*The Economic Journal*, **114** (*April*), 397–420. © Royal Economic Society 2004. Published by Blackwell Publishing, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA.

# PROMISES, THREATS AND FAIRNESS*

*Tore Ellingsen and Magnus Johannesson*

We present experimental evidence that promises and threats mitigate the hold-up problem. While investors rely as much on their own threats as on their trading partner's promises, the latter are more credible. Building on recent work in psychology and behavioural economics, we then present a simple model within which agents are concerned about both fairness and consistency. The model can account for several of our experimental findings. Its most striking implication is that fairmindedness strengthens the credibility of promises to behave fairly, but weakens the credibility of threats to punish unfair behaviour.

*Better is it that thou shouldest not vow, than that thou shouldest vow and not pay.*

Ecclesiastes 5:5

*To breed an animal with the right to make promises – is not this the paradoxical task that nature has set itself in the case of man? Is it not the real problem regarding man?*

Friedrich Nietzsche (1887)[1]

There is little doubt that many people have an opportunistic streak, seeking personal advantage even when the cost to others is greater than the benefit to themselves. Nor is there any doubt that opportunism is harmful to society. Opportunistic activities like shirking and theft are unproductive and in turn erode others' incentive to engage in production.

According to conventional economic theory, with its generally bleak view of human nature, there is only one way to protect society against opportunism, namely to establish institutions that punish socially harmful behaviour. Possible arrangements include formal and informal sanctions as well as production and dissemination of information. (Information is important because it affects cheaters' reputational loss.) To most economists, Nietzsche's above quoted distinction between man and beast does not exist. No promise or threat is credible in and of itself. Only third-party enforcement or reputational concerns make verbal commitments credible.

* An earlier version of this material has been circulated as part of our working paper 'Is There a Hold-up Problem?' We are grateful to Magnus Allgulin, Mats Ekelund, Freddie Henriksson, Douglas Lundin, Arvid Nilsson, Joakim Ramsberg and Niklas Zethraeus for research assistance. Thanks to Juan Carrillo, Ernst Fehr, Oliver Hart, Steffen Huck, Bengt Holmström, David Kreps, Paul Milgrom, Torsten Persson, Randolph Sloof, Torben Tranæs and Jörgen Weibull for helpful discussions. Two editors and two anonymous referees provided very valuable comments and suggestions. The paper has also benefited from the comments of seminar audiences at ECARE (Bruxelles), Norwegian School of Economics and Business Administration, Stockholm School of Economics, University of Amsterdam, University of Copenhagen and University of Stockholm.

[1] Second essay, Sect. 1.

In this paper, we present experimental evidence suggesting that verbal commitment is more credible than economists have been willing to admit. This evidence echoes and extends earlier findings by psychologists and sociologists that non-binding promises to cooperate in social dilemmas are often credible, as was demonstrated already by Loomis (1959).[2] The social dilemma that we consider is known as *the hold-up problem*. Specifically, we conduct a sequence of two-person experiments in which one party, the seller, has the opportunity to invest 60 Swedish kronor (about 8 US dollars at the time) to create a benefit of 100 kronor. If there is investment, the other party, the buyer, proposes how to split 100 kronor. The seller then chooses whether to accept the proposal, realising the proposed split, or to reject it. In the latter case both get zero, leaving the investor with a net loss of 60 kronor. To examine the role of promises and threats, we consider three different treatments. In one treatment, there is no communication except the actions themselves. In a second treatment, the buyer can send a message to the seller before the seller makes the investment decision. The third treatment allows the seller to send a message together with the investment decision. As most subjects realised, the second treatment invited promises whereas the third treatment invited threats.

Under the conventional assumptions that both agents are entirely selfish and that talk is nothing but words, the unique subgame perfect equilibrium outcome in all three treatments is that the first agent chooses not to invest. The reason is simple. At the last stage of the game, an investor should accept any proposal that yields more than 0 kronor. Hence, there is no reason for the trading partner to offer more than 1 krona, leaving the investor with a net loss of (at least) 59 kronor.

Our findings are different. Even without communication about a third of the sellers invest, and the buyers' modal offer is 80:20 rather than 0:100. That is, the buyer proposes to share the net return of $100 - 60 = 40$ equally. In most cases the seller earns a positive return on the investment and, while the average payoff for an investor is about −13 kronor, this is far from the prediction of −59. The finding is of course not entirely unexpected. Evidence from ultimatum bargaining games shows that proposals of an even split are rather common, at least in Western cultures.[3] The effect of communication is also sizeable. The investment rate goes up when one of the parties can communicate, significantly so under seller communication. However, whereas buyers' promises are always kept, sellers' threats while sometimes challenged are rarely pursued.

These results represent a challenge for economic theory. Is there any parsimonious model that can account for them? To address this question, we first investigate whether the behaviour in the no-communication treatment can be explained by a state-of-the-art social preference model due to Fehr and Schmidt (1999). To this model, that emphasises a taste for equality, we then add a taste for consistency, i.e., for keeping one's word. This way of thinking about communication in social dilemmas is not novel to us. Proponents are, among others, Braver

[2] Kerr and Kaufman-Gilliland (1994) and Sally (1995) survey experimental results along these lines.

