terms hidden action and hidden information are often called, respectively, moral hazard and adverse selection. We prefer the terminology in the text, which is more descriptive and which seems less suggestive of the nature of outcomes.

devised by Battigalli & Dufwenberg (2007) (hereafter B&D), who in turn use the toolbox of psychological game theory (Geanakoplos, Pearce & Stacchetti 1989, Battigalli & Dufwenberg 2008).

Our approach complements that of Charness & Dufwenberg (2006), who consider a hidden-action context. The similarity with the current hidden-information exercise is that both approaches transform Nash bargaining settings into certain trust games. However, the games are very different in details, in particular regarding the nature of the trust needed for efficiency to prevail. Under hidden action, a principal must rely on an agent to not act opportunistically but there is no doubt that the agent could deliver in principle. This is different from hidden information, where some agents (with low talent) simply cannot deliver as well as others. Hidden information involves an asymmetry that lacks a counterpart in the hidden-action case. If the efficient outcome is to occur, then a principal must rely on low-talent agents to accept a lower gain than high-talent agents. An extreme special case occurs if low-talent agents have to step aside and forego all gain, to permit principals to have undisturbed top-quality interaction with high-talented agents. We consider both these two cases (accept-a-lower-gain and step-aside), and ask whether guilt and communication promote trust and cooperation.

Besides shedding light on the empirical relevance of some behavioral theory, we note that our results will reveal some seemingly rather stable patterns regarding how language is used strategically, and how words correlate with opportunism and trustworthiness. There may be ‘lessons-for-life’ to take away for both confidence tricksters who wish to improve their deceptive skills and for detectors who wish to build better traps.

The contract-theoretic underpinnings are presented in section 2. The predictions from guilt-from-blame are presented in section 3. The experiment is described in section 4, and the experimental results are presented in section 5. Section 6 offers concluding remarks.

## 2. CONTRACTUAL UNDERPINNINGS

In this section we introduce a Nash bargaining partnership model (2.1), incorporate hidden information (2.2), and finally present a modified game that offers no scope for low-talent agents to have mutually fruitful dealings with the principal (2.3). Our main goal with section 2.1 is to provide a contract-theoretic backdrop to our design. However, the material may have independent value providing a starting point also for other experiments. While we derive a  $2 \times 2 \times 2$  game capturing hidden information, the model could also be used to incorporate hidden action or richer contractual or strategic settings.

### 2.1 Nash bargaining

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 outside-option payoffs (dollars) of  $x$  for the principal and  $y$  for the agent. If the project is carried out, then the contract specifies a ‘wage’  $w$  that the principal must pay the agent. The project generates revenue for the principal. There can be two outcomes: poor or good. A poor outcome generates revenue  $r > 0$ , while a good outcome involves an additional bonus of  $b > 0$  so that total revenue is  $r + b$ . The probability of these outcomes depends on the choice and characteristic of the agent; she chooses ‘effort’,  $e \in [0, 1]$ , has a given ‘talent’,  $t \in [0, 1]$ , and the

probability of a good outcome is  $e \cdot t$ . The agent experiences increasing ‘effort cost’, measured in dollars and equal to  $c \cdot e$ , where  $c > 0$ .

Consider the Nash bargaining solution for risk-neutral and selfish players, assuming effort and wage to be contractible and that all other parameters are commonly known. Following Nash (1950), the solution will be the wage-effort combination  $(w, e)$  that maximizes

$$[(r + e \cdot t \cdot b - w) - x] \cdot [(w - c \cdot e) - y] \quad (1)$$

if it is possible to choose  $w$  and  $e$  so that both factors of the product (1) are positive; otherwise no partnership will form. Three main cases stand out:<sup>2</sup>

- *A partnership forms with high wage & effort:*  $w = (r + t \cdot b - x + y + c)/2$  and  $e = 1$ . This outcome is predicted for high-talent agents if  $r$  and  $b$  are high enough, or more precisely if  $e \cdot t \cdot b > c \cdot e$  for  $e > 0$  and  $r + e \cdot t \cdot b - c \cdot e > x + y$  for large enough  $e$ .
- *A partnership forms with low wage & effort:*  $w = (r - x + y)/2$  and  $e = 0$ . This outcome is predicted for low-talent agents if  $r$  is high enough, or more precisely if  $e \cdot t \cdot b < c \cdot e$  for  $e > 0$  and  $r > x + y$ .
- *No partnership forms:* The outcome is predicted if the partnership is a bad business venture, or more precisely if  $r + e \cdot t \cdot b - c \cdot e < x + y \quad \forall e \in [0, 1]$ .

## 2.2 Incorporating hidden information

From now on, fix the following parameter values (noting that there are two possibilities for  $t$ ):

$$\begin{array}{lll} r = 14 & c = 4 & x = 5 \\ b = 12 & t \in \{0, 5/6\} & y = 5 \end{array} \quad (2)$$

Suppose that just before the principal and the agent meet, the relevant value of  $t$  is determined by chance;  $t = 0$  with probability  $2/3$  and  $t = 5/6$  with probability  $1/3$ . If both the principal and the

---

<sup>2</sup> For generic parameter constellations  $(x, y, r, b, c, t)$ , these are *all* the cases. Other outcomes would be possible if the parameters combine to produce certain equalities. We shall not deal with such cases (and in fact limit attention to the specific parameterizations described in subsections 2.2 and 2.3).

agent know the realization, the first two bullets of section 2.1 imply that the two relevant Nash bargaining solutions are:  $(w,e) = (7,0)$  for  $t = 0$  and  $(w,e) = (14,1)$  for  $t = 5/6$ .

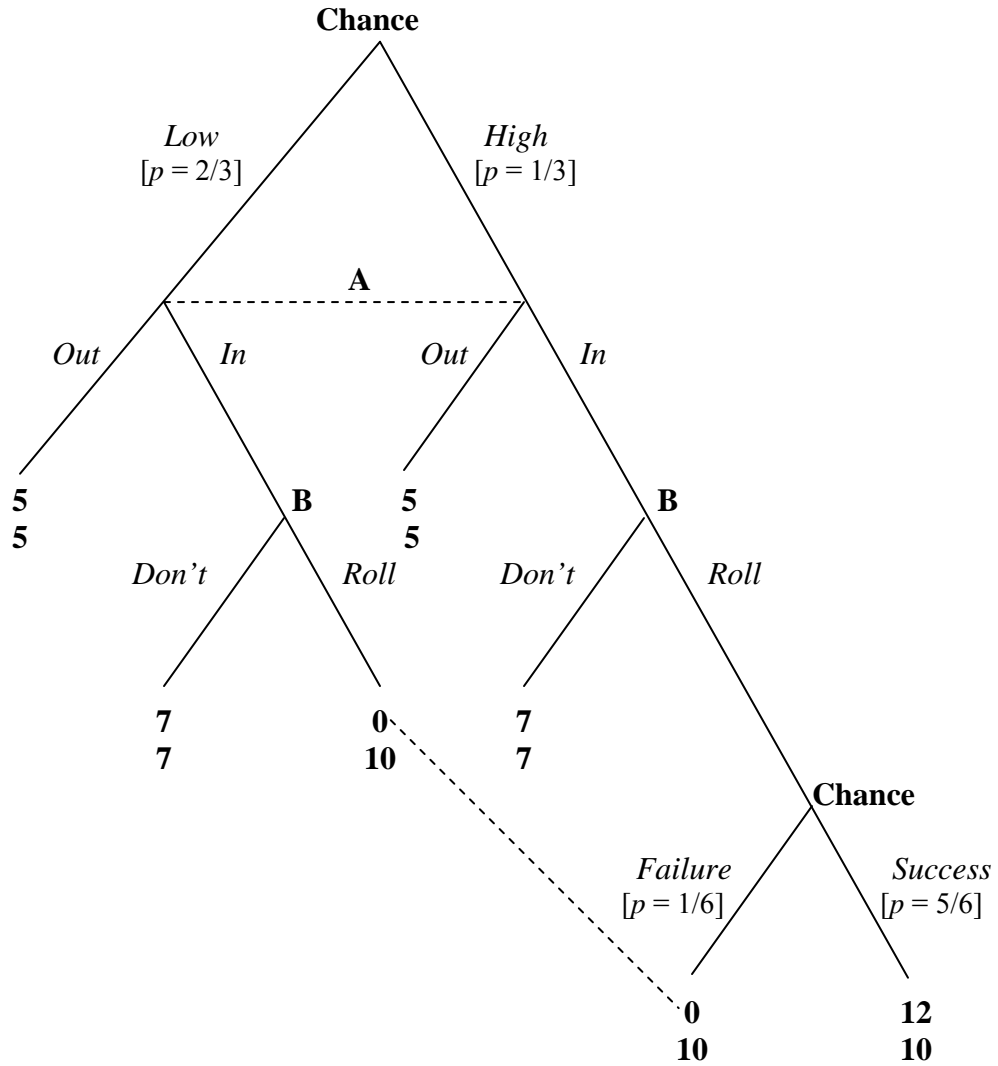
Can these outcomes also be achieved if only the agent knows the value of  $t$ , while the principal knows only the prior probabilities  $1/3$  and  $2/3$ ? There are two good reasons why the answer may be *no*. First, since only the agent knows the value of  $t$ , she will somehow have to (credibly) reveal this information in order for the relevant wage-effort contract to be chosen [*i.e.*,  $(w,e) = (7,0)$  if  $t = 0$  and  $(w,e) = (14,1)$  if  $t = 5/6$ ]. This invites an opportunity for an agent with  $t = 0$  to *represent* she has  $t = 5/6$ , so that she gets the higher wage of 14 (and a net material payoff of 10) rather than a wage of 7. Second, the principal may foresee such a turn of events, dislike it, and not agree to form a partnership.

