Participation

Gary Charness & Martin Dufwenberg*

October 22, 2009

ABSTRACT: We show experimentally that whether and how communication achieves beneficial social outcomes in a hidden-information context depends crucially on whether low-talent agents can participate in a Pareto-improving outcome. Communication is effective (and patterns of lies & truths quite systematic) when this is feasible, but otherwise completely ineffective. We examine the data in the light of two potentially relevant behavioral models: cost-of-lying and guilt-from-blame.

KEYWORDS: Adverse selection, hidden information, participation, communication, cost-of-lying, guilt-from-blame, psychological games, trust, cooperation, lies, deception, social preferences, behavioral economics

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

ACKNOWLEDGMENTS: We would like to thank three referees for their very helpful comments, Geir Asheim, Pierpaolo Battigalli, Stefano DellaVigna, Tore Ellingsen, Shachar Kariv, Ulrike Malmendier, and Matthew Rabin for useful discussions, and the 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 Amsterdam Behavioral & Experimental Economics Workshop, the ESA conferences in Lyon and Innsbruck, and seminar participants at the University of Gothenburg, Brown University, the University of California at Berkeley, and Columbia Business School for more 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).

* CONTACT: Gary Charness, Department of Economics, University of California at Santa Barbara, charness@econ.ucsb.edu; Martin Dufwenberg, Department of Economics, University of Arizona and Department of Economics, University of Gothenburg, 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.[1]

In this paper, we investigate an environment with *hidden information*. Here, while the agent's effort choice is observable and contractible, his production also depends on his ability.[2] A crucial feature is that while the agent knows his ability, the principal does not. Our approach complements that of Charness & Dufwenberg (2006), who consider a hidden-action context. However, the games differ 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.

We consider the interaction of two important issues in our experimental design. The first issue is the extent to which an agent with low-talent can *participate* in an outcome that is a Pareto-improvement for both the principal and the agent. In one environment, there are two possible types of employment available, with more paid for the job requiring high talent; if the

1 For an entry to the literature, see Bolton & Dewatripont (2005). The gloomy outlook can be exemplified with reference to 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. The "hidden" terminology seems more descriptive and less suggestive of the nature of outcomes.

2 In this paper, we shall consider the principal to be female and the agent to be male.

low-ability agent chooses the position not requiring high talent, both the agent and the principal are better off than if the principal chooses not to offer him employment. In the second environment, there is no low-skill position available. In both environments, a principal does poorly if matched with a low-talent agent who chooses the position requiring high ability. In both cases, principals must rely on low-talent agents to voluntarily accept less than could be obtained by acting selfishly and choosing the better-paid position, but in the second case low-talent agents who wish to avoid hurting the principal must step aside and decline the contract.

The second issue is whether *communication* can help to ameliorate the hidden-information problem. If agents have selfish preferences, the prediction in both environments is the same: A low-talent agent will choose the high-skill position and receive more income. Since a vast number of papers have shown that many people have social preferences, we would expect that not all low-talent agents make the selfish choice. But it also may well be the case that some aspect of communication will help to promote trust & cooperation. However, given the qualitative difference in the environments, the character and content of the messages sent are likely to also differ from those in the hidden-action environment.

We find that communication can be effective with hidden information, although this depends critically on low-talent agents having the possibility to participate in a Pareto-improving outcome. We proceed to discuss this result in the light of two behavioral models that can potentially explain such an effect and that have received some support in recent experimental research. One such model involves a cost-of-lying,[3] while the other is Battigalli & Dufwenberg's (2007) model of guilt-from-blame, which has its intellectual home within the

[3] Previous theoretical work considering various forms of cost-of-lying includes Ellingsen & Johannesson (2004), Chen, Kartik & Sobel (2007), Demichelis & Weibull (2008), and Kartik (2008). For some related experimental results see Gneezy (2005), Miettinen (2008), Hurkens & Kartik (2009), Sutter (2009), Vanberg (2008), Charness & Dufwenberg (2008).

framework of psychological game theory (Geanakoplos, Pearce & Stacchetti 1989; Battigalli & Dufwenberg 2009). We present formal predictions for each model in our environments and discuss the extent that the models can encapsulate the observed patterns of behavior.

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 lie-detectors who wish to build better traps.

The remainder of the paper is organized as follows. Our hidden-information games are presented in section 2. The experiment design is described in section 3, and the experimental results are presented in section 4. The two behavioral models are presented in section 5, and section 6 offers concluding remarks.

2. HIDDEN-INFORMATION GAMES

In this section we describe the games that we use in our treatments. The game (form) in Figure 1 models our benchmark scenario (which for reasons explained further below we shall call our (5,7)-game). A principal (player A) considers employing an agent (B) to form a partnership in which a project is carried out. If A passes on this option – an outcome corresponding to A's choice *Out* – then no contract is signed, no project is carried out, and the parties get their outside-option payoffs of 5 (dollars) each. The project is carried out if A chooses *In*, in which case A pays a fixed wage to B and then acts as residual claimant.

<<<Figure 1 about here>>>

Note that there is hidden information, since only the agent knows his own productivity (or talent). If B has low talent – which happens with probability 2/3 as indicated by the initial chance move – then he is only capable of performing a simple task such that if A pays B an appropriate low wage they split the gain and get 7 each. On the other hand, if B has high talent he could take on a more difficult and (in expectation) profitable task at which a low-talent agent would fail. Since only B knows his talent, only he can tell what is the best mutually beneficial contract, and the game in Figure 1 incorporates an opportunity for him to select it: choice *Don't* represents the low-wage simple task and choice *Roll* the high-wage difficult task.[4]

If a high-talent B chooses *Roll* then the outcome is potentially rewarding but risky: with probability 1/6 the project still fails (as it would for sure if low-talent B chose *Roll*). The chance move following path (*High*, *In*, *Roll*) captures this. The dotted line connecting A's payoffs of \$0, following paths (*Low*, *In*, *Roll*) and (*High*, *In*, *Roll*, *Failure*) indicates an information set for A across end nodes.[5] This reflects how A is never told how her payoff of \$0 came about.[6]

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 couldn't even in principle 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

[4] The labeling of players and strategies in Figure 1, which may appear somewhat artificial in light of the principal-agent story, anticipates the upcoming wording of our experimental instructions as described below.

[5] 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 5 will show). See Battigalli & Dufwenberg (2009, section 6.2) for more discussion of this point.

[6] In principle, there should also be dotted lines connecting A's payoffs of \$5 as well as A's payoffs of \$7, but these are omitted for expositional clarity.

principal, or at least not verifiable in court. The chance move justifies a story where a low-type agent could falsely claim that he was in fact a high-type agent but that he had bad luck. Because of the chance move, it cannot be proven in court that he lied.

If the players are selfish and risk-neutral, the (5,7)-game of Figure 1 has a unique sequential equilibrium (henceforth, SE) as defined by Kreps & Wilson (1982): two steps of a backward induction argument yields that B chooses *Roll* independently of his talent, and A's best response is *Out* (this gives A a payoff of 5 whereas *In* would give A an expected payoff of $(1/3) \cdot [(5/6) \cdot 12 + (1/6) \cdot 0] + (2/3) \cdot 0 = 10/3$). The players earn 5 each independently of B's talent. The outcome is inefficient, since A, a low-talent B, and a high-talent B would each receive more (in expectation) if A chose *In* and low-talent B chose *Don't* while high-talent B chose *Roll*.[7] This illustrates how hidden information may undermine efficient contracting.

We also consider a version of the game with an added communication opportunity; B can send a message to A just after chance has determined B's talent and just before A chooses *In* or *Out*. With standard preferences the prediction does not change relative to the no-communication game; words can't change the fact that B gets a higher dollar payoff from *Roll* than from *Don't*, and given this A chooses *Out*.

How should one react to these predictions? One possibility is to take the indicated problem at face value, and examine whether *other* contractual arrangements help overcome the problems. 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,

[7] A would get $8 = (1/3) \cdot [(5/6) \cdot 12 + (1/6) \cdot 0] + (2/3) \cdot 7$; low-talent B would get 7; high-talent B would get 10.

as we are skeptical of the traditional premise that parties are selfish. We stick with the game of Figure 1 with an open mind to whether or not the situation is problematic.

We now move to the important issue of participation. The game in Figure 1 allows a way for each of the two types of the agent to have mutually profitable (Pareto-improving) dealings with the principal. A high-talent agent who chooses *Roll* moves himself and the principal from a payoff of 5 to a payoff of 10 (in expected terms), while a low-talent agent who chooses *Don't* moves the payoff from 5 to 7. Everyone gains. But note how the gains-from-trade are asymmetric as regards different types of agents. One may imagine a more extreme form of such asymmetry, where the low-talent agent is simply incapable of participating in making net additions to partnership profit. Perhaps they lack any helpful trait, or perhaps government taxation is so high that all gains from trade get wasted, or 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. The game in Figure 2 incorporates such a change to the setting:

<<<Figure 2 about here>>>

We call the game in Figure 1 our (5,7)-game because 5 is the value of the outside option (*Out*) and 7 is the value of the low-wage simple-task outcome (path via *Don't*). Accordingly, the game in Figure 2 is our (5,5)-game. Parametrically, the change between games looks small: four 7's are replaced by four 5's. The interpretation of the *Don't* choice changes too, to reflect a "step-aside" move. The prediction for selfish players does not change though: A chooses *Out*, and B chooses *Roll* independently of talent. And again, adding communication (in the same way as described for the (5,7)-game) would not change this dismal prediction.