[3] The seminal study of behaviour in ultimatum games is Güth *et al.* (1982). For an extensive recent survey, see Camerer (2003). For evidence on ultimatum bargaining games with prior production (investment), see Diekmann *et al.* (1996), Königstein and Tietz (2000) and Gantner *et al.* (1998).

(1995), Frank (1987, 1988), Hirshleifer (1987), Kerr (1995) and Ostrom *et al.* (1992).[4] An earlier formalisation is due to Klein and O'Flaherty (1993), who in turn build on Schelling (1960).[5]

Besides the effect of communication on investment, the model allows us to explain the striking difference in credibility between promises and threats. A strong preference for equity makes it difficult for the seller to pursue inequality-generating threats. By rejecting the buyer's offer in our experiment the seller generates the highly uneven outcome (−60, 0). If the buyer has offered an outcome that is more even than this, i.e., an offer of 50+ to the seller, it may take considerable resolve for an inequity averse seller to pursue the threat.[6] In contrast, inequity aversion only helps the buyer keep promises, at least if the promise is for a fair outcome.

The paper is organised as follows. In Section 1, we set up the model and analyse it under conventional economic assumptions. Section 2 presents the experimental evidence. Section 3 proposes an extended model. Section 4 concludes.

## 1. Model

Throughout this Section, we suppose that agents care only about their own monetary payoff.

We consider a seller $S$ and a buyer $B$. At stage 1, the seller can make a fixed, non-contractible investment at cost $F$. This decision is given by the indicator variable $I$, which is 1 if the seller invests and zero otherwise. At stage 2, there is a potential gain from trade $g(I)$, where

$$g = \begin{cases} G & \text{if } I = 1; \\ 0 & \text{if } I = 0. \end{cases}$$

In other words, there is a potential gain from trade if and only if the seller invests. To make the problem interesting, we assume that $G > F$.

If the seller invests, the buyer proposes how to divide the gain $G$. A proposal is a pair $x = (x_S, x_B)$ where we refer to $x_S$ as the *offer* and $x_B$ as the *demand*. We impose the restriction that. $x_B + x_S = G$. Hence, the set of feasible proposals is $X = \{x | x_B + x_S = G\}$.

The seller either agrees to the proposal or rejects it. Thus, the seller's response is denoted $a(x) \in \{Accept, Reject\}$.

[4] The idea that communication can create commitment belongs to a broader strand of psychology which emphasises people's desire for consistency; see e.g., Heider (1946), Newcomb (1953), Festinger (1957) and Cialdini (1993, Chapter 3). A competing perspective emphasises instead the notion that communication creates *identification* among subjects, i.e., it can create altruism. Proponents of this view are, among others, Kramer and Brewer (1984), Dawes *et al.* (1988) and Orbell *et al.* (1991). Kerr and Kaufmann–Gilliland (1994) is a relatively recent attempt at distinguishing between the identification hypothesis and the commitment hypothesis within the framework of a five-player game of voluntary provision of a public good. They find that the commitment hypothesis is superior.

[5] Carrillo and Dewatripont (2000) is the only recent paper (that we know of) which formally models the use of 'costly' promises, i.e., promises which lead to a loss if broken. However, their paper has a different focus, as they consider the problem of a single (but time-inconsistent) individual.

[6] By accepting 50, the seller brings about the payoffs (−10, 50), which again gives a difference of 60.

Let $\mathcal{A}$ be the set of all functions $a : X \to \{Accept, Reject\}$. The sets of pure strategies in the bargaining game are then $P_B = X$ and $P_s = \mathcal{A}$. The payoff functions in the bargaining game are

$$\pi_i = \begin{cases} x_i & \text{if } a(x) = Accept; \\ 0 & \text{if } a(x) = Reject. \end{cases} \tag{1}$$

The bargaining game has many Nash equilibria but only one of them is sub-game perfect: the buyer demands $x_B^* = G_j$ and the seller accepts this proposal. Since the seller has an incentive to accept any offer $x_S > 0$ he ends up with no surplus.

Turning to the whole game, the sets of pure strategies are $P_B = X$ and $P_s = \{0, 1\} \times \mathcal{A}$. The payoff functions are

$$u_B = \begin{cases} x_B & \text{if } I = 1 \text{ and } a(x) = Accept; \\ 0 & \text{otherwise,} \end{cases} \tag{2}$$

and

$$u_S = \begin{cases} G - x_B - F & \text{if } I = 1 \text{ and } a(x) = Accept; \\ -F & \text{if } I = 1 \text{ and } a(x) = Reject; \\ 0 & \text{if } I = 0. \end{cases} \tag{3}$$

Since the seller gets nothing out of the bargaining stage, the result is immediate.

PROPOSITION 1 *The unique subgame perfect equilibrium outcome is* ($I = 0$, $x_B = G$, $a = Accept$).