The extensive “(7,7)-game” in Figure 1 incorporates these two possibilities, in a particular way. The naming of strategies/players anticipates the upcoming wording of the experimental instructions. Chance moves first; the choice of *Low* or *High* corresponds to the determination of the types  $t = 0$  or  $t = 5/6$  for the agent. The principal is player A; his choice of *In* or *Out* reflects whether or not he agrees to forming a partnership. The agent is B; her choice to *Don't (Roll)* or *Roll* (a 6-sided die) reflects the choice between wage-effort combination  $(w,e) = (14,1)$  or  $(7,0)$ . Finally, the chance move following path (*High, In, Roll*) determines whether the project is a failure or a success. The dotted line connecting A's payoffs of \$0, following paths (*Low, In, Roll*) and (*High, In, Roll, Failure*), indicate an information set for A *across endnodes*.<sup>3</sup> This reflects how A is never told how his payoff of \$0 came about.

---

<sup>3</sup> Information sets across endnodes are typically not given in standard game theory as they would have no bearing on equilibrium play. However, in psychological games such information can critically affect play (as our discussion in section 3, about  $\lambda$ , will indicate). See Battigalli & Dufwenberg (2008, section 6.2) for more discussion of this point.

**Figure 1: The (7,7)-game**



Why have we included this chance move that determines the project's success, rather than just replace it with its expected outcome (10,10)? The answer is that this provides a *conceptual* justification for our claim that the game incorporates hidden information. This is a circumstance where a contract *cannot* be conditioned on a party's private information; here this applies to the agent's talent. A typical justification for such a contractual limit, often stressed by contract theorists, is that the agent's type is not observable to the principal, or at least not

verifiable in court.<sup>4</sup> The chance move justifies a story where a low-type agent (talent  $t = 0$ ) could falsely claim that she was in fact a high-type agent (talent  $t = 5/6$ ) but that she had bad luck. Because of the chance move, it cannot be proven in court that she lied.

If the players are selfish and risk-neutral, the game of Figure 1 has a unique subgame-perfect equilibrium: A chooses *Out*; B chooses *Roll* independently of her talent.<sup>5</sup> The players earn 5 each independently of B's talent. The outcome is inefficient, since each party would receive expected payoffs of 8 if A chose *In* and B chose *Don't* with low talent and *Roll* with high talent.<sup>6</sup> This illustrates how hidden information may undermine efficient contracting. Note that the conclusion is independent of whether or not the players communicate before A moves; words can't change the fact that B gets a higher dollar payoff from *Roll* than from *Don't*, and given this A gets a higher dollar payoff from *Out* than from *In*.

How should one react to this? One possibility is to take the indicated problem at face value, and to examine whether *other* contractual arrangements may help overcome the problems caused by hidden information. This sort of approach is typical in contract theory; the optimal choice of contract when a partnership is influenced by hidden information is a major issue, and the assumption that the principal and the agent are selfish is typically maintained.

We do *not* follow that approach. We are skeptical of the traditional premise that parties are selfish, and we test an alternative theory that can allow the rosy Nash bargaining outcome of section 2.1 to obtain even in the face of hidden information (as embodied in Figure 1) and can

---

<sup>4</sup> Thus type-contingent contractual clauses choices are not enforceable. If, however, outcomes under a given contract were perfectly correlated with the agent's type then her type could (arguably) be inferred with certainty, so a type-contingent contract could be enforceable in principle.

<sup>5</sup> Given that B chooses *Roll* independently of his talent, A's best response is *Out*. That choice gives A a payoff of 5, whereas *In* gives A an expected payoff of  $(1/3) \times [(5/6) \times 12 + (1/6) \times 0] + (2/3) \times 0 = 10/3$ .

<sup>6</sup> The value 8 derives from  $(1/3) \times [(5/6) \times 12 + (1/6) \times 0] + (2/3) \times 7$  for A and  $(1/3) \times 10 + (2/3) \times 7$  for B.

also allow communication to play a major role. Section 3 elaborates on what we have in mind, but before that we consider an alternative parameterization.

### 2.3 A step-aside game

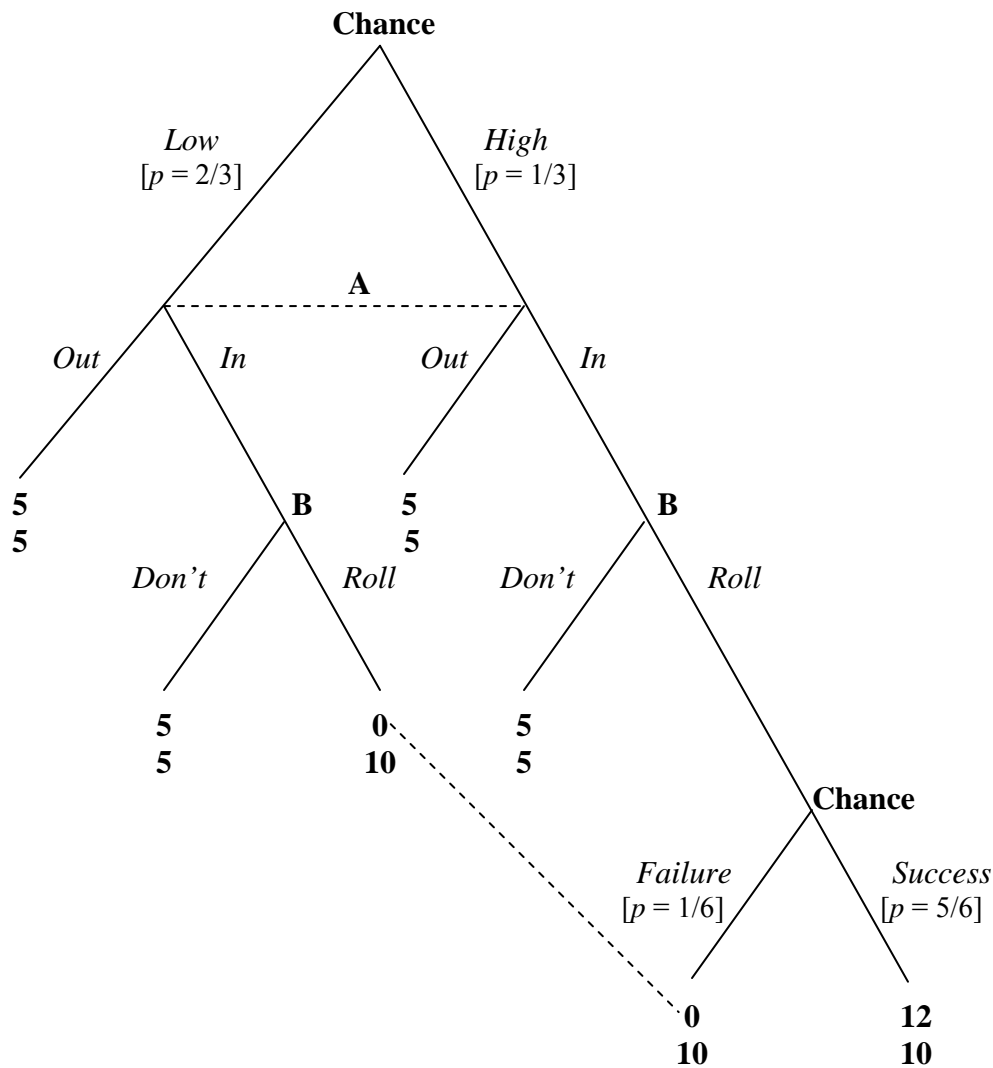
In the Nash bargaining solutions described in section 2.2 (for  $t = 0$  and for  $t = 5/6$ ) high-talent agents who contract move from a (monetary) payoff of 5 to a payoff of 10 (in expected terms) while low-talent agents who contract move from a payoff of 5 to a payoff of 7. This illustrates how with hidden information gains-from-trade may be asymmetric as regards different types of agents. One can imagine a more extreme form of such asymmetry, where the low-talent agents are simply incapable of making net additions to partnership profit. Perhaps they induce lower  $r$ 's, or perhaps government taxation is so high that all gains from trade get wasted when low-talented agents are involved. Alternatively, perhaps there is only one position to fill and many available agents, so that the principal is only interested in hiring a high-talent agent.<sup>7</sup> In such a setting it could be that low-talent agents should step aside, according to the yardstick of Nash bargaining (the third bullet case of section 2.2 would apply).

Suppose nothing changes as regards the probability and productivity of high-talent agents. “The (5,5)-game” in Figure 2 would model such a setting:

---

<sup>7</sup> In section 2.2 we described a situation where the agent was offered a choice between two contracts  $(w,e) = (7,0)$  and  $(w,e) = (14,1)$ . The ‘one job’ interpretation would entail a choice between the contract  $(w,e)=(14,1)$  and a no-partnership option.

**Figure 2: The (5,5)-game**



Parametrically, the change from Figure 1 to Figure 2 looks small: four 7's are replaced by four 5's. The interpretation of the *Don't* choice now changes, to reflect a step-aside move. The backward-induction solution (with and without communication) for selfish players moreover does not change (cf. section 2.2). As we shall discuss in more detail the next section, incorporating behavioral concerns and communication may allow for better outcomes although, as we shall see, the case for this to happen is weaker than in the (7,7)-game.

### 3. GUILT-FROM-BLAME PREDICTIONS

As seen in section 2, if players are selfish then an inefficient outcome is predicted, a conclusion unchanged when agents also can communicate. Models of social preferences such as Fehr & Schmidt (1999), Bolton & Ockenfels (2000), and Charness & Rabin (2002) can accommodate more cooperative behavior if many participants harbor dislike for payoff inequality or have tastes for social efficiency. This holds for both the (7,7)-game and the (5,5)-game, but entering into a contract is more attractive to the principal in the (5,5)-game since a low-talent agent is less likely to be selfish when the choice is between (7,7) and (0,10) rather than between (5,5) and (0,10). Thus, from these models, we would expect more low(-talent)-B's to choose *Don't* and so more A's to choose *In*. However, none of these models predicts a differential effect of communication in these different games. We are interested in the role of communication in fostering trust & communication and focus on and test a different behavioral model.