We find it intuitive that when behavioral concerns are considered it will somehow be easier to foster trust & cooperation in the (5,7)-game than in the (5,5)-game – asking low-talent agents to accept a lesser gain seems easier than asking them to step aside. We explore this. It turns out that for theory-testing purposes we need a third game, a variant of the step-aside scenario called the (7,7)-game. We defer a discussion of the rationale and here just present it:

<<<Figure 3 about here>>>

3. EXPERIMENTAL DESIGN

In line with the presentation in section 2, we have a 3×2 design. The first treatment variable concerns whether subjects played the (5,7)-game, the (5,5)-game, or the (7,7)-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. We provided each potential sender with a blank piece of paper on which he could write any (anonymous) message instead of restricting the message space.

Participants were recruited at UCSB by sending out an e-mail message to the campus community. We conducted 18 sessions, three for each of our six treatments. Sessions were conducted in a large classroom that was divided into two sides by a center aisle, and people were seated at spaced intervals. The number of participants in a session ranged from 20 to 36, for a total of 510 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’ or ‘B’. A coin was tossed to determine which side of the room was A and which side was B. Index cards with identification numbers were drawn from an opaque bag, and participants were informed that these numbers would be used to determine pairings (one A with one B) and to track decisions. Each B first

learned his type, which was determined by his private draw. If his identification number was evenly divisible by three, B had high talent; otherwise he had low talent. Sample instructions are given in Appendix A.

In all of our treatments, we presented a Table to each of the participants, indicating the outcome for every combination of choices and die rolls. 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 treatments, B had an option to send a free-form message to A prior to A's decision. B could also decline to send a message by circling the letter B at the top of the otherwise-blank sheet. Then A chose *In* or *Out*. Finally, B learned A's choice and, if A had chosen *In*, chose *Roll* or *Don't*.

Table 1 shows the experimental presentation of the (5,7)-game; this is identical for the (5,5)- and (7,7)-games, except that each "7" in the fourth and seventh rows is replaced with "5" and "7", respectively, and each "5" in the first row is replaced by "7" in the (7,7)-game.

Table 1: Payoff Outcomes in (5,7) Game

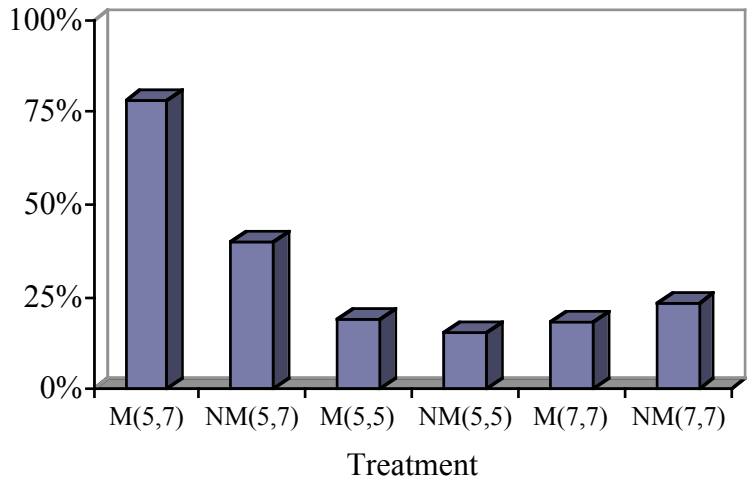
	A receives	**B receives**
A chooses OUT	**$5**	**$5**
A chooses IN and:		
B is LOW type and chooses DON'T ROLL	**$7**	**$7**
B is LOW type and chooses ROLL	**$0**	**$10**
B is HIGH type and chooses DON'T ROLL	**$7**	**$7**
B is HIGH type, chooses ROLL, die = 1	**$0**	**$10**
B is HIGH type, chooses ROLL, die = 2,3,4,5,6	**$12**	**$10**

4. EXPERIMENTAL RESULTS

4.1 Data summary

Even without communication, we find considerably less selfish behavior by low-talent B's when they can participate in a Pareto-improving outcome; A's are also more likely to choose *In* when this is the case. Interestingly, communication is totally ineffective when low-talent B's cannot participate, but has a dramatic effect on low-talent B choices (and leads to a modest increase in A's *In* rate) when participation is feasible.[8] Figures 4 and 5 present low-talent B *Don't* rates and A *In* rates by treatment ("NM" means no message and "M" means message) and Table 2 summarizes the effect of communication on behavior for the (5,7)-, (5,5)-, and (7,7)-games.

Figure 4 - Low-B's Don't Rate Across Treatments



[8] High-talent B choices are omitted, as they are invariably (63 of 63 times) *Roll* in our sessions.

Figure 5 - A's In rate Across Treatments

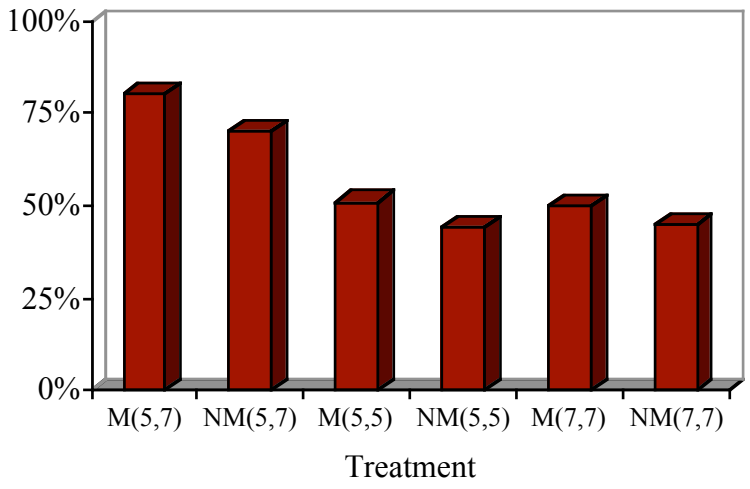


Table 2: Rates by Treatment and Tests for the Effect of Communication

	Low B's *Don't*			A's *In*		
Treatment	M	NM	*Z*-stat	M	NM	*Z*-stat
(5,7)	18/23 (78%)	8/20 (40%)	2.56***	33/41 (80%)	28/40 (70%)	1.09
(5,5)	3/16 (19%)	2/13 (15%)	0.24	24/47 (51%)	20/45 (44%)	0.64
(7,7)	2/11 (18%)	3/13 (23%)	-0.29	21/42 (50%)	18/40 (45%)	0.45

M/NM mean that no messages/messages were feasible. The *Z*-stat reflects the test of proportions across M and NM. *** indicates p < 0.01, one-tailed test.

Summarizing the results, the only case in which communication led to a significant increase was for low-talent B's in the (5,7) game, where the *Don't* rate nearly doubles, to 78%. Note that this rate is more than quadruple the *Don't* rates with communication in the two non-participation games, with statistical significance at $p < 0.001$ in each case.[9] The proportions of *Don't* are very close in the (5,5)- and (7,7)-games, whether or not there is communication. In

[9] Unless otherwise stated, the test used is the test of the difference of proportions (Glasnapp & Poggio 1985); all *p*-values reflect two-tailed tests, unless otherwise stated.

general, it seems that low-talent B's refuse to step aside when there is no available Pareto improvement over A's outside option, but are often willing to accept lower payoffs than high-talent B's when participation is feasible.

Communication only affects A's behavior to a modest and insignificant degree, resulting a slight increase in the *In* rate each of the three games. There is nevertheless a higher *In* rate in the (5,7)-game than in either of the other games both when communication is possible ($Z = 2.88$ and $Z = 2.91$, $p < 0.01$ in both cases) and when it is not ($Z = 2.37$ and $Z = 2.26$, $p < 0.025$ in both cases). The *Don't* rate in the (5,7)-game without communication is about twice as high as in the other games; however, the differences with respect to the other games is no more than marginally significant, perhaps due to the low number of observations. The test of proportions on the no-communication *Don't* rates gives $Z = 1.50$ and $Z = 1.10$, respectively, or $Z = 1.55$ for the pooled data from the (5,5)- and (7,7)-games; if we use one-tailed tests (which seem natural here), we get $p = 0.067$, $p = 0.136$, and $p = 0.061$ for these comparisons.

4.2 Message content

What messages were sent? Free-form messages can potentially be classified in a variety of ways. To simplify the analysis, we assume that B can make a statement regarding his type (*Low* or *High*) and his choice (*Don't* or *Roll*), or stay silent. This produces five possible communication choices LD, LR, HD, HR, and S, where the notation in the first four cases refers to messages "I'm *Low* and I'll choose *Don't*", etc., with S representing silence. Ninety-three percent of all messages (121 of 130) can be assigned to one of these categories; in the other messages B stated that he was a low-talent B without implying an action. There is no doubt

room for discussion in some cases regarding the classification; in any case, the precise messages are presented in Appendix B, where we also provide a richer classification scheme.