Observe also that communication in the form of cheap talk has no effect in this case.

Under the standard assumptions that agents are rational and narrowly self-interested, there is hence a strong prediction from our model: there should never be investment.

## 2. Experiments

We consider three variants of the same theme. In all treatments one agent (the seller) first decides whether to invest SEK 60 or not (SEK = Swedish kronor; in March 1999 USD 1 = SEK 8.13). If the investment is made, the agent's trading partner (the buyer) proposes how to divide a SEK 100 revenue created by the investment. (The seller/buyer terminology was not used in the experiment.) Subjects were recruited among undergraduate business and economics students at the Stockholm School of Economics. They were paid a participation fee of SEK 100 in the experiments without communication and a fee of SEK 60 in the experiments with communication. The reason for the difference in participation fee is that in the latter case we were able to perform the experiments during seminar time, though not in our own course or with students we knew, whence the participation fee was primarily required to ensure non-negative payoff. Subjects who participated during seminar time were still volunteers in the sense that each seminar (and the experiment itself) was voluntary. They were not paired with

opponents in their own classroom (seminar group). Thus social distance between opponents does not vary across treatments. Since all students belong to the same cohort of business students, we do not think that there is a significant difference in the within-classroom social distance either. Nonetheless, the slight difference in conditions means that we are somewhat more confident with respect to comparisons across the two communication treatments than with respect to the communication/no-communication comparison. Below is a description of our procedures; the instructions are reproduced in the Appendix available from the authors.

The subjects are recruited to two different rooms. In each room subjects are given a number between 1 and $N$, where $N$ is the number of students in the room, and subjects with identical numbers form a pair. Anonymity is maintained throughout. The subjects are given the instructions and protocols and are asked to read them. When all subjects have read the instructions, the instructions are read aloud by the experimenter, who also answers clarifying questions but does not answer questions regarding sensible strategies.

The experiments without communication continue as follows. At the first stage, the subjects in the 'investor room' are asked to decide whether to invest SEK 60 to create a revenue of SEK 100 for himself/herself and a trading partner in the other room. The subjects record their investment decision on a form marked 'Investment decision'. When everybody has filled in the form, it is collected by the experimenter. The experimenter then hands out the forms to the respective trading partners in the other room. At the second stage, there is bargaining over the surplus, and this stage differs across treatments.

Observe that each subject plays only a single round. We want to understand the willingness of people to make unique relationship-specific investments when contracts are inevitably incomplete. Hence, the isolation of the event was desirable. At the same time it is clear that the prospects for observing equilibrium behaviour is worse, a feature that will show up in the evidence.[7]

Bargaining proceeds as follows. The trading partner (buyer) writes down a proposed division of the revenue of SEK 100. This is done on a form marked 'Bargaining'. For pairs where no investment was made, the buyer marks the box 'No investment' on the bargaining form. The forms are then collected by the experimenters and handed out to the respective sellers. Sellers accept or reject the proposed division by marking the 'accept' or 'reject' box on the bargaining forms. The forms are then collected again by the experimenter and handed out to the respective buyers in the other room.

The subjects in both rooms record the proposed division and whether it was accepted or not in the protocol. Both subjects also estimate their revenue from the bargaining and record this in their protocols. The revenue is equal to 0 if the proposed division is rejected. If the proposed division is accepted, the revenue is equal to the proposed amount. The subjects finally estimate their earnings from

[7] Even if subjects understand the game form perfectly, i.e., the available strategies and the material payoffs, they do not necessarily know much about each others', utilities. Repeated play would allow learning about this feature of the game.

the experiment and record their earnings on the protocol. For subjects in the 'investor room' (sellers) the earnings are equal to the revenue from the bargaining minus the investment cost of SEK 60. For subjects in the other room (buyers), the earnings are equal to the revenue from the bargaining. The subjects in the experiment are paid their earnings in the experiment plus their participation fee. The experiment is then over.

### 2.1. *Statistics*

To investigate whether the proportion of investors differs between the experimental groups, we use a contingency table Pearson chi-square test.[8] The null hypothesis is no difference and we report two-sided p-values.

We also want to test whether an investor's payoff differ across experiments. Statistically, this is a tricky issue. Bargaining experiments usually lead to a highly skewed payoff distribution. For this reason, many authors have rejected standard parametric tests in favour of non-parametric tests, which do not invoke normality assumptions. For a careful discussion of this issue, see, for example, Roth *et al.* (1991). The drawback of non-parametric tests is that they do not utilise the rich cardinal information in the data. With recent advances in econometric theory and computer power, it has become possible to conduct parametric testing without imposing normality, i.e., by inferring the underlying distribution from which the data has emerged using bootstrap techniques.[9] The significance levels for the payoff comparisons that we report below have all been obtained by generating 1,099 bootstrap replications. According to Davidson and MacKinnon (2000), this number of replications is high enough to guarantee a reasonable confidence in the estimated p-values, compared to the 'ideal' bootstrap with infinitely many replications.[10] Although the bootstrap technique has been widely used recently, we are not aware of previous studies which have used it for experimental bargaining games.