B&D develop theories of guilt aversion, defined for a general class of game forms that includes those of Figures 1 and 2. Due to the chance move following choice *Roll*, the following critical issue comes up: Say the strategy profile played is (*In*, (*Roll*, *Roll*)), meaning that A chooses *In* and the strategy of both type of B's is *Roll*. A's dollar payoff is 0 and B is responsible. Yet A does not know this for a fact, and B knows A doesn't know. To the extent that B is thus not 'exposed', perhaps this should shelter B from some pangs of guilt? B&D develop two notions of guilt, simple guilt and guilt-from-blame, that answer that question *no* and *yes*, respectively. In light of several recent studies (Dana, Cain & Dawes 2006, Dana, Weber & Kuang 2007, Broberg, Ellingsen & Johannesson 2007, Tadelis 2008) indicating that

opportunistic behavior is stymied particularly when subjects know they can be held accountable, we will go with B&D's notion of guilt-from-blame, according to which guilt is felt by  $i$  to the extent that  $j$  blames  $i$  for causing  $j$  to get a lower dollar payoff than  $j$  expected to get. This entails that a player will be sheltered from pangs of guilt to the extent that other players cannot infer that he hurt them.<sup>8</sup> The description of such guilt relies on higher-order beliefs, which need to be modeled up to the fourth order.<sup>9</sup>

The exposure issue becomes especially pertinent when one factors in the additional role of communication. Consider a version of the (7,7)-game where after learning his type, B sent a message to A. Suppose B lies about his talent and manages to completely convince A his talent is high. When A gets his \$0 payoff he will think it was just bad luck, and therefore not blame B. According to the guilt-from-blame concept, B would not feel guilty at all. This feature of the theory is so stark that one may suspect it to be unrealistic, and yet for the purposes of what we shall test in this paper we stick with this prediction in the name of keeping a clear focus.

While applying guilt-from-blame is somewhat tricky, we demonstrate how the machinery can be used in our environment. The key idea is that the pangs a low-B must suffer depend on the degree to which B thinks A can figure out that B chose *Roll*. Suppose A gets a \$0 payoff and assigns probability  $\lambda$  to a low-B having chosen *Roll* and probability  $1 - \lambda$  to a high-B having chosen *Roll*, but with bad luck. The variable  $\lambda$  is key to describing how exposure affects guilt-from-blame. The smaller B believes  $\lambda$  is, the less B suffers from bad conscience, *ceteris paribus*.

---

<sup>8</sup> B&D state (pp. 171-2): "Whereas with simple guilt, a player cares about the extent to which he lets the other player down, [guilt-from-blame] assumes that a player cares about others' inferences regarding the extent to which he is willing to let them down". This distinction had not yet been conceptualized when we wrote Charness & Dufwenberg (2006); in retrospect we see that we focused on simple guilt in that paper.

<sup>9</sup> As we felt it would be rather unrealistic to collect meaningful fourth-order beliefs, we chose not to elicit beliefs in the experiment. As we shall see, the theory generates testable predictions based on many other observables.

B&D develop a notion of sequential equilibrium (SE), which gives guilt-from-blame-based predictions for general game forms. In Appendix A we derive predictions for the games in Figures 1 and 2 in some detail; here in the main text, we give a *Reader's Digest* version. For this purpose we refer to a key parameter: B's sensitivity-to-guilt parameter  $\theta \geq 0$ . If  $\theta = 0$  we are back in a selfish setting where guilt-from-blame is irrelevant. Otherwise, the higher is  $\theta$ , the more sensitive is B to guilt-from-blame.

*(7,7)-Game without Communication*

Refer to the (7,7)-game in Figure 1. It is impossible to sustain 'full cooperation' in the sense that the strategy profile in which A chooses *In*, low-B's choose *Don't*, and high-B's choose *Roll* is not a SE for any value of  $\theta$ . The intuition: If inferences were based on low-B's choosing *Don't* with probability 1, then low-B couldn't be blamed for choosing *Roll* as we'd get  $\lambda=0$ . A would credit the \$0-payoff to path (*High, In, Roll, Failure*). A would thus not blame B, so B would suffer no guilt-from-blame choosing *Roll*.

However, if guilt sensitivity is sufficiently high it is possible to come *close* to full cooperation: there is a SE such that A chooses *In* and a low-B's choose *Roll* with probability  $x = \frac{1}{28\theta-12} \leq \frac{3}{14}$ ; note that  $x$  tends to 0 as  $\theta$  goes to infinity.<sup>10</sup> The intuition: Low-B's are indifferent between *Don't* (which gives relatively low material payoff and no guilt) and *Roll* (which gives higher material payoff but some guilt). The probability of *Roll* (i.e.,  $x$ ) is precisely at the level that makes the two choices equally attractive.<sup>11</sup>

---

<sup>10</sup> This SE requires  $\theta \geq \frac{25}{42}$ ; see Appendix A for details.

<sup>11</sup> The formulation presumes a given  $\theta$  and either direct randomization by B or a population interpretation where  $x$  is a fraction of a population from which the low-B player is drawn. On the latter interpretation, one can alternatively have incomplete information regarding  $\theta$  and the fraction  $x$  with the lowest such values choosing *Roll*.

However, for a range of guilt sensitivities there are multiple SEs that can be Pareto ranked in terms of (expected) utilities. The nice SE described in the previous paragraph co-exists with another SE in which A chooses *Out*, low-B's choose *Roll*, and high-B's choose *Roll*.<sup>12</sup> In light of this multiplicity, it thus seems guilt-from-blame *may or may not* help improve efficiency.

*(7,7)-Game with Communication*

In order to derive predictions with communication, we adapt to our context with guilt-from-blame an idea from the literature on cheap talk in standard games: natural language conveys exogenously given *meaning* that is focal in the sense that players tend to believe what is said as long as such belief is consistent with rationality and the incentives given in the game.<sup>13</sup>

We focus on the role of *promises*. Consider the following two B-to-A messages:

***LD-message:*** “I am *Low* and I will choose *Don't*”

***HR-message:*** “I am *High* and I will *Roll*”

One may imagine variants, but we will focus on the LD- and HR-messages as representative of respective classes of promises with similar meaning. Suppose B gets to send either message to A, before play proceeds as in Figure 1. We make three observations.

First, we ask if it is consistent with rationality and the given incentives (including concern for guilt-from-blame) to have full revelation + separation + honesty: low-B's send LD-messages and choose *Don't*, high-B's send HR-messages and choose *Roll*. The answer is *no*.

<sup>12</sup> This requires  $\theta \leq \frac{39}{40}$  so (with reference to footnote 10) there are multiple equilibria if  $\frac{25}{42} \leq \theta \leq \frac{39}{40}$ .

<sup>13</sup> This rules out “babbling equilibria” where the message content is neglected. For relevant work that explores similar assumptions, see Rabin (1990), Farrell (1993), Farrell & Rabin (1996), Crawford (2003), Blume & Ortmann (2007), and Demichelis & Weibull (2008).

The argument is analogous to that which ruled out, for the (7,7)-game without communication, the profile in which A chooses *In*, low-B's choose *Don't*, and high-B's choose *Roll* as a SE. The key concern is that the truth of HR-messages can't be taken for granted: If inferences were based on HR-messages never coming from low-B's, then low-B's would be safe choosing *Roll* as they wouldn't be blamed for sending a HR-message and then choosing *Roll* (as we'd get  $\lambda = 0$ ).

Second, having thus seen what communication cannot achieve, we ask if it can nevertheless alleviate the equilibrium multiplicity problem described for the (7,7)-treatment without communication. The answer is *yes*. To see this, note two things: (i) an LR-message is self-committing for a low-B with a high enough  $\theta$ , in the sense that if he sent that message and if he believed that A believed it then he would not want to renege (because  $\lambda$  would equal 1 so he would be fully exposed). (ii) We already know that HR-messages cannot be fully trusted; see the previous paragraph. However, we can consider the possibility that HR-messages are as credible as possible, under the constraint that they be consistent with rationality and the given incentives. Suppose therefore that while all high-B's send HR-messages, so do low-B's with probability  $x$  who then go on to choose *Roll*. With remaining probability  $1-x$  low-B's send LD-messages and then choose *Don't*. For there to be no incentives to deviate, we need  $x$  (and, via

Bayesian updating,  $\lambda = \frac{\frac{2}{3}x}{\frac{2}{3}x + \frac{1}{3} \cdot \frac{1}{6}} = \frac{12x}{12x + 1}$  to be just high enough that low-B's are

indifferent between the first option (which gives relatively low material payoff and no guilt) and the second option (which gives higher material payoff but some guilt). On balance, play proceeds just as in the best SE of the (7,7)-game without communication. If promises are believed to the extent that this is consistent with the given incentives, this selects the best SE.

Third, we make observations regarding communication tactics and the costs of being caught lying. Consider these two patterns of low-B behavior:

***LD-then-D***: send a LD-message followed up by choice *Don't*

***HR-then-R***: send a HR-message followed up by choice *Roll*

Because of the nature of the subsequent choice, we may think of approaches LD-then-D and HR-then-R as, respectively, *trustworthy* and *opportunistic*. We also point out that the following alternative approaches to being trustworthy and opportunistic seem available:

***HR-then-D***: send a HR-message followed up by choice *Don't*

***LD-then-R***: send a LD-message followed up by choice *Roll*

Note that if we consider a situation where A responds to either message (LD or HR) with *In*, then tactics LD-then-D and HR-then-D lead to the same strategy profile in the game of Figure 1, and so do tactics HR-then-R and LD-then-R. We now argue that there is a good reason to believe that tactics LD-then-D and HR-then-R will nevertheless be more prevalent than tactics HR-then-D and LD-then-R. Namely, HR-then-D and LD-then-R entail *exposed lies*, which LD-then-D and HR-then-R do not. To see this, note that LD-then-D entails a promise that A's payoff will be 7, and A will indeed get 7. By contrast, HR-then-D entails a promise that A's payoff will be 0 or 12, but A will get 7 and so it will become evident that B lied. Similarly, HR-then-R entails a promise that A's payoff will be 0 or 12, and A will indeed get 0 or 12 (in fact, 0). By contrast, LD-then-R entails a promise that A's payoff will be 7, but A will get 0 so it will become evident that B lied. If B does not like to be caught lying, LD-then-D is thus preferable to HR-then-D and HR-then-R is preferable to LD-then-R.