In Tables 3-5 below, we break down our results with communication according to the type of message sent and the actions that were observed thereafter. Notice that we never observe a LR or HD message.

Table 3: Messages and Outcomes in (5,7)-Treatment

		LD	LR	HD	HR	S	Other	Total
Low B	*Out*	1	0	0	0	3	1	5
	In, R	0	0	0	5	0	0	5
	In, DR	13	0	0	1	4	0	18
	Total	14	0	0	6	7	1	28
High B	*Out*	0	0	0	1	2	0	3
	In, R	2	0	0	8	0	0	10
	In, DR	0	0	0	0	0	0	0
	Total	2	0	0	9	2	0	13

LD = *Low* & *Don't*; LR = *Low* & *Roll*; HD = *High* & *Don't*, HR = *High* & *Roll*, S = Silence

Table 4: Messages and Outcomes in (5,5)-Treatment

		LD	LR	HD	HR	S	Other	Total
Low B	*Out*	0	0	0	2	11	3	16
	In, R	1	0	0	4	8	0	13
	In, DR	1	0	0	0	2	0	3
	Total	2	0	0	6	21	3	32
High B	*Out*	0	0	0	6	1	0	7
	In, R	0	0	0	5	3	0	8
	In, DR	0	0	0	0	0	0	0
	Total	0	0	0	11	4	0	15

LD = *Low* & *Don't*; LR = *Low* & *Roll*; HD = *High* & *Don't*, HR = *High* & *Roll*, S = Silence

Table 5: Messages and Outcomes in (7,7)-Treatment

		LD	LR	HD	HR	S	Other	Total
Low B	*Out*	1	0	0	3	10	4	18
	In, R	1	0	0	5	3	0	9
	In, DR	1	0	0	0	0	1	2
	Total	3	0	0	8	13	5	29
High B	*Out*	1	0	0	0	2	0	3
	In, R	0	0	0	6	4	0	10
	In, DR	0	0	0	0	0	0	0
	Total	1	0	0	6	6	0	13

LD = *Low* & *Don't*; LR = *Low* & *Roll*; HD = *High* & *Don't*, HR = *High* & *Roll*, S = Silence

First, consider the messages of the low-talent B's. In the (5,5)-game, two chose LD, while six chose HR, and 21 chose Silence. The distribution of messages was similar for the low-talent B's in the (7,7)-game (the Chi-square test gives $\chi^2_2 = 1.91$, $p = 0.384$), where three chose LD, eight chose HR, and 13 chose Silence. However, the patterns are quite different in the (5,7)-game, where 14 low-talent B's chose LD, four chose HR, and seven chose Silence (the Chi-square test gives $\chi^2_2 = 16.19$, $p = 0.000$ and $\chi^2_2 = 10.23$, $p = 0.006$ for the two comparisons). Overall, the rate of LD messages from low-talent B's is much higher when they can potentially participate in a Pareto-improvement than when they cannot (50% versus 8%, $Z = 4.46$, $p = 0.000$), while the rate of Silence is much lower (25% versus 56%, $Z = -2.70$, $p = 0.007$).[10]

With respect to the responses of the A's to these messages, we see that 'promise' (HR and LD) messages induce *In* 53% of the time in the (5,5)-game, 72% of the time in the (7,7)-game, and 94% of the time in the (5,7)-game. As the rate in the (5,7)-game is significantly higher

[10] Regarding the messages of the high-talent B's, nine chose HR, two chose Silence, and two chose LD in the (5,7)-game; 11 chose HR and four chose Silence in the (5,5)-game; six chose HR, six chose Silence, and one chose LD in the (7,7)-game. The proportions of HR-messages in the three games do not differ significantly from any other.

than the rate in either of the other games ($Z = 3.33$, $p = 0.001$ and $Z = 2.01$, $p = 0.045$ for the respective comparisons), A's seem to believe that HR messages are more credible in this case.

4.3 Patterns of lies, truth & action

We now proceed to present some observations regarding the very interesting structure of lies, truth, and action in our data set. As we mentioned in the introduction, these systematic patterns may offer some 'lessons-for-life' for impostors interested in how best to deceive as well as for people who want to engage in lie detection.

We shall find it useful to refer to 'plans-of-action', equivalence classes of strategies that specify a message plus subsequent *Don't* or *Roll* choice, as in the following examples that explain our associated notation:

LD-then-D	= LD-message + *Don't* (in response to *In*)
HR-then-R	= HR-message + *Roll*
S-then-R	= Silence + *Roll*

In fact, we see some striking patterns in the message-action combinations for the low-talent B's. We focus on the messages LD and HR, which may be viewed as forms of promises each of which might make A choose *In*. Notice that given these two message options there are two possible ways for low-talent B's to act 'trustworthy', either LD-then-D or HR-then-D; the first of these does not involve being exposed as having sent a deceitful message, while the second one does, but in each case the low-talent B at last makes the non-opportunistic choice. Overall, low-talent B's choose LD-then-D 15 times, while choosing HR-then-D only once; a binomial test shows that this difference is not random ($Z = 3.21$, $p = 0.001$).

There are also two possible ways to act 'opportunistically', either HR-then-R or LD-then-R; once again, the first of these does not involve being exposed as having sent a deceitful message, while the second one does. For low-talent B's overall, LD-then-R occurred only twice,

while HR-then-R occurred 14 times; a binomial test shows that this difference is not random (Z = 3.00, p = 0.003). These patterns strongly suggest that people prefer to avoid exposed as liars, whether they choose to act trustworthy or opportunistically.[11]

In section 4.2, we noted that LD-messages are much more frequent in the (5,7)-game than in the other two games. In some sense it is not surprising that low-talent B's send so many more LD-messages when this can lead to a Pareto improvement or that low-talent B's who don't wish to lie are at a relative loss for words when it cannot. The surprise is rather the strong degree of trust and trustworthiness behavior that is induced by these messages in the (5,7)-game. Not only is it the case that A responds with *In* 13 of 14 times when a low-talent B sends a LD message, but it is also true that every (13 of 13) low-talent B who sends a LD-message chooses *Don't* when given the option. In fact, all five low-talent B's who chose *Roll* were amongst the six low-talent B's who had sent a HR-message. The difference in the *Don't* rates (100% versus 17%) is of course highly significant (Z = 3.83, p = 0.000).

In fact, the only time in the (5,7)-game with communication that it was dangerous to choose *In* was after a HR-message, since all other low-talent B's chose *Don't* when given the opportunity. The difference in both the messages and the subsequent choices of the low-talent B's when we compare the (5,7)-game to the two non-participation games drives the outcomes in the game, particularly as the proportion of low-talent B's who send a HR message differs little across games.[12]

---

[11] Overall, high-talent B's chose LD-then-R twice, HR-then-R 19 times, and S-then-R seven times. Once again, B's generally chose to avoid sending messages that would be exposed as being untrue.

[12] Low-talent B's sent HR messages six of 32 times (19%) in the (5,5)-treatment, eight of 29 times (28%) in the (7,7)-treatment, and six of 28 times (21%) in the (5,7)-treatment. None of these rates differ significantly from each other; furthermore, if we pool the non-participation treatments, 14 of 61 low-talent B's (23%) sent HR messages, nearly the same as in the (5,7)-treatment (Z = 0.16).

So, if a principal can offer a low-talent agent the chance to participate in a Pareto-improving outcome, we find that the agent will always be trustworthy. This is true despite the fact that the material payoff for acting opportunistically is nearly half again as large as the payoff from choosing *Don't*. Those people who can potentially participate in a Pareto-improvement and who confess to having low talent will perform up to the level of their ability. On the other hand, one should be skeptical of those who claim to be the best, as liars lurk among them.

5. BEHAVIORAL THEORY

When players are selfish, inefficient outcomes are predicted. This conclusion is unchanged when agents communicate. Models of distributional preferences such as Fehr & Schmidt (1999), Bolton & Ockenfels (2000), and (part of) Charness & Rabin (2002) provide an alternative approach that can accommodate more cooperative behavior if players dislike payoff inequality or have tastes for social efficiency. However, these models can not explain why communication makes low-talent B's more likely to choose *Don't* in the (5,7)-game, as the material payoff distributions do not depend on the preceding words.

Instead, we examine two behavioral models that permit communication to foster trust & cooperation: cost-of-lying and guilt-from-blame. We are not claiming that these are the only relevant behavioral theories,[13] only that recent developments suggest that they are worth scrutiny. Charness & Dufwenberg (2006) found support for guilt aversion in a hidden-action

[13] For example, in sociology and social psychology there is the notion of impression management, which is the process through which people attempt to influence how other people perceive them. The earliest reference in this area is Goffman (1956); for related contributions see Schlenker (1980), Tedeschi & Riess (1981), and Hannan, Rankin & Towry (2006). As impression management has not been formalized mathematically, we chose not to analyze the predictions of this theory. (Or perhaps we do, if avoiding guilt-from-blame, as described below, may be seen as one form of impression management.)

context, so it is natural to see how related ideas fare with hidden information.[14] Vanberg (2007) presents evidence suggesting that a preference for promise-keeping may explain our old data, and the more general concept of cost-of-lying has been emphasized by several scholars (see footnote 3) so we also find it natural to look at that concept.