### 2.2. *Aggregate Outcomes*

Table 1 reports aggregate outcomes for each of the three treatments. Table 2 reports p-values for pairwise comparisons between the three treatments.

The hypothesis that there should be no investment is clearly rejected. Moreover, the investment rate is higher in the two treatments with communication than in the no communication treatment but significantly so only in the

[8] D'Agostino *et al.* (1988) have shown that the commonly used Yates correction, as well as Fisher's exact test, are overly conservative.

[9] For an introduction to bootstrap methods, see e.g. Efron and Tibshirani (1993). Some powerful recent characterisations are reported by Davidson and MacKinnon (1999).

[10] Davidson and MacKinnon (2000) argue that at least 399 replications is needed to avoid a power loss of more than 1% for computed p-values of about 0.05, whereas at least 1,499 replications are required to avoid a similar power loss for p-values of about 0.01. (The reason for choosing numbers of replications ending with 99 is that Monte Carlo tests are exact only if $p(B + 1)$is an integer, where $p$ is the significance level and $B$ is the number of replications. The exact choice of $B$ thus matters if special significance is attached to particular p-values, such as 0.05.)

Table 1

*Investment and Profit*

| | No communication | Seller communication | Buyer communication |
|---|---|---|---|
| Number of pairs | 40 | 33 | 30 |
| Proportion investors | 0.35 | 0.64 | 0.53 |
| Mean offer | 48.57 | 63.33 | 70.00 |
| Mean profit of investor | −12.86 | −6.43 | 8.75 |

Table 2

*The Probability that Two Treatments Yield Statistically Indistinguishable Averages*

| | Proportion investors | Offer | Profit |
|---|---|---|---|
| No comm. vs Seller comm. | 0.015 | 0.125 | 0.562 |
| No comm. vs Buyer comm. | 0.125 | 0.017 | 0.033 |
| Seller comm. vs Buyer comm. | 0.407 | 0.229 | 0.048 |

seller communication treatment. The offers are highest under buyer communication; the difference between 70 (buyer communication) and 48.6 (no communication) is statistically significant, whereas the difference in average offers across the two communication treatments is not. Interestingly, the investor's mean profit is significantly higher under buyer communication than in any of the other two treatments. In order to learn more, we now look at the detailed findings.

### 2.3. *Results without Communication*

The bargaining behaviour for 14 pairs in which there was investment is depicted in Figure 1. In the Figure, a number such as $z/y$ indicates that there were $z$ proposals of this kind, of which $y$ were accepted. If there is only one number, all proposals of this kind were accepted.

On average the buyers offer SEK 48.57 to the sellers. The most common offer, made by four of the buyers, is SEK 80. These offers stand in striking contrast to the prediction that offers should be zero (or the smallest recognised money unit, which is clearly not larger than 5 in the experiment, since two offers of 75 are observed). The offer distribution is also very different from that of conventional ultimatum games, where the modal offer is typically 50 or sometimes 40, and offers above 50 are extremely rare (at least in Western cultures). It thus seems obvious that some buyers take the seller's investment cost into consideration when making their offer.

The proposed division is rejected by three out of 14 investors (21%), and these rejections are for the three lowest offers. The two rejections of 10 strongly suggest that these agents do not exclusively maximise their own monetary payoff. On average the investors lose SEK 12.86 from the investment but this loss is not significantly different from 0 ($p = 0.126$).