The preference derived here is not based on guilt-from-blame in that it relates to the truth-value of messages rather than the degree to which B believes A believes B lets A down, but it is similar in that it relates to a form of exposure regarding the extent to which A can draw inferences regarding B's motives (as captured via  $\lambda$  before, and now via inferences via realized payoffs for A). We also find it natural to explore the LD-then-D versus HR-then-D and HR-then-R versus LD-then-R issues in light of several recent experimental contributions that suggest that cost-of-lying may matter to human motivation,<sup>14</sup> even if this work has not focused on a distinction between cost-of-lying *per se* and cost-of-lying that depends on the degree to which others can infer that one lied.

*(5,5)-Game, with and without Communication*

Refer to the (5,5)-game in Figure 2. The analysis proceeds analogously as in the (7,7)-games; to save space we do not go through the motions again (Appendix A contains formal details). Again, we can come arbitrarily close to full cooperation if  $\theta$  is high enough, although for each relevant value of  $x$  we now need a higher value of  $\theta$  than in the case of the (7,7)-game. Intuitively, it is harder to sustain cooperation in the (5,5)-games than in the (7,7)-games, as going for (7,7) is a more attractive option for a low-B than going for (5,5). Guilt-from-blame still allows for multiple equilibria for a (higher) range of values of  $\theta$ .

Adding communication may again help select the more efficient equilibria also in the (5,5)-game. However, the argument for this to happen is weaker in the case of the (5,5)-game for the following reason: In both the (7,7)-game and the (5,5)-game an LD-message is self-

---

<sup>14</sup> See Gneezy (2005), Hurkens & Kartik (2007), Vanberg (2008), and Charness & Dufwenberg (2008). For some theoretical work on cost-of-lying, see Ellingsen & Johannesson (2004), Chen, Kartik & Sobel (2007), Demichelis & Weibull (2008), and Kartik (2008).

committing for a low-B with a high enough  $\theta$  in the sense that if he sent that message and if he believed that A believed it then he would not want to renege. However, the incentive to embark on the associated LD-then-D approach is less strong for B in the (5,5)-game; in the (7,7)-game B changes the monetary-payoff vector from (5,5) to (7,7) if he makes A choose *In* rather than *Out*, whereas in the (5,5)-game the monetary-payoff vector will be (5,5) regardless of A's choice.

So on balance we expect less trust & cooperation in the (5,5)-game than in the (7,7)-game, and somewhat less benefit from communication. In addition to the theoretical arguments presented so far, the only way for a low-B to gain in the (5,5)-game is at the expense of A. Perhaps this makes low-B's less trustworthy, perhaps A's become less willing to trust, and perhaps communication won't help.

#### **4. EXPERIMENTAL DESIGN**

In line with the presentation in section 3, we have a 2×2 design. The first treatment variable concerns whether subjects played the (7,7)-game or the (5,5)-game. In each case we have one-shot interaction, to rule out any reputation or repeat-game effects. The second treatment variable concerns whether or not communication from B to A was allowed. While our main interest concerns the LD- and HR-message categories, we also wish to see if such messages are chosen spontaneously, so we give the sender a blank piece of paper on which to write any (anonymous) message instead of restricting the message space.<sup>15</sup>

Participants were recruited at UCSB by sending out an e-mail message to the campus community. We conducted 12 sessions, three for each of our treatments. Sessions were

---

<sup>15</sup> In Charness & Dufwenberg (2008), we show that restricting this message space can make a dramatic difference in the effectiveness of messages in a hidden-action environment.

conducted in a large classroom that was divided into two sides by a center aisle, and people were seated at spaced intervals. The number of participants in a session ranged from 24 to 36, for a total of 344 people; each person could only participate in one of these sessions. Average earnings were about \$14, including a \$5 show-up fee; each session was one hour in duration.

In each session, participants were referred to as ‘A’ (principal) or ‘B’ (agent). A coin was tossed to determine which side of the room was A and which side was B. Identification numbers were shuffled and passed out face down, and participants were informed that these numbers would be used to determine pairings (one A with one B) and to track decisions. Sample instructions are given in Appendix C.

Our first two treatments concern the (7,7)-game. We presented Table 1 to each of the participants, indicating the outcome for every combination of choices and die rolls. In one treatment, no messages were permitted, while in the other treatment, each B had an option to send a non-binding message to A prior to A’s decision concerning *In* or *Out*. B could decline to send a message by circling the letter B at the top of the otherwise-blank sheet. If an agent’s private identification number was evenly divisible by three, the agent was high talent; otherwise the agent was low talent.

**Table 1: Payoff Outcomes with (7,7) Option for Low B’s**

	<b>A receives</b>	<b>B receives</b>
A chooses OUT	<b>\$5</b>	<b>\$5</b>
A chooses IN and:		
B is LOW type and chooses DON’T ROLL	<b>\$7</b>	<b>\$7</b>
B is LOW type and chooses ROLL	<b>\$0</b>	<b>\$10</b>
B is HIGH type and chooses DON’T ROLL	<b>\$7</b>	<b>\$7</b>
B is HIGH type, chooses ROLL, die = 1	<b>\$0</b>	<b>\$10</b>
B is HIGH type, chooses ROLL, die = 2,3,4,5,6	<b>\$12</b>	<b>\$10</b>

After the decisions had been collected, a 6-sided die was rolled for each agent; this was explained in advance, to avoid the anticipated loss of public anonymity for agents who chose *Don't*. This roll was determinative if and only if (*In, Roll*) had been chosen.

Our third and fourth treatments concern the (5,5)-game. We presented Table 2 to each of the participants, indicating the outcome for every combination of choices and die rolls. In the first treatment, no messages were permitted, while in the second treatment, each B had an option to send a non-binding message to A prior to A's decision concerning *In* or *Out*. B could decline to send a message by circling the letter B at the top of the otherwise-blank sheet. Again, if an agent's private identification number was evenly divisible by three, the agent was high talent; otherwise the agent was low talent.

**Table 2: Payoff Outcomes without (7,7) Option for Low B's**

	<b>A receives</b>	<b>B receives</b>
A chooses OUT	<b>\$5</b>	<b>\$5</b>
A chooses IN and:		
B is LOW type and chooses DON'T ROLL	<b>\$5</b>	<b>\$5</b>
B is LOW type and chooses ROLL	<b>\$0</b>	<b>\$10</b>
B is HIGH type and chooses DON'T ROLL	<b>\$5</b>	<b>\$5</b>
B is HIGH type, chooses ROLL, die = 1	<b>\$0</b>	<b>\$10</b>
B is HIGH type, chooses ROLL, die = 2,3,4,5,6	<b>\$12</b>	<b>\$10</b>

The outcome corresponding to a successful project occurred if and only if the die came up 2-6 after a *Roll* choice. After the decisions had been collected, a six-sided die was rolled for each B; this was made clear to the participants in advance, to avoid the anticipated loss of public anonymity for B's. This roll was determinative if and only if (*In, Roll*) had been chosen.

In all treatments, each B first learned her type. After answering questions, the experimenter chose individuals at random to state the outcome for each possible case, starting

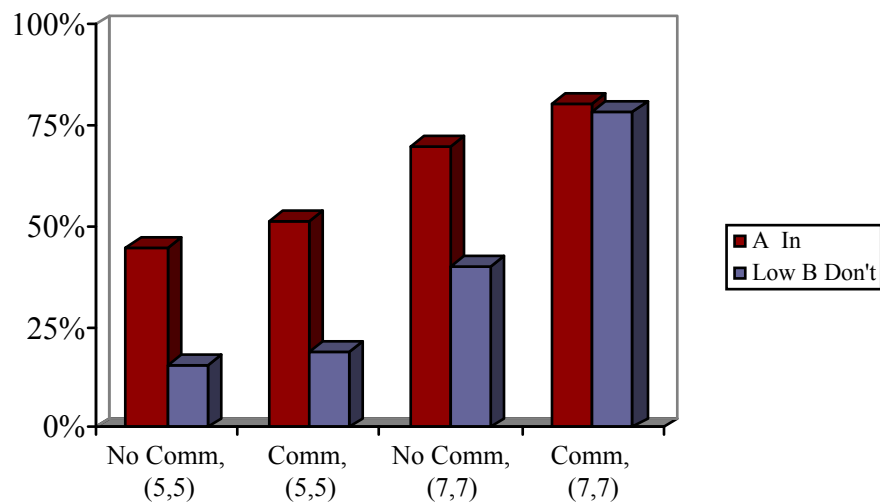
the session when it seemed clear that everyone understood the rules. In the message treatment, B had an option to send a free-form message to A prior to A's decision. Then A chose *In* or *Out*. Finally, B learned A's choice and, if A had chosen *In*, chose *Roll* or *Don't*.

## 5. EXPERIMENTAL RESULTS

### 5.1 Aggregated data

We find dramatically more type-revealing behavior for agents when the (7,7) outcome is feasible; A's are also more likely to choose *In* in this case. High-B choices are omitted, as they are invariably (48 of 48 times) *Roll* in our sessions. We also find that communication is quite effective in affecting low-B choices when there is a Pareto-improving option for low-B's, but it is almost totally ineffective when no such an option is available. Figure 3 presents A and low-B choices by session and treatment, and Table 3 summarizes the effect of communication on behavior for the (5,5)- and (7,7)-treatments:

**Figure 3 - Behavior Across Treatments**



**Table 3: Tests for the Effect of Communication**

Treatment	A's <i>In</i>		Z-stat	Low B's <i>Don't</i>		Z-stat
	M	NM		M	NM	
(7,7)	33/41 (80%)	28/40 (70%)	1.09	18/23 (78%)	8/20 (40%)	2.56***
(5,5)	24/47 (51%)	20/45 (44%)	0.64	3/16 (19%)	2/13 (15%)	0.24

M/NM mean that messages/no messages were feasible. The Z-stat reflects the test of proportions for the two populations. \*,\*\*and \*\*\* indicate  $p < 0.10$ ,  $0.05$  and  $0.01$ , respectively, one-tailed tests.