Throughout we make the admittedly unrealistic assumption that the key psychological parameters involved (k in the case of cost-of-lying and θ in the case of guilt-from-blame) are commonly known among the players. As we deal with some fairly non-standard theory, we hope this approach helps highlight key insights regarding the psychological mechanisms at work, uncluttered by complicated signaling issues that otherwise would have to be addressed alongside.

One more comment before we proceed: Previous work testing for guilt aversion has elicited or induced beliefs, and a recent paper by Ellingsen, Johannesson, Tjøtta & Torsvik (2009) calls to question the accuracy of some of the measures and the conclusions drawn from them. The issues are hardly settled,[15] but there is surely a concern to acknowledge. We note that nothing in this study hinges on belief elicitation. The results of section 4 did not mention beliefs, and the theoretical implications below that we take to the data concern choices only.

5.1 Cost-of-lying

The key idea is that a person who utters a lie experiences an associated cost $k>0$. If there can be no communication there can be no cost-of-lying, so in the no-communication games the predictions correspond to the case with selfish preferences described in section 2: A chooses *Out* and B chooses *Roll* independently of talent.

---

[14] We say "related ideas" because guilt-from-blame differs somewhat from the form of guilt aversion considered by Charness & Dufwenberg (2006). Below we explain further and justify our focus, in light of recent theory-development, as well as empirical evidence.

[15] Reuben, Sapienza & Zingales (2009) present evidence that to some extent goes against Ellingsen *et al*'s.

In the communication games, however, the outcome may be improved. To see this let us first state precisely how we assume payoffs are affected. For player A (who cannot lie) payoffs will be as indicated in Figures 1-3 for any corresponding path of play. For each type of player B, payoffs will be as indicated in Figures 1-3 except that we must deduct k following paths that entail lies. For example, in the (5,7)-game, following path (*Low*, HD, *Out*) low-talent B's payoff is 5-k because he lied about his talent; following path (*Low*, LD, *In*, *Roll*) low-talent B's payoff is 10-k because he lied about his choice.[16]

Observation 1: In a (5,7)-communication game with cost-of-lying:

(i) If $k>3$ the strategy profile where A chooses *Out* and B chooses *Roll* independently of talent (and message) is not a SE.

(ii) If $k>3$ there is a SE where low-talent B uses plan-of-action LD-then-D, high-talent B uses HR-then-R, and A responds to messages LD and HR with *In*.

(iii) If $0<k<3$ the pattern of behavior described in (ii) can not appear in any SE.

All proofs (also of subsequent results) are in Appendix C. Parts (i) and (ii) of Observation 1 imply that adding communication when players have high cost-of-lying *fundamentally alters the prediction* relative to the case with selfish preferences. (As we shall see below, the guilt-from-blame theory discussed below does not have the analogous property.) The SEs described are not unique.[17] However, the prediction described in part (ii) is most compelling because it could also be obtained via solution concepts that do not assume equilibrium behavior,

[16] A list of all cases where B's payoff is decreased by k comprises those end nodes reached by the following paths: (*Low*, HD, *Out*), (*Low*, HR, *Out*), (*High*, LD, *Out*), (*High*, LR, *Out*), (*Low*, LD, *In*, *Roll*), (*Low*, LR, *In*, *Don't*), (*Low*, HD, *In*, *Roll*), (*Low*, HD, *In*, *Don't*), (*Low*, HR, *In*, *Don't*), (*Low*, HR, *In*, *Don't*), (*High*, HD, *In*, *Roll*), (*High*, HR, *In*, *Don't*), (*High*, LD, *In*, *Don't*), (*High*, LD, *In*, *Roll*), (*High*, LR, *In*, *Don't*), and (*High*, LD, *In*, *Roll*).

[17] For example, with $k\in(3,5)$ pooling by low- and high-talent B's on message LD is sustainable in SE (say with out-of-equilibrium inferences assigning probability 1 to messages LR, HD, HR, and S coming from low-talent B). With $k<3$ there exist mixed strategy SEs where A chooses *Out* except in response to HR where he chooses *In* with probability $k/5$; low-talent B uses HR-then-R with probability 1/2 and S-then-R with probability 1/2; high-talent B uses HR-then-R.

e.g. iterated elimination of weakly dominated strategies (applied to the game's normal form, treating low- and high-talent B as separate players) or extensive-form rationalizability (Pearce 1984; see also Battigalli 1995). It may also be seen as capturing an idea from the literature on cheap talk (non-binding costless communication): language conveys exogenously given meaning and players tend to believe what is said as long as such belief is consistent with rationality and the incentives given in the game.[18] Ponder the following story of *commitment* captured by the SE highlighted in part (ii): Each agent reveals his talent and cooperative choice-intention, and he neither lies nor reneges because that would trigger too much cost-of-lying.

The predictions for the (5,5)- and (7,7)-games are similar. We focus on the high k cases:

Observation 2: In a (5,5)- [(7,7)-]communication game with cost-of-lying, if $k>5$ [$k>3$] there is a SE where low-talent B uses plan-of-action LD-then-D, high-talent B uses HR-then-R, and A responds to messages LD and HR with *In*. There is also a a SE where low-talent B uses S-then-R, high-talent B uses HR-then-R, and A chooses *Out* except in response to message HR.

Observation 2 does not single out a particular choice for a low-talent B. Low-talent B may in SE use either LD-then-D or S-then-R; A would respond with *In* or *Out*, respectively, and A and the low-talent B would both get the same payoff regardless so there are no welfare consequences. The essence of Observation 2 is that high-talent B can signal his presence and intention with message HR, which is credible since low-talent B won't copy as k is too high. Player A chooses *In* in response, and efficiency is obtained.[19]

---

[18] For previous work that explores similar assumptions, see Rabin (1990), Farrell (1993), Farrell & Rabin (1996), Crawford (2003), Blume & Ortmann (2007), and Demichelis & Weibull (2008).

[19] As with Observation 1, the described SEs are not the only ones, just the plausible ones. There is also a SE where low-talent B uses S-then-R, high-talent B uses HD-then-D (!), and A assigns probability 1 to any messages except HD coming from low-talent B and responds to every message by *Out.* This pattern of behavior is, however, not plausible in the sense that it is again ruled out by iterated elimination of weakly dominated strategies or extensive-

The difference in dollar payoffs for a low-talent B between choices *Don't* and *Roll* is higher in the (5,5)-game than in the (5,7)- and (7,7)-games (10-5=5 instead of 10-7=3) and we need $k>5$ rather than $k>3$ to argue in favor of an efficient outcome. As we argued in section 2, the (5,7)- and (5,5)-games compare well, in the sense that one moves from the former to the latter through a subtle change in the underlying economic story (moving from asymmetric-but-positive agent gains to a step-aside-completely scenario). We suggested that it was intuitive that that change alone may cause trust & cooperation to deteriorate. A comparison of Observations 1 & 2 highlights why, with respect to testing that idea experimentally, a comparison of the (5,7)- and (5,5)-games is confounded in that different costs-of-lying are needed to support efficient outcomes in the two cases.[20] This explains why we also consider the (7,7)-game, which avoids this confound.

Let us finally, then, recall the data from section 4 and reflect on how well the cost-of-lying model accommodates it. First, while cost-of-lying may help explain why communication fosters trust & cooperation in the (5,7)-game (Observation 1), it provides equally strong support for an efficiency-enhancing effect in the (7,7)-game. This prediction was not borne out by the data, as trust & cooperation are distinctly lower in the (7,7)-game than in the (5,7)-game. Cost-of-lying alone does not help us explain why it matters whether we have asymmetric-but-positive gains or a step-aside-completely scenario. Second, the results reported in section 4.3, concerning patterns of lies & truth, suggest that decision makers *avoid being caught lying*. This is a nuance

---

form rationalizability, or the idea that players tend to believe what is said (here applied to message HR) as long as such belief is consistent with rationality and the incentives given.

[20] A comparison of the two games would be similarly confounded were we to take distributional preferences into account. For example, if a low-talent B is inequity averse he is more prone to choose *Don't* in the (5,7)-game than in the (5,5)-game. And an analogous confound arises with guilt-from-blame (cf. below).

that is not picked up by the cost-of-lying theory, according to which an uttered lie carries the consequences, rather than whether one has been caught.