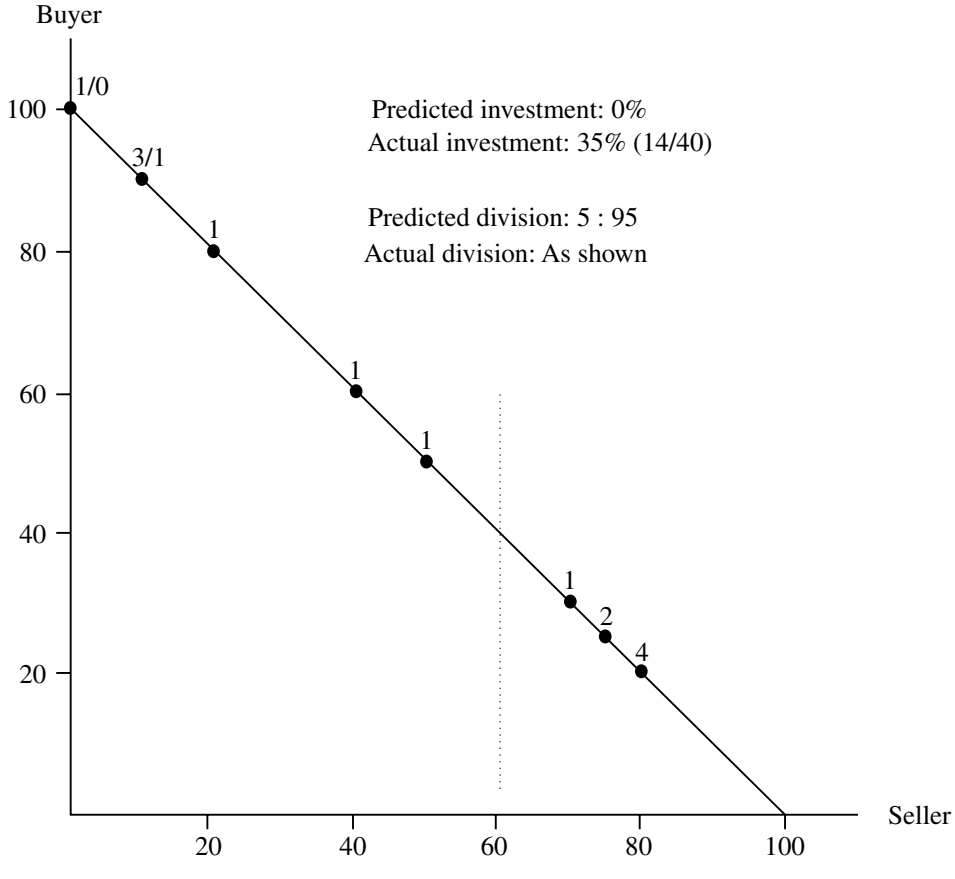


Fig. 1. *Bargaining without Communication*

2.4. *Results with Communication*

We introduced communication in the simplest possible way, namely by having one of the parties sending a written message. In order to investigate whether it matters who sends the message, we conducted one set of experiments in which the seller had the opportunity to communicate and one set of experiments in which the buyer had the opportunity to communicate. In the former case, the message was sent simultaneously with the investment decision; in the latter case, the buyer sent the message (and the seller received it) before the investment decision had been made. No restrictions were put on the content of the message, nor did we suggest what it may contain. Otherwise, the experiments were carried out exactly as in the case of no communication.

Figure 2 displays the bargaining outcome when the seller could send a message in advance. In this case, buyers' proposals vary considerably. A quarter of the offers give the investor a net loss. The two most popular offers are 60:40, which just covers the investment cost, and 80:20. Investors sometimes reject low offers, but not always. One seller rejects a 50:50 offer, one rejects a 60:40 offer and one even

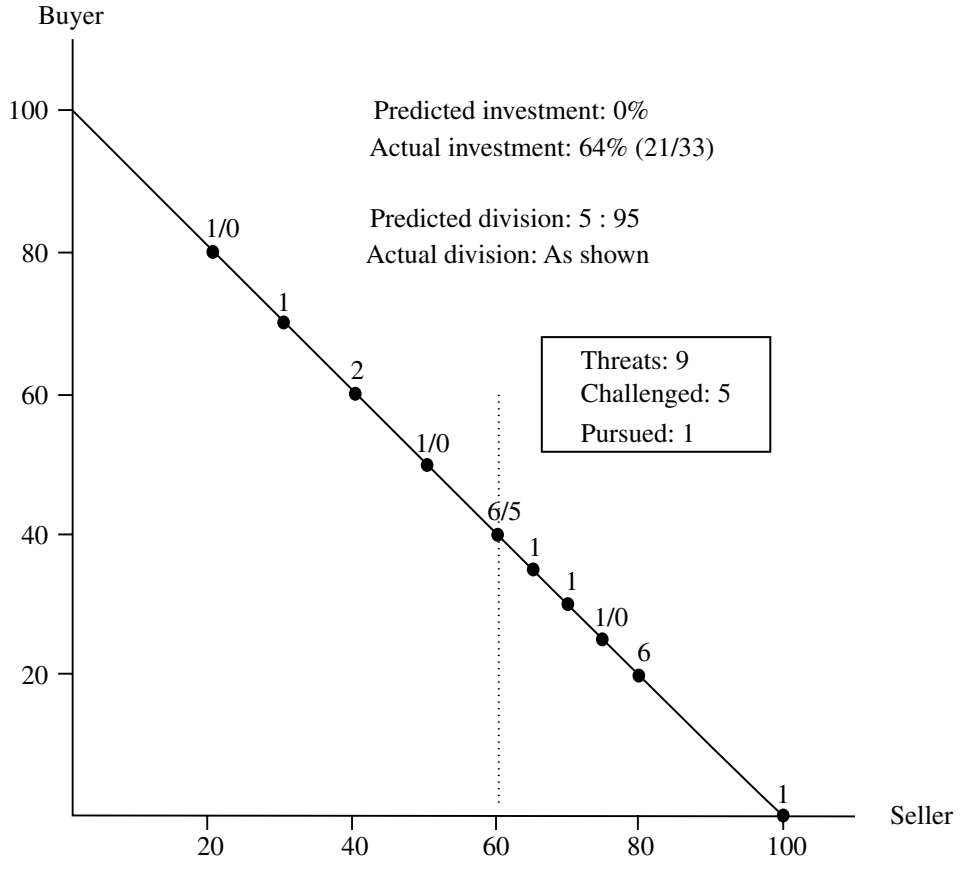


Fig. 2. *Bargaining with Seller Communication*

rejects a 75:25 offer (leaving us slightly unsure whether this subject had actually understood the offer); at the same time, two loss-making 40:60 offers and one 30:70 offer are accepted. Compared with the no-communication treatment, the data indicates both that the buyers' offers are more generous and that the sellers are more prone to reject offers in the interval [20, 75]. While the fraction of generous offers (70:30 or better), does not increase much, there is a reduction of very meagre offers and an increase in intermediate offers, notably 60:40. The latter is close to being the optimal selfish offer, given the empirical rejection behaviour.