We note that the difference in A's *In* rates is significant across the (7,7)- and (5,5)-treatments, both with and without messages. The test of the difference in proportions (see Glasnapp & Poggio 1985) gives  $Z = 2.88$ ,  $p = 0.002$ , and  $Z = 2.37$ ,  $p = 0.009$ , respectively (one-tailed tests, as these are the natural predictions). Nevertheless, a substantial proportion of A's choose *In* even in the (5,5) treatment.<sup>16</sup> The difference in Low B's *Don't* rates across the (7,7)- and (5,5)-treatments is very large and significant ( $Z = 3.67$ ,  $p = 0.000$ ) with messages, but, while not insubstantial, is at best marginally significant ( $Z = 1.50$ ,  $p = 0.066$ ) without messages.

Communication only affects A's behavior to a modest and not significant degree. However, there is a powerful and significant effect of communication on the behavior of low-B's when (7,7) is an attainable outcome. It seems that many message senders behave differently after sending a message in the (7,7)-treatment, where it is possible to achieve Pareto improvement over the outside option.

---

<sup>16</sup> Note that this action is own-income maximizing (in expected terms) if A believes that as many as 1/6 of low-B's will choose *Don't* and that high-B's will choose *In*; in addition, some A's may be attracted by the social efficiency of the possible Pareto improvement if the agent ctually has high talent.

## 5.2 Message types and behavior

In Table 4 and Table 5 below, we break down our results with communication according to the type of message sent (all of the messages are shown in Appendix B):

**Table 4: Messages and Outcomes in (7,7)-Treatment**

		HR	LD	PNS	PH	PL	PD	E	NM	Total
Low B	<i>Out</i>	0	1	0	0	1	0	3	0	5
	<i>In, R</i>	4	0	0	1	0	0	0	0	5
	<i>In, DR</i>	0	7	2	1	0	4	3	1	18
	<b>Total</b>	4	8	2	2	1	4	6	1	28
High B	<i>Out</i>	1	0	0	0	0	0	1	1	3
	<i>In, R</i>	6	0	1	2	0	1	0	0	10
	<i>In, DR</i>	0	0	0	0	0	0	0	0	0
	<b>Total</b>	7	0	1	2	0	1	1	1	13

HR = Promise High & Roll, LD = Promise Low & Don't, PH = Promise High, PL = Promise Low, PNS = Promise not selfish, PD = Promise Don't, E = Empty Talk

**Table 5: Messages and Outcomes in (5,5)-Treatment**

		HR	LD	PNS	PH	PL	PD	E	NM	Total
Low B	<i>Out</i>	2	0	0	0	3	0	6	5	16
	<i>In, R</i>	4	0	1	0	0	0	3	5	13
	<i>In, DR</i>	0	1	0	0	0	0	1	1	3
	<b>Total</b>	6	1	1	0	3	0	10	11	32
High B	<i>Out</i>	6	0	0	0	0	0	1	0	7
	<i>In, R</i>	5	0	0	0	0	0	1	2	8
	<i>In, DR</i>	0	0	0	0	0	0	0	0	0
	<b>Total</b>	11	0	0	0	0	0	2	2	15

HR = Promise High & Roll, LD = Promise Low & Don't, PH = Promise High, PL = Promise Low, PNS = Promise not selfish, PD = Promise Don't, E = Empty Talk

There were 28 low-B's in the (7,7)-treatment. Of these, 18 made either an explicit or implicit promise to the paired A (see Appendix B); eight of these promises were of the form LD, while four were lies of the form HR. The remaining six promises do not fit the LD or HR categories; in four cases, B promised to choose *Don't* without specifying her type, and in two

cases there was a promise that B would not be selfish. Three B's stated their type, without designating an action, with a low-B claiming to be a high-B on occasions and a low-B stating she was a low-B on one occasion.<sup>17</sup> A's responded to these promises quite favorably – 17 of 18 (94%) chose *In* after receiving a promise (the choice of *Out* followed a LD message). Of the 10 A's who did not receive some form of promise from low-B's, six (60%) chose *In*. Even with this small sample, the difference is significant ( $Z = 2.28, p = 0.011$ , one-tailed test). Fourteen of the 18 Low B's (78%) who had a choice after sending a promise chose *Don't*.<sup>18</sup> Note that there were no instances of LD-then-R behavior and one instance of HR-then-D behavior.

Next, we consider the (5,5)-treatment. Communication does not lead to any substantial or significant differences for A's *In* rates. In striking contrast to the (7,7)-treatment, here there is virtually no effect from communication on B's *Don't* rates. A's who received a HR message chose *In* only eight of 17 times (47%). Sixteen of the 30 A's (53%) who did not receive an HR message chose *In*. These proportions are quite similar and of course are not significantly different ( $Z = 0.41$ ).<sup>19</sup> It appears that A's were more suspicious of HR messages in the (5,5) treatment, since the difference in *In* rates across treatments (in the (7,7)-treatment, A's chose *In* 10 of 11 times (91%) after a HR message) is significant ( $Z = 2.37, p = 0.018$ , two-tailed test).

Regarding B's behavior, in the (7,7)-treatment with communication, there were 18 cases where a low-B chose *Don't*. In 13 of these 18 cases (72%), low-B's sent a LD message, promised to choose *Don't*, or promised not to be selfish. Thus, the preponderance of those people (all low-B's) who chose *Don't* advertised that they would not take advantage of A. One

---

<sup>17</sup> It is tempting to simply classify these as HR and LD messages, respectively, as the meaning seems fairly clear. If we do so, then there were nine LD messages and six LD messages out of 28 possible messages.

<sup>18</sup> As we shall see, all five of the low-B's who chose *Roll* had claimed to have high talent.

<sup>19</sup> Note that the *In* rate of A's when communication was not permitted (44%) was also very similar.

low-B claimed to have high talent and then chose *Don't*, one sent no message, and three others chose messages classified as Empty Talk. There were five low-B's who chose *Roll*; in every case, the agent explicitly claimed to have high talent. Overall, of the six low B's who claim to have high talent, five chose *Roll*. All of the 17 other low-B's with choices in the communication treatment choose *Don't* when responding; this difference is highly significant ( $Z = 4.25, p = 0.000$ ). We see that B's who don't claim to have high talent can be relied upon to cooperate with paired A's who have trusted them. Furthermore, the comparison of the non-lying Low-B's in the communication treatment with the behavior of low-B's in the no-communication treatment, where 12 of 20 choose *Roll*, is also quite significant ( $Z = 3.89, p = 0.000$ ).

It appears that low-B's generally refuse to step aside when there is no available Pareto improvement over A's outside option. In fact, the behavior of the non-lying low B's is the driving difference between the (5,5)-treatment and the (7,7)-treatment. A comparison of the rate of lies supports this view: In the (5,5)-treatment, six explicit or implicit claims of high talent by the 32 low-B's (19%), while in the (7,7)-treatment, we observe six claims of high talent by the 28 low-B's (21%). There is no significant difference across communication treatments in the likelihood that a low-B will claim to have high talent ( $Z = 0.26$ ). However, we do see a major difference in the proportion of low-B's who choose to make no claims or promises, instead sending either an empty-talk message or none at all: while seven of 28 (25%) made such a choice in the (7,7)-treatment, 21 of 32 (66%) did so in the (5,5)-treatment. This difference is statistically significant ( $Z = 3.15, p = 0.002$ , two-tailed test). Clearly, low-B's in the (5,5)-treatment who did not wish to lie were at a loss for (meaningful) words.

How can we explain the difference in the effectiveness of communication across the (7,7)- and (5,5)-treatments? We appeal again to the line of argument put forth in section 3,

where higher values of the guilt-sensitivity parameter  $\theta$  are needed to support the more favorable mixed-strategy SE. In addition, we point out that all B's have the capability of being involved in a Pareto-improving combination of actions (at least in expected terms) in the (7,7)-treatment. Therefore, every agent can make a promise (such as a LD message) concerning a favorable outcome for the paired principal. However, in the (5,5)-treatment, low-B's can make no such claim about being able to help A's. A low-B who wishes to induce A to choose *In* must either lie or send an ambiguous message.

Harking back to the patterns of behavior that we delineated in section 3, LD-then-D behavior is much more prevalent than HR-then-D behavior, as there were eight instances of the former pattern (seven of these in the (7,7)-treatment) and no instances of the latter pattern. Similarly, HR-then-R behavior is much more common than LD-then-R, as we observed 19 instances of the former pattern (eight of these with low-B's) and there were no examples of the latter pattern. This supports our argument that, all else equal, being exposed as a liar is a strong deterrent to HR-then D and LD-then-R behavior; exposure does seem to matter to people.

## 6. DISCUSSION

Samuel Goldwyn quipped “an oral contract isn't worth the paper it is written on.” Contract theorists mainly agree, if not explicitly in writing, at least in the spirit of their work. Their basic models typically possess a unique equilibrium, which can't be upset by the addition of communication.<sup>20</sup> Yet, the human side of contracting seems a bit less dismal.

---

<sup>20</sup> We refer only to *one-shot* games where under traditional assumptions the backward-induction solution is unique. We do not consider repeated games in which communication may serve as an equilibrium-selection device. Some contract-theoretic models, in particular those based on repeated games such as MacLeod & Malcolmson (1989) have multiple equilibria, in which one may think that communication can help select say a Pareto-dominant equilibrium.

We incorporate guilt feelings to a contract-theoretic model with hidden information, using B&D's theory of guilt-from-blame. We show that guilt may foster trust & cooperation, and argue that communication may strengthen these effects further given that a few complementary assumptions are satisfied: language has focal meaning and people do not like to be caught lying. We predicted specific patterns regarding what confidence tricksters say (they lie and claim to be better than they are and then blame bad luck, rather than reveal to have low talent and then manifest opportunistic behavior) and about what trustworthy people say (they truthfully reveal their level of talent and then do as well as they can).

We examined these predictions in an experiment and found that the predictions were largely supported in a setting where there is an opportunity for a Pareto-improvement in payoffs for everyone (i.e., in the (7,7)-treatment). Beyond supporting theory, these results entail some 'useful lessons' which on extrapolation may offer useful guidance for confidence tricksters who want to fool others and for whom feelings of guilt is not a great worry as well as for people trying to tell if someone else is being honest. When someone claims to have high talent, this is often 'the big lie', which should be viewed with some suspicion. However, the claim that someone has low talent but will do his best turns out to be completely reliable, and is in fact almost always believed by the principal.