5.2 Guilt-from-blame

Under this theory of Battigalli & Dufwenberg (2007), player *i* experiences guilt depending on the degree to which player *j* blames *i* for being willing to disappoint *j*. To develop this formally and give intuition we proceed as follows: Consider first the (5,7)-game without communication. Summarize the players' mixed strategies by p^{In}, p_L^R, and p_H^R, denoting the probability that A chooses *In*, low-talent B's choose *Roll*, and high-talent B's choose *Roll*, respectively. We assume that high-talent B's and A's cannot feel guilt, as they have no choice that can in expectation hurt another player. Anticipating the upcoming SE-definition,[21] we also assume that players have correct beliefs. Hence in SE, $p_H^R = 1$ and p^{In} must maximize A's subjectively expected material payoff, which we denote by α. With $p_H^R = 1$ we get:

$$\alpha = 5 \cdot (1 - p^{In}) + (\tfrac{2}{3} \cdot [7 \cdot (1 - p_L^R) + 0 \cdot p_L^R] + \tfrac{1}{3} \cdot [\tfrac{5}{6} \cdot 12 + \tfrac{1}{6} \cdot 0]) \cdot p^{In} = 5 \cdot [1 - p^{In}] + [8 - \tfrac{14}{3} \cdot p_L^R] \cdot p^{In}$$

To state and explain low-talent B's utility we need α as well as two more key variables, which we label λ and θ. λ is the probability A assigns to the leftmost node in the information set where she receives a 0 payoff. In SE, applying Bayes' rule, and using $p^{In} = 1$, we get:

$$\lambda = \frac{\frac{2}{3} \cdot p_L^R}{\frac{2}{3} \cdot p_L^R + \frac{1}{3} \cdot \frac{1}{6}} = \frac{12 p_L^R}{12 p_L^R + 1}$$

We can now state low-talent B's utility and best response (and in the process introduce θ) in a SE where A chooses *In* (p^{In}=1). Low-talent B experiences guilt only if he chooses *Roll*,

[21] Battigalli & Dufwenberg (2009) extend Kreps & Wilson's SE definition to psychological games.

the choice that hurts player A and that might lead A to blame low-talent B. To determine his best response low-talent B compares the (guilt-free) payoff of 7 from choosing *Don't* to the payoff from *Roll*, which is

$$10 - \theta \cdot \lambda \cdot \min\{7, \alpha\}$$

This expression describes utility as material payoff (=10) minus guilt-from-blame ($=\theta \cdot \lambda \cdot \min\{7, \alpha\}$). We explain the latter term walking through its factors from right to left. The expression $\min\{7, \alpha\}$ measures how much A would blame low-talent B (and how much guilt low-talent B would then experience) *were it known* that low-talent B chose *Roll*; α is the difference between what A initially expected ($=\alpha$) and what he actually received (=0) due to low-talent B's opportunistic choice. The 7 is present in the expression because the blame/guilt is capped at 7, since this is the full payoff difference that low-talent B actually controls. Regarding λ, note that because of A's information set across the end nodes where he receives 0, he will actually never know for certain that low-talent chose *Roll*. λ captures an assumption that a low-talent B is sheltered from guilt to the extent that A isn't sure that B is blameworthy. Notice that A assigns probability 1- λ to the event that she received a payoff of 0 due to path (*High*, *In*, *Roll*, *Failure*), which would just be bad luck and no fault of a low-talent B. Finally, θ is a non-negative constant, describing how sensitive *i* is to feelings of guilt-from-blame. If $\theta = 0$, a low-talent B would be selfish.

At this point we wish to make two important comments about guilt-from-blame. First, one may model guilt in many ways. Battigalli & Dufwenberg (2007) offer two models. In one variety (simple guilt) player *i* internalizes the emotion in the sense that he feels guilt when he

believes he disappoints another player *j*, regardless of what *j* believes *i*'s intentions are.[22] Guilt-from-blame is the other variety, where guilt is driven rather by what *i* believes *j* believes about *i*'s intentions as regards disappointing *j*. The goal of our paper is *not* to test simple guilt against guilt-from-blame. Rather we focus only on the latter concept (which we compare with cost-of-lying), the reason being a recent string of papers (Dana, Cain & Dawes 2006, Dana, Weber & Kuang 2007, Broberg, Ellingsen & Johannesson 2007, Lazear, Malmendier & Weber 2009, Tadelis 2008) that suggest in various ways that players are more prone to selfless choice to the extent that others will know about it.[23] Guilt-from-blame caters to such concerns through the way λ affects utility. Second, since the key elements α and λ depend on beliefs, specifying low-talent B's utility requires the framework of psychological game theory. In principle, this could be complicated. For example, A's subjectively expected payoff α conceptually should depend on A's *beliefs* about p_L^R and p_H^R, not on p_L^R and p_H^R themselves. However, our focus on equilibrium (SE) simplifies matters considerably, as players have correct beliefs about p^{In}, p_L^R, and p_H^R.

Drawing on the above notations and calculations, we now state SE conditions formally:

Definition 1: Let $\alpha = 5 \cdot [1 - p^{In}] + [8 - \frac{14}{3} \cdot p_L^R] \cdot p^{In}$ and $\lambda = \dfrac{12 p_L^R}{12 p_L^R + 1}$. A SE in the (5,7)-game, when low-talent B is sensitive to guilt-from-blame, is a triple (p^{In}, p_L^R, p_H^R) such that:

---

[22] Although the distinction with guilt-from-blame had not yet been conceptualized when Charness & Dufwenberg (2006) was written, in retrospect we see that simple guilt was in focus in that paper.

[23] For example, dictators can either divide $10 (in which case the recipient learned of the dictator game and the dictator's choice) or choose to exit and take a smaller amount, in which case the would-be recipient would not learn of the dictator game. Many people choose to exit. In fact, Dana, Cain & Dawes-find that 43% exit when the would-be recipient would learn of the dictator game without exit, but only 4% exit when the would-be recipient would never learn of the dictator game even if exit is forgone. Tadelis uses the same game as Charness & Dufwenberg (2006), but varies whether the principal will learn of the actual choice made by the agent. In two separate comparisons, he finds that *Roll* rates are nearly twice as high with this exposure than when the agent knows that the principal will not learn his choice.

(i) p^{In} maximizes α
(ii) p_L^R maximizes $(1- p_L^R)\cdot 7 + p_L^R \cdot (10 - p^{In}\cdot\theta\cdot\lambda\cdot\min\{7,\alpha\})$
(iii) $p_H^R = 1$

We can state analogous definitions for the (5,5)- and (7,7)-games. For the (7,7)-game, the definition is *identical*, except that the two numbers "5" in Definition 1 should be replaced by "7". As regards the (5,5)-game the specification changes more:

Definition 2: Let $\alpha = 5\cdot[1 - p^{In}] + [\frac{20}{3} - \frac{10}{3}\cdot p_L^R]\cdot p^{In}$ and $\lambda = \dfrac{12p_L^R}{12p_L^R + 1}$. A SE in the (5,5)-game, when low-talent B is sensitive to guilt-from-blame, is a triple (p^{In}, p_L^R, p_H^R) such that:

(i) p^{In} maximizes α
(ii) p_L^R maximizes $(1- p_L^R)\cdot 5 + p_L^R \cdot (10 - p^{In}\cdot\theta\cdot\lambda\cdot\min\{5,\alpha\})$
(iii) $p_H^R = 1$

Applying these definitions we get multiple SEs once θ is high enough:

Observation 3: In both the (5,7)-game and the (7,7)-game, when low-talent B is sensitive to guilt-from-blame:

(i) For any $\theta \geq 0$ there is a SE with $(p^{In}, p_L^R, p_H^R) = (0, 1, 1)$
(ii) If $\theta \geq \frac{25}{42}$ there is a SE with $(p^{In}, p_L^R, p_H^R) = (1, \frac{1}{28\theta - 12}, 1)$

Observation 4: In the (5,5)-game, when low-talent B is sensitive to guilt-from-blame:

(i) For any $\theta \geq 0$ there is a SE with $(p^{In}, p_L^R, p_H^R) = (0, 1, 1)$
(ii) If $\theta \geq \frac{7}{6}$ there is a SE with $(p^{In}, p_L^R, p_H^R) = (1, \frac{1}{12\theta - 12}, 1)$

Note several things: First, parts (i) of Observations 3 & 4 describe inefficient zero-trust play by A and no cooperation by low-talent B's. The intuition for why this pattern is allowed for any θ is that if low-talent B initially expects A to choose *Out*; then B believes that A then can't blame low-talent B, who therefore does not feel guilt. It is true that if A were to deviate, then a

low-talent B would realize that he can affect A's payoff, so in principle one might imagine that guilt could come into play. However, as the theory is constructed (through the presence of factor p^{In} in parts (i) of Definitions 1 & 2) blame & guilt is only relevant to the extent that A believes low-talent-B believes *initially* that low-talent B set out to disappoint A. Second, it is impossible in each of the games to have a full-trust-&-cooperation SE with (p^{In}, p_L^R, p_H^R) = (1, 0, 1). In that case we would get $\lambda = 0$ and low-talent B would be entirely sheltered from blame & guilt and so choose *Roll,* i.e. $p_L^R = 1$, a contradiction. Instead, the SEs reflecting the most trust & cooperation involve mixing by low-talent B. For example, in the (5,7)-game, he chooses *Roll* with probability $p_L^R = \frac{1}{28\theta - 12}$. Note that $p_L^R \to 0$ as $\theta \to \infty$. Third, the SEs described for the (5,7)- and (7,7)-games coincide (Observation 3; Appendix A also comments on some additional SEs for the (5,7)-game which are not covered in Observation 3). The (5,5)-game is different (Observation 4); because of the difference between parts (ii) of Definitions 1 & 2 there is a confound for comparing behavior in the (5,5)- and (5,7)-games analogous to what we discussed for cost-of-lying. So again our main comparison as regards whether the theory can explain why it matters whether we have asymmetric-but-positive gains or a step-aside-completely scenario will center on comparing the (5,7)- and (7,7)-games. Fourth, unlike in the case with cost-of-lying, rationalizability will not help pin down a clear prediction; one can show that for any $\theta \geq \frac{25}{42}$ each of A's and low-talent B's strategies is rationalizable (as defined by Battigalli & Dufwenberg 2009 who extend Pearce's extensive form rationalizability notion to psychological games). Fifth, in light of the presence of multiple SEs when θ is large enough, we face an equilibrium-selection problem.