Reading the messages provides some additional information. The first interesting observation is that nine out of the sellers' 21 messages contained explicit threats that any offer smaller than the seller's suggested split would be rejected. However, only in four of these cases did the buyer's proposal respect the threat. In the remaining five cases, the buyer made a lower offer, which the seller accepted in all but one case, where the seller rejected a 50:50 offer. The most glaring neglect of a threat was perpetrated by the buyer who successfully proposed a 30:70 split when the seller had threatened not to accept lower offers than 80:20. Hence, the experiment indicates that threats are not very credible. We observe that a majority,

16 out of 21, of the sellers' messages suggest 80:20 but in nine of these cases the buyer's offer is lower.[11]

Turning to the case in which buyers could send a message, the investment rate also goes up compared to the no communication treatment but the effect is statistically insignificant ($p = 0.125$). However, the distribution of offers is rather different, as shown in Figure 3 (and documented statistically in Tables 1 and 2).

Now, all proposals except one allows the seller to recoup his investment cost, and half of the offers are for an even split of the net surplus, i.e., 80:20. The explanation for the change in buyer behaviour is quite clear. Ten out of the sixteen investments followed explicit promises by the buyer; in seven cases the promise was 80:20, in the remaining three cases 70:30. None of these promises were violated. (The other six investments followed messages which did not make explicit promises.) Apparently, there is little difference between the messages that

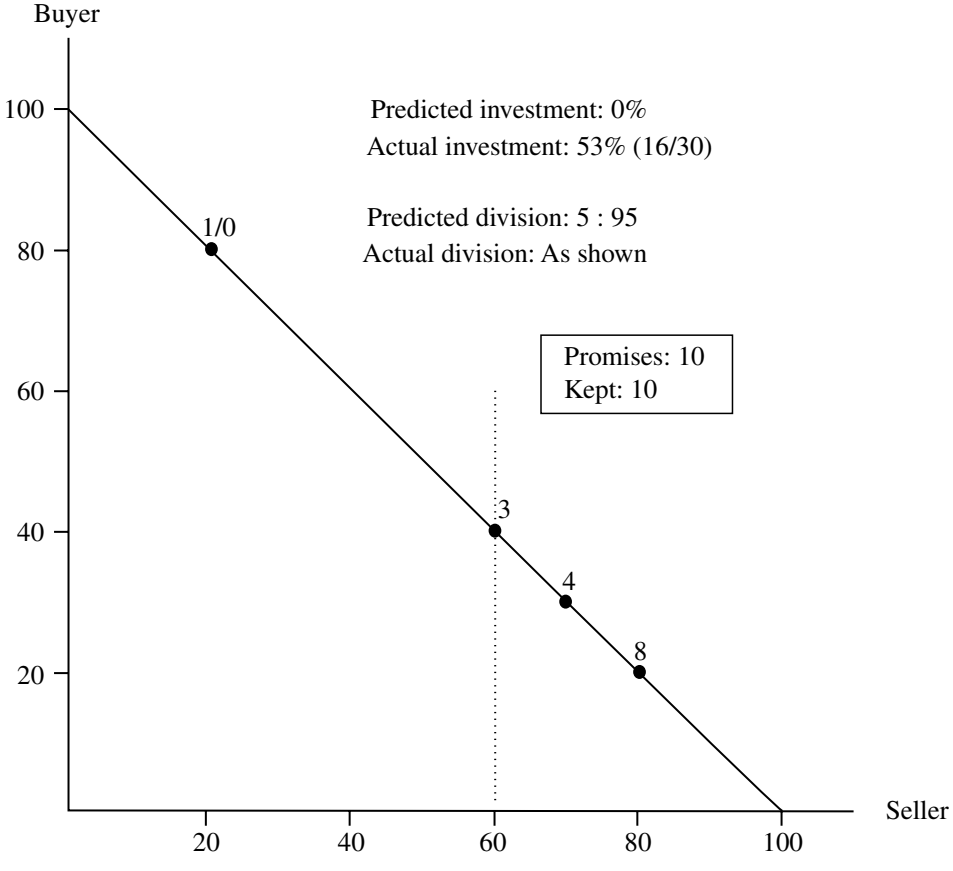


Fig. 3. *Bargaining with Buyer Communication*

[11] In two cases did the partner make a more generous offer than that suggested by the seller. In one case, the buyer proposed 100:0, giving away all his surplus, when the seller's message suggested 80:20. Another buyer rewarded a modest suggestion of 60:40 by a proposal of 65:35

attract investment and those that do not. The total number of 80:20 promises was twelve and the number of 70:30 promises was five.