In the alternative treatment, where no Pareto improvement is possible for most of the agents (i.e., in the (5,5)-treatment), we see a dramatic breakdown of trust & cooperation relative to the (7,7)-treatment. Even with communication, very few (3 out of 16) low agents get-out-of-the-way (by choosing *Don't*). A reduction of trust & cooperation is in line with the predictions we derived in the theory section, but it is perhaps still surprising that the effects are so extreme. The breakdown of cooperation between the (7,7)- and (5,5)-treatments is mainly due to the low

willingness of low-B's to step aside in the (5,5)-treatment relative to the higher willingness of low-B's to accept a lower payoff than high-B's in the (7,7)-treatment.<sup>21</sup> <sup>22</sup> It seems that either the sense of disadvantage or the necessity of voluntarily stepping aside is enough to overcome whatever sense of guilt might be present for a low-B choosing *Roll* and thereby hurting the trusting A.

Perhaps this principle applies in the field as well, so that low-talent people in the real world will also manifest this sort of behavior, figuring that life has not given them an even break, so why should they cooperate and help out the principal? It appears to be the case that people are substantially more prone to be cooperative when they have a voice and an action that yields improvements in material payoffs for all parties involved than when the only way to gain is at the expense of others.

## REFERENCES

- Akerlof, George (1970), "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism," *Quarterly Journal of Economics*, **84**, 488–500.
- Battigalli, Pierpaolo & Martin Dufwenberg (2007), "Guilt in Games", *American Economic Review Papers and Proceedings*, **97**, 170-176.
- Battigalli, Pierpaolo & Martin Dufwenberg (2008), "Dynamic Psychological Games", forthcoming, *Journal of Economic Theory*.
- Baumeister, Roy, Arlene Stillwell & Todd Heatherton (1994), "Guilt: An Interpersonal Approach", *Psychological Bulletin*, **115**, 243-67.
- Baumeister, Roy, Arlene Stillwell & Todd Heatherton (1995), "Personal Narratives about Guilt: Role in Action Control and Interpersonal Relationships", *Basic & Applied Social Psychology*, **17**, 173-98.

---

<sup>21</sup> The difference does not derive from an increased tendency of low-B's to pretend that they have high talent. Roughly the same proportion of low-talent agents in the (5,5) and (7,7) treatments with communication lie and claim to have high talent.

<sup>22</sup> As we noted earlier, another difference is that principals are significantly less likely to trust HR messages in the (5,5)-treatment.

- Blume, Andreas & Andreas Ortmann (2007), “The Effect of Costless Pre-play Communication: Experimental Evidence for Games with Pareto-ranked Equilibria.” *Journal of Economic Theory* **132**, 274–90.
- Bolton, Patrick & Mathias Dewatripont (2005), *Contract Theory*, MIT Press.
- Bolton, Gary & Axel Ockenfels (2000), “ERC: A Theory of Equity, Reciprocity, and Competition”, *American Economic Review*, **90**, 166-93.
- Broberg, Tomas, Tore Ellingsen & Magnus Johannesson (2004), “Is Generosity Involuntary?”, *Economics Letters*, **94**, 32-37.
- Charness, Gary & Martin Dufwenberg (2006), “Promises and Partnership,” *Econometrica*, **74**, 1579-1601.
- Charness, Gary & Martin Dufwenberg (2008), “Broken Promises: An Experiment”, mimeo.
- Charness, Gary & Matthew Rabin (2002), “Understanding Social Preferences with Simple Tests”, *Quarterly Journal of Economics*, **117**, 817-69.
- Chen, Ying, Navin Kartik & Joel Sobel (2007), “Selecting Cheap-Talk Equilibria,” forthcoming in *Econometrica*.
- Crawford, Vincent (2003), “Lying for Strategic Advantage: Rational and Boundedly Rational Misrepresentation of Intentions”, *American Economic Review*, **93**, 133-149.
- Dana, Jason, Daylian Cain & Robyn Dawes (2006), “What You Don’t Know Won’t Hurt Me: Costly (but Quiet) Exit in Dictator Games”, *Organizational Behavior and Human Decision Processes*, **100**, 193-201.
- Dana, Jason, Roberto Weber & Jason Xi Kuang (2007), “Exploiting Moral Wiggle Room: Experiments Demonstrating and Illusory Preference for Fairness”, *Economic Theory*, **33**, 67-80.
- Demichelis, Stefano & Jörgen Weibull (2008), “Language, Meaning, and Games: A Model of Communication, Coordination, and Evolution”, *American Economic Review*, **98**, 1292–1311.
- Ellingsen, Tore & Magnus Johannesson (2004) “Promises, Threats, and Fairness”, *Economic Journal*, **114**, 397-420.
- Farrell, Joseph (1993), “Meaning and Credibility in Cheap-Talk Games.” *Games and Economic Behavior*, **5**, 514 –31.
- Farrell, Joseph & Matthew Rabin (1996), “Cheap Talk,” *Journal of Economic Perspectives*, **79**, 1214 –23.
- Fehr, Ernst & Klaus Schmidt (1999), “A Theory of Fairness, Competition, and Cooperation”, *Quarterly Journal of Economics*, **114**, 817-68.

- 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 (2005), “Deception: The Role of Consequences”, *American Economic Review*, **95**, 384-394.
- Hurkens, Sjaak & Navin Kartik (2007), “Would I Lie to You? On Social Preferences and Lying Aversion”, forthcoming in *Experimental Economics*.
- Kartik, Navin (2008), “Strategic Communication with Lying Costs,” mimeo.
- MacLeod, W. Bentley & James Malcolmson (1989), “Implicit Contracts, Incentive Compatibility, and Involuntary Unemployment”, *Econometrica*, **57**, 447-80.
- Nash, John (1950), “The Bargaining Problem”, *Econometrica*, **18**, 155-62.
- Rabin, Matthew (1990), “Communication Between Rational Agents,” *Journal of Economic Theory*, **51**, 144-70.
- Tadelis, Steven (2007), “The Power of Shame and the Rationality of Trust”, mimeo.
- Tangney, June (1995), “Recent Advances in the Empirical-Study of Shame and Guilt,” *American Behavioral Science*, **38**, 1132-45.
- Vanberg, Christoph (2008), “Why Do People Keep Their Promises? An Experimental Test of Two Explanations”, forthcoming in *Econometrica*.

## Appendix A: Guilt-from-blame derivations

*B&D develop the theory of guilt-from-blame for a large class of game forms. For a full account we refer to the original source. Here we apply B&D's sequential equilibrium (SE) concept to our games. We start with the (7,7)-game of Figure 1:*

- We assume away bad-conscience issues on behalf of a high-B player; she simply chooses *Roll*, thereby maximizing the (expected) payoff of A as well as of B.<sup>23</sup>
- For simplicity we also abstract away from bad-conscience issues on behalf of player A; he simply chooses what maximizes his (expected) monetary payoff.
- Note that the expected payoff to A if a low-B chooses *Roll* with probability  $x$  (and a high-B chooses *Roll*) equals  $\frac{2}{3} \cdot ([1-x] \cdot 7 + x \cdot 0) + \frac{1}{3} \cdot (\frac{5}{6} \cdot 12 + \frac{1}{6} \cdot 0) = 8 - \frac{14}{3} \cdot x$ .
- The guilt-from-blame of a low-B choosing *Roll* is proportional to the minimum of  $8 - \frac{14}{3} \cdot x$  (which in SE is all A expects when choosing *In*) and 7 (since that is all a low-B actually controls when he moves).
- The guilt-from-blame of a low-B choosing *Roll* is moreover proportional to the probability A assigns to B having low talent conditional on getting a \$0 payoff. Let  $\lambda$  denote this probability. Note that if a low-B chooses *Roll* with probability  $x$  (and high-B chooses *Roll*), Bayesian updating implies that  $\lambda = \frac{\frac{2}{3}x}{\frac{2}{3}x + \frac{1}{3} \cdot \frac{1}{6}} = \frac{12x}{12x + 1}$ .
- The guilt-from-blame of a low-B choosing *Roll* is finally also proportional to an exogenously given parameter  $\theta \geq 0$ , which reflects B's sensitivity to guilt-from-blame. If  $\theta = 0$ , then B is a selfish own-money maximizer; else, she has a conscience.
- Combining the past three bullets, if low-B chooses *Roll* he suffers from guilt-from blame in proportion to  $\theta \cdot \lambda \cdot \min\{8 - \frac{14}{3} \cdot x, 7\}$ . His utility if he chooses *Roll* equals  $10 - \theta \cdot \lambda \cdot \min\{8 - \frac{14}{3} \cdot x, 7\}$ . His utility if he chooses *Don't* equals 7. Other utilities throughout our game equal dollar payoffs, for all players and at all other end nodes.

---

<sup>23</sup> Theoretically, if a high-B were *extremely* sensitive to guilt (and if  $\lambda$  were high enough) then he might choose *Don't*, preferring the (7,7) outcome to the gamble following *Roll* that would render him a pang of guilt-from-blame with probability 1/6. In retrospect we see that we are justified in ruling out this case as we observed no instances of a high-B choosing *Don't*.