What happens when the communication stage is added (with messages LD, LR, HD, HR, and S, just as before)? The first thing to note is that (unlike in the case with cost-of-lying) we cannot hope to get an automatic move of the set of SEs in the direction of enhanced efficiency. In particular, there is no SE with full revelation + separation + honesty. To see this, imagine for example that low-talent B uses LD-then-D, high-talent B uses HR-then-R, and A chooses *In* if and only if she gets message LD or HR. The argument regarding why this cannot be part of a SE is analogous to that which ruled out, for the games without communication, an SE with $(p^{In}, p_L^R, p_H^R) = (1, 0, 1)$. If inferences were based on HR-messages never coming from low-talent B's, then a low-talent B would be safe choosing *Roll* as he wouldn't be blamed if he actually sent a HR-message and then chose *Roll*. Following HR, we'd have $\lambda = 0$ and a complete shelter for low-talent B's feelings of guilt.

On the other hand, every pattern of SE play described for the games without communication is also attainable via some SE in the games with communication. For example, consider the (5,7)-game and suppose $\theta > \frac{25}{42}$. The SE with $(p^{In}, p_L^R, p_H^R) = (0, 1, 1)$ in part (i) of Observation 3 could be matched if low- and high-talent talent both use use HR-then-R while A responds to any message message with *Out*. The SE with $(p^{In}, p_L^R, p_H^R) = (1, \frac{1}{28\theta - 12}, 1)$ in part (ii) of Observation 3 could be matched if low-talent B uses HR-then-R with probability $\frac{1}{28\theta - 12}$ and otherwise LD-then-D; high-talent B uses HR-then-R; A responds to both LD and HR with *In* but would respond to any other message with choice *Out*.

Communication may, however, help the players coordinate on a favorable SE. One-way communication has been found to lead to coordination on a strictly Pareto-superior equilibrium in papers such as Charness (2000). This could be relevant to the two SEs for the (5,7)-game with

$\theta > \frac{25}{42}$ described in the previous paragraph, which are indeed strictly Pareto-ranked. But a key insight is that this idea does not extend to the (5,5)- and (7,7)-games! While these games also exhibit multiple SEs, no strict Pareto-gains are available. In particular, a low-talent B lacks a strict incentive to sway A away from his choice *Out* by promising A he will choose *Don't*. The low-talent B gets exactly the same payoff from choosing *Don't* after A chooses *In* as when A chooses *Out*, as does A.

Let us finally, then, recall the data from section 4 and reflect on how well the guilt-from-blame theory accommodates it. First, even without communication selfless choice is possible if players are motivated by guilt-from-blame, so guilt-from-blame can help explain why in the experimental (5,7)-game we saw considerable deviations from the selfish SE. Second, guilt-from-blame may help explain why communication fosters additional trust & cooperation in the (5,7)-game as well as the observed differential effect of communication in this game in comparison with the (5,5)- and (7,7)-games, if we add the idea (admittedly from outside the guilt-from-blame theory proper) that one-sided communication helps players coordinate on a strictly Pareto-superior SE. That idea does not apply to the (5,5)- and (7,7)-games. Third, regarding the patterns of lies & truth reported in section 4.3, we did not derive these through Observations 3 or 4, as doing so would require some extra structure on out-of-equilibrium inferences. However, the following pattern of inferences would naturally produce the result that no low-talent B uses LD-then-R: Suppose B uses LD-then-R, so that A receives 0. A knows B lied, but whether A blames B depends on whether she thinks B had low or high talent. On the presumption that high-talent B's always choose HR-messages, A would interpret an LD-message as coming from a low-talent B. Given that inference, a low-talent B would refrain from LD-then-R, in line with the data.

6. CONCLUSION

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 cannot be upset by the addition of communication. Yet, the human side of contracting seems a bit less dismal.

In this paper we explore whether and how communication can achieve beneficial social outcomes in a hidden-information context. It turns out that whether communication affects behavior depends crucially on whether low-talent agents can participate in an outcome that, compared to no contractual agreement, is a Pareto-improvement for the principal and the agent. When low-talent agents can participate in this way, communication is quite effective; the great majority of these agents behave cooperatively, foregoing the additional earnings that could be pocketed. However, when participation for low-talent agents is infeasible, selfish behavior on the part of these agents predominates, whether or not communication is feasible.

We present the predictions from two relevant behavioral models, one that involves a cost-of-lying and one that involves guilt-from-blame. When communication is allowed, both theories offer some scope for trust & cooperation, although the mechanisms differ. Under (high enough) cost-of-lying, incorporating communication leads to new equilibria that embody Pareto-gains (predictions that also obtain with solution concepts like iterated weak dominance, extensive form rationalizabilty, or full permissibility). The cost-of-lying theory does not, however, predict a difference depending on whether or not Pareto-gains are feasible for both talent levels for B, as all those who gain can unilaterally credibly separate.

With guilt-from-blame on the other hand, allowing communication does not add new patterns of equilibrium play. There are multiple equilibria both with and without

communication. Communication may, however, enhance trust & cooperation not by expanding the possible patterns of equilibrium play, but rather by facilitating equilibrium coordination. Previous experimental studies (e.g. Charness 2000) have suggested that communication has this efficiency-enhancing property only when equilibria can be strictly Pareto-ranked. The coupling of this idea with the guilt-from-blame theory squares nicely with the data, and can shed light on why it is easier to obtain efficient outcomes when everyone gains than when some are excluded.

Our data exhibit some systematic patterns regarding how people lie and tell truth, in the game where there are gains for all (the (5,7)-game). Liars claim to be better then they are, as if they meant to suggest that the subsequent bad outcome was due to bad luck rather than opportunistic choice. Trustworthy people, on the other hand, truthfully reveal their level of talent and can then be relied upon to do as well as they can. These results provide some 'useful lessons' that, on extrapolation, may offer useful guidance for those who want to deceive others, as well as for people trying to tell if someone else is being honest. A claim that the agent has high talent should be viewed with some suspicion, as it often 'the big lie'. However, when participation is possible regardless of the agent's talent, 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; it seems that one can trust people who confess imperfections.

Perhaps the notion that permitting participation in Pareto-improvements applies in the field as well, so that low-talent people in the real world will also manifest this sort of behavior. The principle can also be extended into other realms in which the quality level is not readily observable, such as e-commerce.[24] It appears to be the case that people are substantially more

[24] We thank Ulrike Malmendier for the following example. If an internet seller expects buyers to only be interested in a brand-new item, he is likely to claim that the item for sale is new, whether or not it is. However, if the seller believes that there is a market for used items in good condition, perhaps he is much more likely to confess the item is used, but claim that it is nevertheless in excellent condition. Of course, this argument requires that online

prone to be cooperative when they can participate by having a voice and choosing 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 (1997), "On Rationalizability in Extensive Games", *Journal of Economic Theory,* **74**, 40–61.

Battigalli, Pierpaolo & Martin Dufwenberg (2007), "Guilt in Games", *American Economic Review Papers and Proceedings*, **97**, 170-176.

Battigalli, Pierpaolo & Martin Dufwenberg (2009), "Dynamic Psychological Games", *Journal of Economic Theory*, **144**, 1-35.

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 (2007), "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*.

---

reputation systems are less-than-perfect, but people do to some extent game this system by tactics such as changing online identities after misbehavior.

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.

Ellingsen, Tore, Magnus Johannesson, Sigve Tjøtta & Gaute Torsvik (2007) "Testing Guilt Aversion", forthcoming in *Games and Economic Behavior*.

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.

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.

Goffman, Erving (1956), "Embarrassment and Social Interaction," *American Journal of Sociology*, **62**, 264-271.

Hannan, R. Lynn, Frederick Rankin & Kristy Towry (2006), "The Effect of Information Systems in Managerial Reporting: A Behavioral Perspective", *Contemporary Accounting Research*, **23**, 885-918.

Hurkens, Sjaak & Navin Kartik (2009), "Would I Lie to You? On Social Preferences and Lying Aversion", *Experimental Economics*. **12**, 180-192.

Kartik, Navin (2008), "Strategic Communication with Lying Costs", mimeo.

Kreps, David & Robert Wilson (1982), "Sequential Equilibrium", *Econometrica*, **50**, 863-894.

Lazear, Edward, Ulrike Malmendier & Roberto Weber (2009), "Sorting and Social Preferences," mimeo.

Miettinen, Topi (2008), "Contracts and Promises - An Approach to Pre-play Agreements", SSH/EFI Working Paper No 707, Stockholm School of Economics.

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

Rabin, Matthew (1990), "Communication Between Rational Agents," *Journal of Economic Theory*, **51**, 144-70.

Reuben, Ernesto, Paolo Sapienza & Luigi Zingales (2009), "Is Mistrust Self-Fulfilling?", *Economics Letters*, **104**, 89-91.