As pointed out by an anonymous referee, the apparent difference in credibility between promises and threats could in principle be due to subjects' ability to detect false statements. If false promises and threats are mostly disbelieved, we might observe something like the above pattern even though the fraction of credible threats were the same as the fraction of credible promises. Since there does not seem to be much difference in the suggestions of the messages that were believed and those that were not, the referee's explanation requires that our subjects are able to pick up subtle clues about truthfulness from the messages' wording or from the senders' handwriting. We are reluctant to believe that people are that good at detecting deceptive intentions in brief written messages, and in the following we will maintain the view that promises are more credible than threats. However, a different experiment would be required to settle the issue definitely.

## 3. Behavioural Explanations

We have identified four main violations of the theory. First, some agents reject positive (and 'large') material payoffs. Second, many agents propose an equal split of the net surplus, even in circumstances when their expected surplus would almost certainly be larger under a different proposal. Third, communication increases investment. Fourth, promises are much more credible than threats. Are there alternative theories of human behaviour which can explain our findings?

One line of argument is that people simply do not play subgame perfect equilibria. Instead, they play according to some other Nash equilibrium strategy. While this argument might explain why some sellers invest and why some buyers make 'generous' offers, it does not convincingly explain why some meagre offers are in fact rejected. Faced with a choice between a positive amount and zero, the choice of a material payoff maximiser is trivial. Another objection to the 'imperfect equilibrium' view is that there are so many Nash equilibria that the theory loses all predictive power unless some other refinement is imposed.[12]

To us it seems plausible that our subjects' behaviour is best accounted for by admitting a different specification of agents' preferences, allowing for considerable differences across agents.[13] Recently, several authors have proposed formulations of preferences which can account for 'social' behaviour while preserving much of the parsimony and predictive content of earlier models.[14] Bolton and Ockenfels (2000) and Fehr and Schmidt (1999) are two leading examples of models which emphasise agents' concern for equality, and hence might potentially

[12] A candidate equilibrium refinement is stochastic stability. However, as Ellingsen and Robles (2002, Proposition 4.3) show, stochastic stability has very little cutting power in this kind of game.

[13] For further justifications of social preference models, see Camerer (2003), Fehr and Schmidt (2002) and Rabin (2002).

[14] To the extent that social preferences have descriptive power, it is a formidable challenge to understand their origin. We neglect this problem here. For a general discussion about evolutionary foundations, see Ostrom (2000). A specific foundation for inequity aversion (in the form of vengefulness) is provided by Huck and Oechssler (1999).

Table 3

*Fehr and Schmidt's Preference Distribution*

| Distribution of $\alpha$ | | Distribution of $\beta$ | |
|---|---|---|---|
| $\alpha = 0$ | 30% | $\beta = 0$ | 30% |
| $\alpha = 0.5$ | 30% | $\beta = 0.25$ | 30% |
| $\alpha = 1$ | 30% | $\beta = 0.6$ | 40% |
| $\alpha = 4$ | 10% | | |

explain the high incidence of equal splits in our experiments.[15] These two models are quite similar for two-player games but both due to its simplicity and greater predictive content we have chosen primarily to investigate the predictions from Fehr and Schmidt's model.

### 3.1. *Inequity Aversion*

In Fehr and Schmidt's model, for the case of two players, agent $i$ has utility function

$$w_i = u_i - \alpha_i \max(u_j - u_i, 0) - \beta_i \max(u_i - u_j, 0), \quad i \neq j, \tag{4}$$

where $u_i$ denotes agent $i$'s net monetary payoff and $\alpha_i$ and $\beta_i$ are non-negative parameters. A narrowly self-interested agent is given by the special case $\alpha_i = \beta_i = 0$. If $\alpha_i > 0$, agent $i$ dislikes having a lower monetary payoff than his opponent. We shall say that $\alpha_i$ measures agent $i$'s degree of *inferiority aversion*. If $\beta_i > 0$, agent $i$ dislikes having a higher payoff than his opponent; hence $\beta_i$ measures agent $i$'s degree of *superiority aversion*. Agents for whom both parameters are zero are called *selfish*.

In order to make quantitative predictions, we need to assume something about the distribution of preference parameters in the population. Fehr and Schmidt (1999) argue that earlier experimental evidence can be used to calibrate their model, and that the parameter distribution described in Table 3 matches available evidence quite well.

The Table indicates large preference heterogeneity, with sizeable fractions of both selfish and strongly inequity averse agents. Overall, inferiority aversion is somewhat stronger than superiority aversion. Indeed, Fehr and Schmidt assume that $\alpha_i \geq \beta_i$ for all agents but the implied correlation between the two parameters is going to be irrelevant for our purposes.