- What was said so far regarding guilt-from-blame is true only in SE; absent the equilibrium restriction we would have to work harder and keep explicit track of the players' higher-order beliefs.
- We next give SE predictions: A SE is a strategy profile such that each player maximizes his utility at his information set, along with beliefs for each information set and player that are consistent with Bayes' rule. For a more formal definition, see B&D. Here we just focus on what SE implies for our game, and the supporting intuition:
- If  $\theta \leq \frac{39}{40}$ , strategy profile  $(Out, (Roll, Roll))$  – meaning A chooses *Out*, low-B chooses *Roll*, high-B chooses *Roll* – is a SE. Proof: Low-B chooses *Roll* implies  $x = 1$ . Therefore  $\lambda = \frac{12}{13}$  and  $8 - \frac{14}{3} \cdot x = \frac{10}{3}$  and  $10 - \theta \cdot \lambda \cdot \min\{8 - \frac{14}{3} \cdot x, 7\} = 10 - \theta \cdot \frac{12}{13} \cdot \frac{10}{3}$ . Choosing *Roll* is optimal for the low B if  $10 - \theta \cdot \frac{12}{13} \cdot \frac{10}{3} \geq 7$ , which implies  $\theta \leq \frac{39}{40}$ .
- Following up on the previous bullet,  $(Out, (Roll, Roll))$  cannot be an SE if  $\theta > \frac{39}{40}$ , because the inequality  $10 - \theta \cdot \frac{12}{13} \cdot \frac{10}{3} \geq 7$  would not hold.
- Regardless of  $\theta$ , the strategy profile  $(In, (Don't, Roll))$  is not a SE. Proof: Low-B chooses *Don't* implies  $x = 0$ , so  $\lambda = 0$ , so by deviating to *Roll* low-B would get utility  $10 - \theta \cdot \lambda \cdot \min\{8 - \frac{14}{3} \cdot x, 7\} = 10 - \theta \cdot 0 \cdot 8 = 10$ . Since  $10 > 7$ , a low-B would thus gain.
- If  $\theta \geq \frac{25}{42}$ , the strategy profile  $(In, (x, Roll))$  is a SE where  $x = \frac{1}{28\theta - 12} \leq \frac{3}{14}$ . Proof: If  $x \leq \frac{3}{14}$ , then  $\min\{8 - \frac{14}{3} \cdot x, 7\} = 7$ . A low-B is willing to randomize only if her utility from choices *Don't* and *Roll* is equal, so we have:

$$7 = 10 - \theta \cdot \lambda \cdot \min\{8 - \frac{14}{3} \cdot x, 7\} = 10 - \theta \cdot \frac{12x}{12x+1} \cdot 7.$$

Solving this inequality for different combinations of  $\theta$  and  $x$  shows that as  $\theta$  ranges from  $\frac{25}{42}$  to infinity,  $x$  ranges from  $\frac{3}{14}$  down to 0.<sup>24</sup>

---

<sup>24</sup> There are also equilibria such that  $\frac{3}{14} < x < 1$ . In these cases,  $\min\{8 - \frac{14}{3} \cdot x, 7\} = 8 - \frac{14}{3} \cdot x$ . The dependence on  $x$  means that solving for SE requires solving the quadratic equation  $7 = 10 - \theta \cdot \lambda \cdot (8 - \frac{14}{3} \cdot x)$ ,

- There is SE multiplicity if  $\frac{25}{42} \leq \theta \leq \frac{39}{40}$ . Proof: Compare the bullets one and three steps up. This shows that incorporating guilt-from-blame may or may not alleviate the problems of adverse selection, as for a range of intermediate values of  $\theta$  there are multiple SE that can be Pareto ranked in terms of the players' (expected) utilities.

We next move to the (5,5)- game of Figure 2:

- Again, we assume away bad-conscience issues on behalf of a high-B player as well as player A.
- In SE, a low-B's utility equals 5 if he chooses *Don't* and it equals  $10 - \theta \cdot \lambda \cdot \min\{\frac{20}{3} - \frac{10}{3} \cdot x, 5\}$  if he chooses *Roll*. The guilt-from-blame of a low-B choosing *Roll* is proportional to the minimum of  $\frac{20}{3} - \frac{10}{3} \cdot x$  (which in SE is all A expects when choosing *In*) and 5 (which is all a low-B actually controls when he moves).
- If  $\theta \leq \frac{13}{8}$ , strategy profile (*Out*, (*Roll*,*Roll*)) – meaning A chooses *Out*, low-B chooses *Roll*, high-B chooses *Roll* – is a SE. Proof: Low-B chooses *Roll* implies  $x = 1$ . Therefore  $\lambda = \frac{12}{13}$  and  $\frac{20}{3} - \frac{10}{3} \cdot x = \frac{10}{3}$  and  $10 - \theta \cdot \lambda \cdot \min\{\frac{20}{3} - \frac{10}{3} \cdot x, 5\} = 10 - \theta \cdot \frac{12}{13} \cdot \frac{10}{3}$ . Choosing *Roll* is optimal for the low B if  $10 - \theta \cdot \frac{12}{13} \cdot \frac{10}{3} \geq 5$ , which implies  $\theta \leq \frac{13}{8}$ .
- Following up on the previous bullet, (*Out*, (*Roll*,*Roll*)) cannot be an SE if  $\theta > \frac{13}{8}$ , because the inequality  $10 - \theta \cdot \frac{12}{13} \cdot \frac{10}{3} \geq 5$  would not hold.
- Regardless of  $\theta$ , the strategy profile (*In*, (*Don't*,*Roll*)) is not a SE. Proof: Low-B chooses *Don't* implies  $x = 0$ , so  $\lambda = 0$ , so  $10 - \theta \cdot \lambda \cdot \min\{\frac{20}{3} - \frac{10}{3} \cdot x, 5\} = 10 - \theta \cdot 0 \cdot \frac{20}{3} = 10$ . Since  $10 > 5$ , a low-B would optimize by deviating to choose *Roll*.

---

which describes the relevant indifference condition for low-B. Some manipulations show that relevant roots satisfy  $(96\theta - 36) \pm ((96\theta - 36)^2 - \frac{3}{56\theta})^{\frac{1}{2}}$  as well as  $\frac{3}{14} < x < 1$ .

- If  $\theta \geq \frac{7}{6}$ , the strategy profile  $(In, (x, Roll))$  is a SE where  $x = \frac{1}{12\theta - 12} \leq \frac{1}{2}$ . Proof: If  $x \leq \frac{1}{2}$ , then  $\min\{\frac{20}{3} - \frac{10}{3} \cdot x, 5\} = 5$ . A low-B is willing to randomize only if her utility from choices *Don't* and *Roll* is equal:

$$5 = 10 - \theta \cdot \lambda \cdot \min\{\frac{20}{3} - \frac{10}{3} \cdot x, 5\} = 10 - \theta \cdot \frac{12x}{12x + 1} \cdot 5.$$

Solving this inequality for different combinations of  $\theta$  and  $x$  shows that as  $\theta$  ranges from  $\frac{7}{6}$  to  $\infty$ ,  $x$  ranges from  $\frac{1}{2}$  down to 0.<sup>25</sup>

- There is SE multiplicity if  $\frac{7}{6} \leq \theta \leq \frac{13}{8}$ . Proof: Compare the bullets one & three steps up.

---

<sup>25</sup> There are also equilibria such that  $\frac{1}{2} < x < 1$ . In these cases,  $\min\{\frac{20}{3} - \frac{10}{3} \cdot x, 5\} = \frac{20}{3} - \frac{10}{3} \cdot x$  and the dependence on  $x$  means that solving for equilibria requires dealing with the quadratic equation  $5 = 10 - \theta \cdot \lambda \cdot (\frac{20}{3} - \frac{10}{3} \cdot x)$ , which describes the relevant indifference condition for a low-B. Some manipulations

show that relevant roots satisfy  $\frac{(48\theta - 84) \pm ((48\theta - 84)^2 - 672\theta)^{\frac{1}{2}}}{48\theta}$  as well as  $\frac{1}{2} < x < 1$ .

## Appendix B: Messages

### The (7,7)-game

In the tables below: R = *Roll*, D = *Don't*.

We classify messages as HR = Promise High & *Roll*, IHR = Implicit promise High & *Roll*, LD = Promise Low, then *Don't*, ILD = Implicit promise Low, then *Don't*, PH = Promise High, PL = Promise Low, PNS = Promise not selfish, PD = Promise *Don't*, E = Empty Talk

Session	ID	Message	Class	Type	A	B
1	1	I am a B that is Low.	PL	L	OUT	-
1	2	I won't be selfish—Choose In.	PNS	L	IN	D
1	3	I hope you had a good summer so far. Santa Barbara's a neat place to be...I'm here with my wonderful boyfriend...I feel very blessed & content. I hope you are feeling great too. Bye!	E	H	IN	R
1	4	I won't roll.	PD	L	IN	D
1	5	Trust me. I am a nice person. ☺	PNS	L	IN	D
1	6	You have a 5/6 chance of getting \$12. Wanna go for it? <u>Live dangerously.</u> ☺ Either way, free cash. Enjoy.	IHR	H	IN	R
1	7	Hello A, you should choose IN. After doing this I will agree to select DON'T ROLL considering I am a LOW B. The outcome will be that we both receive \$7 each. I am a very altruistic person and would receive more gratification from both of us receiving \$7 each than me only being the one to receive \$10. You have my word and we will both leave happy!	LD	L	IN	D
1	8	Choose IN, I want you to get money too! Everybody wins when A chooses in ☺	E	L	IN	D
1	9	I'll take care of you, so don't worry. I am not a "Pursue wealth forgetting all but self" type of person.	PNS	H	IN	R
1	10	I am a <u>Low Type</u> . Please choose IN because I <u>will not</u> choose ROLL. I will not use you for my benefit.	LD	L	IN	D
1	11	I am <u>LOW TYPE</u> . I am going to choose "Don't roll" because I want <u>both</u> of us to get \$7. (I would feel bad if I got \$10 & you got nothing b/c we all came here to get something). Have a great weekend, whoever you are.	LD	L	IN	D
1	12	-	-	H	OUT	-
1	13	High type.	PH	L	IN	D
1	14	I am a high number so choosing IN would work out best for both of us so I hope you choose it. I plan to roll. I think it is worth the risk.	HR	L	IN	R
1	15	5/6 CHANCE FOR \$12 FOR YOU IF YOU CHOOSE IN...	IHR	H	OUT	-
1	16	If you choose in we will both win.	E	L	IN	D
2	1	I need TP for my Bunghole! Let's choose IN and don't roll so we both get \$7.	PD	L	IN	D
2	2	I will choose don't roll.	PD	L	IN	D