Schlenker, Barry (1980), *Impression Management: The Self-Concept, Social Identity, and Interpersonal Relations*, Monterey/California: Brooks/Cole.

Sutter, Matthias (2009), "Deception Through Telling the Truth?! Experimental Evidence from Individuals and Teams", *Economic Journal*, **119**, 47-60.

Tadelis, Steven (2007), "The Power of Shame and the Rationality of Trust", mimeo.

Tedeschi, J. & M. Riess (1981), "Identities, the phenomenal self, and laboratory research", in *Impression Management Theory and Social Psychological Research*, ed. J. Tedeschi, 3-22, New York: Academic Press.

Vanberg, Christoph (2008), "Why Do People Keep Their Promises? An Experimental Test of Two Explanations", *Econometrica*, **76**, 1467-1480.

Appendix A: Sample Instructions [(5,7)-game with communication]

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 $7 and B receives $7. 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	**$5**	**$5**
A chooses IN and:		
B is LOW type and chooses DON'T ROLL	**$7**	**$7**
B is LOW type and chooses ROLL	**$0**	**$10**
B is HIGH type and chooses DON'T ROLL	**$7**	**$7**
B is HIGH type, chooses ROLL, die=1	**$0**	**$10**
B is HIGH type, chooses ROLL, die= 2,3,4,5, or 6	**$12**	**$10**

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.

Appendix B: Messages

The (5,7)-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, and N = No Message. This classification scheme maps onto the scheme in the text as follows: 1) HR, IHR, and PH are considered to be HR messages, 2) LD, ILD, PD, and PNS are considered to be LD messages (promises one will choose *Don't* or not be selfish only really make sense for a low-talent agent), and 3) E and N are considered to be Silence.

Session	ID (B)	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? Live dangerously. ☺ 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 Low Type. Please choose IN because I will not choose ROLL. I will not use you for my benefit.	LD	L	IN	D
1	11	I am LOW TYPE. I am going to choose "Don't roll" because I want both 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	-	N	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	-	N	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</u> B. 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 amound 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. IN = 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 GIVE YOU MY WORD. 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 In you have a 5 in 6 chance of getting 17. Much better than 10.	IHR	H	IN	R

The (5,5)-game

Session	ID (B)	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	-	N	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 high 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 in; 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 ~~$10~~ $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	-	N	H	IN	R
2	10	-	N	L	IN	R
2	11	-	N	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	-	N	L	IN	R
2	14	-	N	L	OUT	-
2	15	I'm a B high type – I'm rolling.	HR	H	OUT	-
2	16	-	N	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 IN! If I am a low type I'll choose Don't Roll and we both get $5. But, if I am a High type I'll choose roll and you'll have a 5/6 chance to win $12.	PNS	L	IN	R
3	3	Hey, Choose in because we will "both" make a lot of money, or at least we both have an extremely high probability of making a lot of money.	IHR	H	OUT	-
3	4	Choose 'In' you have a good chance of making more than me. It will benifit both of us.	IHR	L	IN	R
3	5	-	N	L	IN	D
3	6	You should choose IN and since I am HIGH I will roll and the likelihood of me getting a 2-6 is 5/6 (about (84%) chance that you'll get 12$ 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	-	N	L	IN	R
3	9	-	N	H	IN	R
3	10	☺ g'day!	E	L	IN	D
3	11	-	N	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	-	N	L	OUT	-
3	14	-	N	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

The (7,7)-game

Session	ID (B)	Message	Class	Type	A	B
1	1	Dear numbers 1,2,3,4,5,6,7, I'm so sorry. You know I need this money for my life. My life is humble. Sorry I've killed you	E	L	OUT	-
1	2	not sure what to write here	E	L	OUT	-
1	3	Chose IN and you have a high chance of making cash money. "Skinner said the teachers will crack any minute Purple monkey Dishwasher" - Simpsons	IHR	H	IN	R
1	4	Let's make money. I'm a low type B	PL	L	IN	DR
1	5	I'll choose DON'T ROLL, so you're good!	PD	L	IN	R
1	6	Hi Partner – Please Choose in. I am a high type so you have a very h chance of receiving 12$ instead of 7$ dollars if you choos because I am planning on choosing in. Thanks!	PH	H	IN	R
1	7	-	N	L	IN	R
1	8	Let's go for $12 & $10 ☺	IHR	L	IN	R
1	9	If you choose IN we can both benefit more!	IHR	H	IN	R
1	10	We'll break even	ILD	L	OUT	-
1	11	-	N	L	OUT	-
1	12	Lets gamble!	E	H	IN	R
1	13	Your decision is based upon whether I turn out to be a high or low B. If I'm low, then you have the possibility of getting only $5 today … doesn't matter to me, I'm not mean … so I'd give you money. $3 bucks isn't a big deal ☺	ILD	L	IN	DR
1	14	-	N	L	OUT	-
1	15	[In-In]-out? ☺	E	H	IN	R
1	16	Good luck on this.	E	L	OUT	-
2	1	Dear A, I am a low B.	PL	L	OUT	-
2	2	#12 high Go IN	HR	L	OUT	-

		I ROLL				
2	3	☺	E	H	OUT	-
2	4	-		L	OUT	-
2	5	Choose in please so we can get some money	E	L	IN	R
2	6	-	N	H	IN	R
2	7	I'm a Low. Probably.	PL	L	OUT	-
2	8	I am a Low type B.	PL	L	OUT	-
2	9	Please choose In. I'm high and not in the illegal way. Please believe it. ☺	PH	H	IN	R
2	10	-	N	L	IN	R
2	11	We both have a good chance for a nice payout	IHR	L	IN	R
2	12	-	N	H	OUT	-
2	13	FYI: you have no/little chance for an extra payout. Decide accordingly.	PL	L	OUT	-
2	14	Choose "In" for a better pay out	IHR	L	OUT	-
2	15	My dear friend The decision you are about to make is going to change your life forever! Just kidding ☺ take it easy pal	E	H	IN	R
3	1	Good luck	E	L	OUT	-
3	2	I'm high (not from weed … just my number). Choose in. There will only be a 1/6 chance you get nothing and a 5/6 chance to get $12. We will both make more money. I won't roll a one I promise lol.	HR	L	IN	R
3	3	I am a high type As long as you choose IN, we have 5/6 chance of getting max money	HR	H	IN	R
3	4	If you choose "IN", I will choose ROLL.	E	L	OUT	-
3	5	I am 3!!! Lets leave with 17 and 15	PH	L	OUT	-
3	6	I am <u>High</u>. I don't think there is an interest to lie. So believe it I'm High: the only way you get $0 is if I roll 1 otherwise you get $7 or $12: let's do it!	PH	H	IN	R
3	7	-	N	L	OUT	-
3	8	I'm High	PH	L	IN	R
3	9	I am not going to roll the dice.	PDR	H	OUT	-
3	10	-	N	L	OUT	-
3	11	Think positive & let's both leave here with some major cash flow ☺	IHR	L	IN	R

Appendix C: Proofs

Proof of Observation 1

(i) In SE, if a low-talent B sends message LD he must follow up with choice *Don't* (since 7>10-*k*). (Note that it now also follows that A must respond by *In* to message LD; in SE high-talent B chooses *Roll* after message LD so by choosing *In* after LD player A gets at least 7>5).

(ii) The described SE profile describes sequentially rational play, as no player has a profitable unilateral deviation. To pin down a SE just add an appropriate specification for out-of-equilibrium beliefs, e.g. probability 1 to low-talent B following messages LR, HD, or S, and choices following those messages for each type of player B.

(iii) A low-talent B would have a unilateral incentive to deviate to HR-then-R.

Proof of Observation 2:

It is straightforward to assert that no player has a unilateral deviation incentive. We leave the specification of complete strategies and out-of-equilibrium inferences for the reader.

Proof of Observation 3:

(i) In the text we specified B's utility in a SE where A chooses *In* (p^{In}=1) and a low-talent B chooses *Roll* as $10 - \theta \cdot \lambda \cdot \min\{7, \alpha\}$. That omitted cases where p^{In}<1; the more general statement (cf. (2) in Battigalli & Dufwenberg 2007, p. 172) is

that in a SE where A chooses *In* with probability p^{In} the utility of a low-talent B who chooses *Roll* equals $10 - p^{In} \cdot \theta \cdot \lambda \cdot \min\{7, \alpha\}$. (This is seen also in parts (ii) of Definition 1 & 2.) The guilt term thus vanishes when p^{In}=0. A low-talent B thus chooses *Roll* since 10>7, so p_L^R=1. We know from Definition 1(iii) that p_H^R = 1, and we see that A's best response is *Out*. All in all, (p^{In}, p_L^R, p_H^R) = (0, 1, 1).

(ii) A low-talent B randomizes and so must be indifferent between *Don't* and *Roll*: $7 = 10 - p^{In} \cdot \theta \cdot \lambda \cdot \min\{7, \alpha\}$. This equation can be simplified:

- $p^{In} = 1$ is given by the SE
- $\alpha = 5 \cdot [1 - p^{In}] + [8 - \frac{14}{3} \cdot p_L^R] \cdot p^{In} = 8 - \frac{14}{3} \cdot p_L^R$ since $p^{In} = 1$
- $p_L^R = \frac{1}{28\theta - 12}$ is given by the SE and since $\theta \geq \frac{25}{42}$ we get $p_L^R \leq \frac{3}{14}$, which in turn implies that $\alpha = 8 - \frac{14}{3} \cdot p_L^R \geq 7$, so that $\min\{7, \alpha\} = 7$.