2	3	Choose IN, I'm high. Big chance of winning high amount.	IHR	H	IN	R
2	4	You will not get \$0. I promise. Do not choose out.	PD	L	IN	D
2	5	Want to win money, attend a psych test. Huzzah! I don't know what to write you, but I suppose you just need to trust I'm going to act in your best interests... In the meantime! There are two sausages in a frying pan. One of them says "It's getting hot in here." The other says "Ahh! Tolling sausage!"	E	L	IN	D
2	6	Hi, anyway I think you should know that I am a high type because we can both make more money if you choose in.	PH	H	IN	R
2	7	-	-	L	IN	D
2	8	If you choose "In" I will choose "don't Roll" so we both get \$7.	LD	L	IN	D
2	9	I AM A HIGH TYPE AND I'M GOING TO CHOOSE ROLL IF YOU CHOOSE IN, BECAUSE 17 IS THE BEST CHANCE OF DOING VERY WELL.	PH	H	IN	R
2	10	Best of luck on this game. I hope we both win lots of \$, so please choose <u>IN</u> !	E	L	OUT	-
2	11	I am a <u>LOW B</u> . If you choose <u>IN</u> , and I choose don't roll, the 7 bucks is better than the 5...win/win.	LD	L	IN	D
2	12	I'm happy with my current situation lucky today. You can make your own decision.	E	H	OUT	-
2	13	♪ Imagine no possessions, no religion, too.  Imagine all the people living life in peace...  You may say I'm a dreamer but I'm not the only one  I hope someday you will join us And the world will live as one.	E	L	OUT	-
3	1	I'm low. But do select IN. I will opt <u>not to Roll</u> so we can get \$& each. Trust me ☺ .	LD	L	IN	D
3	2	I'm a low type, and am planning on choosing not to roll so if you choose in, that was we both get a good amount of \$.	LD	L	OUT	-
3	3	No risk, no money. You choose in and I can give you 5 out of 6 odds that your wallet will be fatter.	IHR	H	IN	R

3	4	Hi! I am lucky to have got HIGH type B; which means we can get higher pay-offs! It only makes max sense if you choose IN so that we get more than \$5 each. Also you may tell me if you want me to roll or not by underlining your choice. <u>IN</u> = ROLL IN = DON'T ROLL Doing this just to involve you in the complete decision process. Good luck!	PH	L	IN	R
3	5	Hi, I am a HIGH B type. So I think we can earn a \$7 each. You choose IN & I'll choose DON'T ROLL. Instead of choosing to roll, I think not choosing it is better as there is no risk of you getting a \$0 at all. I loose \$3 but I guess it's better than you loosing \$7. Good luck!	HR	L	IN	R
3	6	Hi! I'm a high B, so probabilistically, it's to your advantage to choose IN, regardless of whether I choose roll or not. Of course, it wouldn't be if I was a low B, but I'm not. Pick IN and, chances are, you'll be pleasantly surprised. Have a nice day!	HR	H	IN	R
3	7	Nice day for an experiment. ☺ Good luck.	E	L	OUT	-
3	8	I AM A LOW TYPE. I <u>GIVE YOU MY WORD</u> . THAT I WILL CHOOSE "DON'T ROLL" IF YOU CHOOSE "IN." IN THAT CASE, WE WILL EACH MAKE \$7.	LD	L	IN	D
3	9	It's probably a good idea to take a chance...I'm a high type, so you're almost guaranteed to make \$12.	HR	H	IN	R
3	10	I am of high type, so you have a good chance of \$12 if you choose IN.	HR	L	IN	R
3	11	I am a high type!! Let's roll - there's a great chance you'll get \$12 plus the \$5 show-up fee.	HR	L	IN	R
3	12	If you choose <u>In</u> you have a 5 in 6 chance of getting 17. Much better than 10.	IHR	H	IN	R

### The (5,5)-game

Session	ID	Message	Class	Type	A	B
1	1	Hope you take a good decision Good luck	E	L	OUT	-
1	2	I'm a low type, you may as well choose out. Even if you choose in, I won't roll. \$5 extra ... yay! ☺ If I were an A, I'd want someone to do the same for me.	LD	L	IN	D
1	3	Well, I think that it would be in your best interest to choose IN, cause I'm a high type of B, that's just how I roll, and I'm pretty sure that the roll of the dice will go your way too! Ok well, I hope it goes well.	HR	H	OUT	-
1	4	-	-	L	IN	R
1	5	Low Type	PL	L	OUT	

1	6	Sure I'm high type, odds are that you will get 12 dollars if you choose "in" and allow me to roll. ☺	HR	H	IN	R
1	7	Choose in, thanks	E	L	OUT	-
1	8	You're at pretty good odds to make some money. Take a risk, you know?	E	L	OUT	-
1	9	I am B <u>high</u> type. Your odds are good, 2 out of 3, do it, go in. Odds of die roll are 5/6, 5 out of 6 for a winner. Pretend your in Vegas!!	IHR	H	IN	R
1	10	I'm high, so pick IN, because that would give you a 5/6 chance of receiving 12 dollars, so that's an average of 10 dollars per person, instead of 5.	IHR	L	IN	R
1	11	I am a low type	PL	L	OUT	-
1	12	I'm high type, so choose IN so I can roll. You've got a 5/6 chance in scoring 12 bucks, pretty good odds.  P.S. I love you. If only you'd stop staring at me, maybe it could work.	HR	H	OUT	-
1	13	I'm a high type, and I'm planning to roll if you go in. I suggest you do, not only so I can get \$10, but also because you have a 5/6 chance of getting \$12 rather than \$5.	HR	L	IN	R
2	1	I'm a Low.	PL	L	OUT	-
2	2	Choose In, I'll hook you up. 5/6 chance we both win big.	IHR	L	OUT	-
2	3	Go in. I am high; 5/6 chance of a higher payoff.	IHR	H	IN	R
2	4	I've done these experiments before and it tends to pay off when you take risks.  I'm high type & it's unlikely for you to roll a 1.	IHR	L	IN	R
2	5	I'm a high type. 1 n 6 chance of both of us getting a good payoff.	IHR	L	OUT	-
2	6	I am a high type. So please choose <u>in</u> ; we will have a higher chance to win more money. \$17 for me and \$15 for you id I roll anything other than 1 ☺ Even if I roll a 1 you will still get <del>\$10</del> \$15.	HR	H	OUT	-
2	7	Choose 'IN'? I don't know what to write .....	E	L	OUT	-
2	8	Go in.	E	L	IN	R
2	9	-	-	H	IN	R
2	10	-	-	L	IN	R
2	11	-	-	L	OUT	-
2	12	A,  Choose IN, because I am a high. I will role. We have a 5/6 chance to make some serious \$\$\$\$. You can make \$17 minimum. And I heard he actually gives \$2 extra, in addition to the \$5 entry money. That's \$19. Do it!  This is for you! & me! - B	HR	H	IN	R

2	13	-	-	L	IN	R
2	14	-	-	L	OUT	-
2	15	I'm a B high type – I'm rolling.	HR	H	OUT	-
2	16	-	-	L	IN	R
2	17	You should choose to stay in. ☺	E	L	IN	R
2	18	Hey. I'm a high type so most likely you're going to make some money. Lucky you!	IHR	H	OUT	-
3	1	You should go in to get more money.	E	L	OUT	-
3	2	Hey!  Choose <u>IN</u> !  If I am a low type I'll choose <u>Don't Roll</u> and we both get \$5.  <u>But</u> , if I am a High type I'll choose roll and you'll have a 5/6 chance to win \$12.	PNS	L	IN	R
3	3	Hey, Choose in because we will “both” make a lot of money, or at least we both have an extremely <u>high</u> probability of making a lot of money.	IHR	H	OUT	-
3	4	Choose 'In' you have a good chance of making more than me. It will benefit both of us.	IHR	L	IN	R
3	5	-	-	L	IN	D
3	6	You should choose IN and since I am HIGH I will roll and the likelihood of me getting a 2-6 is 5/6 (about (84%) chance that you'll get 12\$ and I will get 10\$	HR	H	IN	R
3	7	We will likely both leave very happy as long as you are in.	E	L	OUT	-
3	8	-	-	L	IN	R
3	9	-	-	H	IN	R
3	10	☺ g'day!	E	L	IN	D
3	11	-	-	L	OUT	-
3	12	I encourage you to choose “IN”, it is more likely to be financially rewarding, and just makes life more interesting.  Good luck to us!	E	H	IN	R
3	13	-	-	L	OUT	-
3	14	-	-	L	OUT	-
3	15	I think you should choose IN. It can give us a better opportunity for more money.	E	H	OUT	-
3	16	Hi,  I don't really like the idea of not knowing who you are or not being able to tell you my name, but we can't do anything about it.	E	L	IN	R

## Appendix C: Sample Instructions

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, as a show-up fee 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. There are 2 types for B; call these HIGH and LOW. Each B participant will draw a number from a bag to determine his or her type. Each B who draws a number that is a multiple of three (for example: 3, 6, 9, etc.) will be a HIGH type; all other B's are LOW types. Thus, there are about twice as many LOW types as HIGH types. Information about B's type is not conveyed to A.

On the designated decision sheet, each person A will indicate whether he or she wishes to choose IN or OUT. If A chooses OUT, each of A and B receives \$5 (in addition to the show-up fee).

We will collect these sheets after the choices have been indicated. We will then convey to each B the choice made by the A with whom he or she is paired. If A chose OUT, B has no choice to make. If A has chosen IN, B will indicate whether he or she wishes to ROLL.

If A chooses IN and B chooses DON'T ROLL, A receives \$5 and B receives \$5. If A chooses IN and B chooses ROLL, the result depends on B's type. If B is the LOW type and chooses ROLL, then A receives \$0 and B receives \$10. If B is the HIGH type and chooses ROLL, then B receives \$10 and the outcome of the roll of a 6-sided die determines 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	<b>\$5</b>	<b>\$5</b>
A chooses IN and:		
B is LOW type and chooses DON'T ROLL	<b>\$5</b>	<b>\$5</b>
B is LOW type and chooses ROLL	<b>\$0</b>	<b>\$10</b>
B is HIGH type and chooses DON'T ROLL	<b>\$5</b>	<b>\$5</b>
B is HIGH type, chooses ROLL, die=1	<b>\$0</b>	<b>\$10</b>
B is HIGH type, chooses ROLL, die= 2,3,4,5, or 6	<b>\$12</b>	<b>\$10</b>

### *A Message*

Prior to the decisions by A and B, 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; these will be collected when people are ready. Please print clearly if you wish to send a message to A. We will convey this message to the appropriate A participant (without the identifying number) and then A and B will proceed as described above.

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.