Thus, we have that $7 = 10 - \theta \cdot \lambda \cdot 7$. Plug in $\lambda = \frac{12 p_L^R}{12 p_L^R + 1}$ and solve for p_L^R as a function of θ to verify that $p_L^R = \frac{1}{28\theta - 12}$. We know from Definition 1(iii) that p_H^R = 1, and we see that A's best response is indeed *In*. All in all, (p^{In}, p_L^R, p_H^R) = (1, $\frac{1}{28\theta - 12}$, 1).

Comment on Observation 3: All of the SEs described under part (ii) give A a payoff of at least 7, but in the (5,7)-game (and not the (7,7)-game) there are also SEs where A chooses *In* and receives a payoff in the range (5, 7). In these cases, $\frac{3}{14} < p_L^R \leq \frac{9}{14}$, and $\min\{\alpha, 7\} = \min\{8 - \frac{14}{3} \cdot p_L^R, 7\} = 8 - \frac{14}{3} \cdot p_L^R$. The dependence on p_L^R means that when calculating

the SEs one must solve the quadratic equation $7 = 10 - \theta \cdot \frac{12 p_L^R}{12 p_L^R + 1} \cdot (8 - \frac{14}{3} \cdot p_L^R)$, which describes the relevant indifference condition for a low-talent B. Additional SE appear for values of θ slightly lower than 25/42 (down to slightly more than 0.54), as well as much higher values of θ (<61/18). Some manipulations show that relevant roots satisfy $p_L^R =$ $\frac{(24\theta - 9)}{28\theta} \pm ((\frac{24\theta - 9}{28\theta})^2 - \frac{3}{56\theta}))^{\frac{1}{2}}$ as well as $\frac{3}{14} < p_L^R < \frac{9}{14}$.

Proof of Observation 4: The technique is analogous to that in the proof of Observation 3 and is therefore omitted.

Figure 1: The (5,7)-game

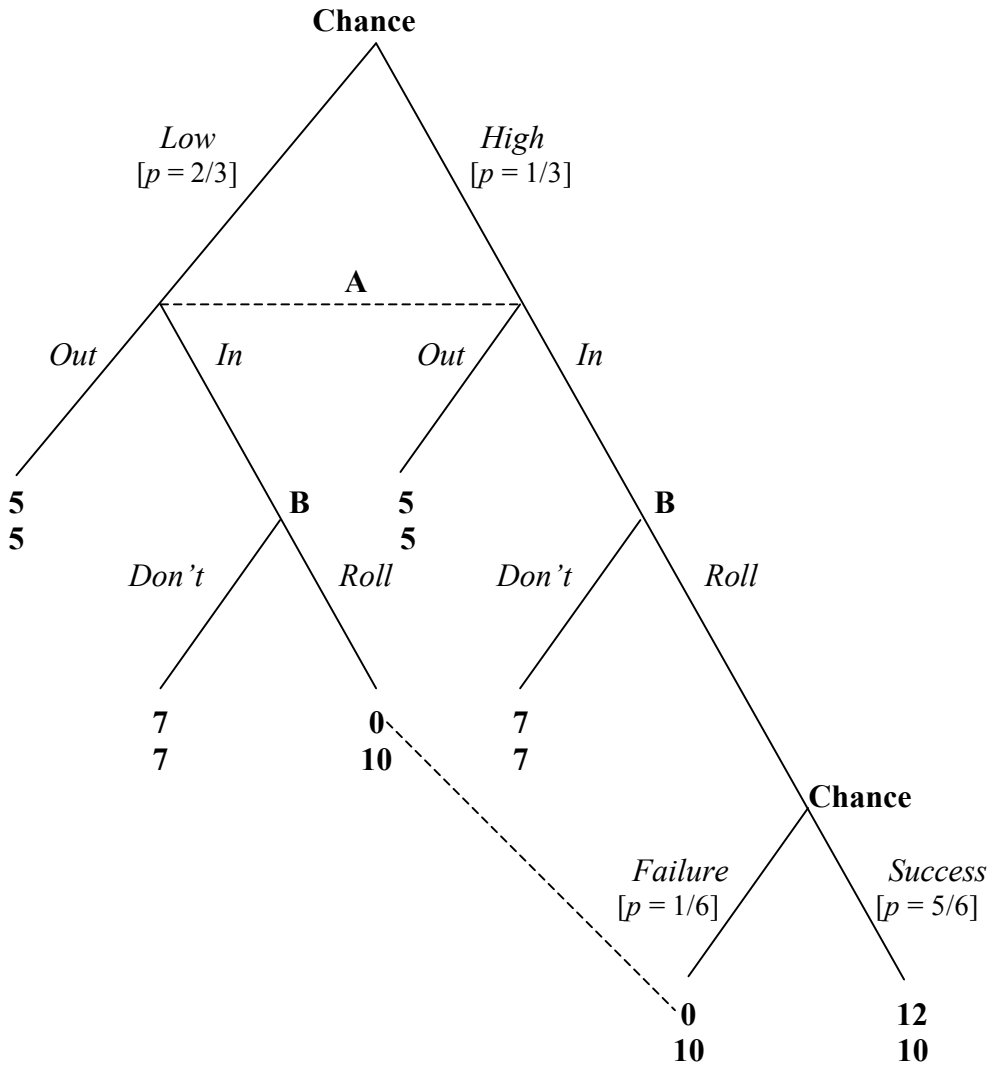


Figure 2: The (5,5)-game

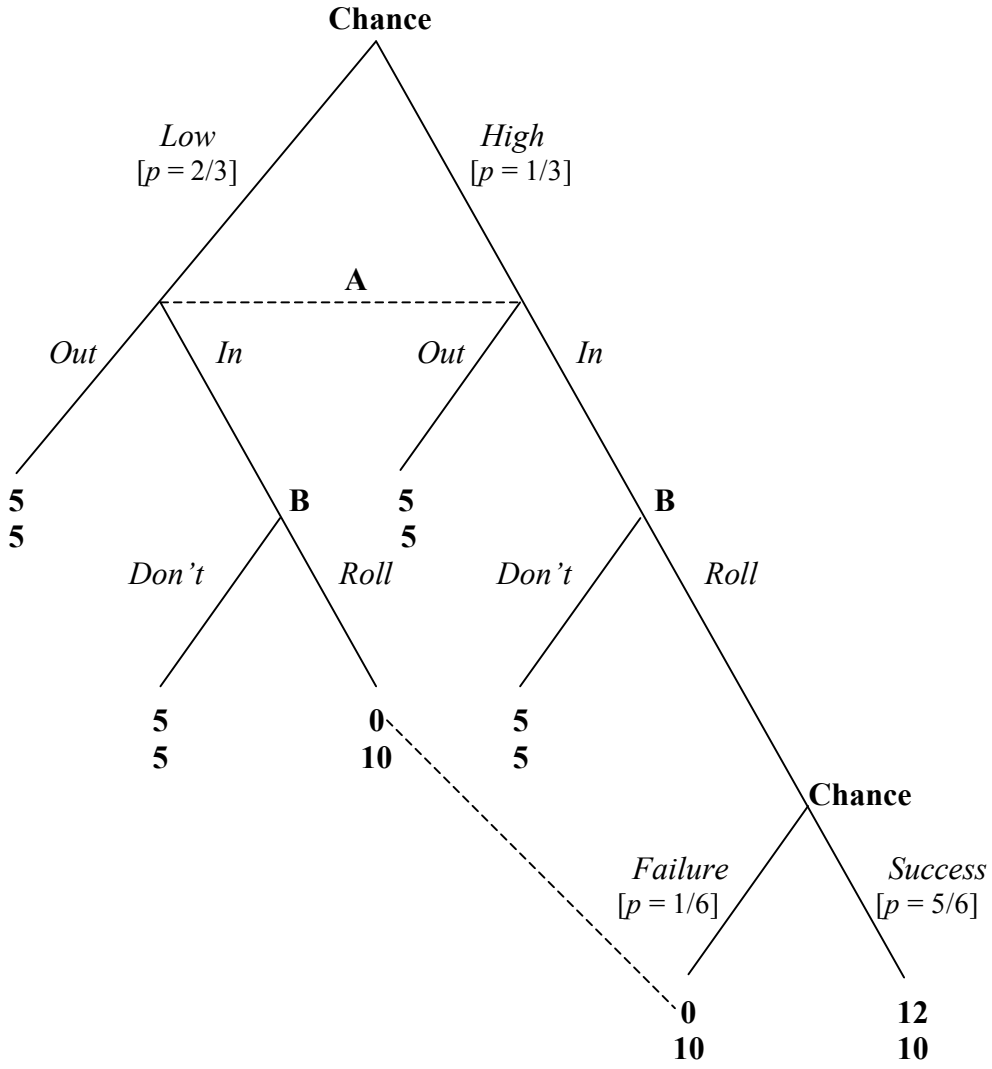


Figure 3: The (7,7)-game