Since agents are heterogeneous, there should be incomplete information about preferences. Thus, the standard solution concept is perfect Bayesian equilibria. While we shall apply this concept here, we doubt very much that agents have identical beliefs about the distribution of preferences.

[15] Other authors, such as Rabin (1993), have argued that players care not only about payoffs *per se* but also about the intentions of their opponents. It is by now also well understood that the Fehr/Schmidt model works poorly on some domains; see for example Charness and Rabin (2002). Still, we are not aware of any model that has done better in organising the data from bargaining experiments.

3.2. *Qualitative Analysis*

Let us first note some general forces that operate in this model. The first observation is that sunk costs potentially affect behaviour at the bargaining stage if: (*i*) these costs enter the agent's utility functions at that stage, *and* (*ii*) at least one agent is inequity averse. Conversely, if agents forget about sunk costs when considering their bargaining strategy, *or* it is common knowledge that no agent is inequity averse, then sunk costs are irrelevant. We shall make the assumption that agents include sunk costs in their computations.

A second general point is that inequity aversion narrows the set of strategies that agents are willing to consider. There are two different forces at work. First, there is a direct effect of inferiority aversion. A seller who has made an investment is only willing to accept a split, $x = \{(x_B, x_S) | x_B + x_S = G\}$ if it yields a higher utility than rejection, i.e. if

$$x_S - \alpha_S[G - x_S - (x_S - F)] \geq -F\alpha_S. \tag{5}$$

Letting $X_S(\alpha_S)$ denote the *smallest* amount that is acceptable to the seller we have from the above inequality that

$$X_S = \frac{\alpha_S G}{1 + 2\alpha_S}. \tag{6}$$

Interestingly, this acceptance threshold is independent of the investment cost $F$. The reason is that preferences are linear in inequity, which implies that a change in $F$ affects both the fair offer and the loss associated with rejection equally.

The seller's inferiority aversion limits the scope for buyer opportunism. Superiority aversion could also exert a moderating influence on the buyer's demands. To illustrate the latter effect, consider the problem of a buyer who considers demanding more than half the net surplus. Let $p_B(x_B)$ denote the buyer's subjective probability that a demand $x_B$ is successful. The buyer's problem is to choose $x_B \geq (G - F)/2$ to maximise

$$\begin{aligned} &p_B(x_B)\{x_B - \beta_B[x_B - (G - F - x_B)]\} + [1 - p_B(x_B)](-\beta_B F) \\ &\quad = p_B(x_B)[(1 - 2\beta_B)x_B + \beta_B G] - \beta_B F. \end{aligned} \tag{7}$$

Since $p_B(x_B)$ will be a non-increasing function taking values in the interval [0, 1], the buyer prefers the equal split $x_B = (G - F)/2$ if $\beta_B \geq 1/2$.

We see immediately that superiority aversion could have beneficial effects on investment.

PROPOSITION 2 *Suppose the buyer's superiority aversion is known to the seller.*

(*i*) *If* $\beta_B > 1/2$, *the unique subgame perfect equilibrium outcome is* $(I = 1, x_B = (G - F)/2, x_S = (G + F)/2)$.

(*ii*) *If* $\beta_B < 1/2$, *there is no investment*; $I = 0$.

The point of part (*i*) is that a strongly inequity averse buyer has a dominant strategy at the bargaining stage, namely to offer an equal split of the net surplus and this is clearly enough to make the seller invest. This result contrasts sharply

with the case of purely selfish preferences, when investment is not sustainable. Could inferiority aversion also, by itself, motivate investment? (The idea would be that a seller who is committed to rejecting low offers could extract a sufficiently large offer to cover the investment cost.) The answer, as stated in part (*ii*) of Proposition 2 is no. While it is true that a large $\alpha_S$ would persuade the buyer that a high $x_S$ is required, the lowest $x_S$ that would be acceptable *ex post* is smaller than the lowest $x_S$ that is acceptable *ex ante*. The former is the solution to the equation

$$x_S - \alpha_S[G - x_S - (x_S - F)] = \alpha(0 - F),$$

whereas the latter solves

$$x_S - F - \alpha_S[G - x_S - (x_S - F)] = 0.$$

Thus, the seller only invests if the expected offer is at least

$$\underline{x}_S^A = \frac{F + (F + G)\alpha_S}{1 + 2\alpha_S}, \tag{8}$$

which is greater than $\underline{x}_S$: as long as $\beta_B$ is below $1/2$, we see from (7) that the buyer would not make an offer higher than the lowest offer that is accepted for sure. Thus, we conclude that the implicit threat embodied in a seller's aversion to being badly treated is not *by itself* enough to sustain investment.

Let $b_0$ denote the share of buyers with $\beta_B = 0$. By the previous argument, we know that no seller can invest in equilibrium if $b_0 = 1$. More generally, there is a number $\bar{b}_0(\alpha) < 1$ such that a seller with inequity aversion $\alpha$ will not invest for any $b_0 > \bar{b}_0$.

PROPOSITION 3 *A necessary condition for investment in equilibrium is* $b_0 \leq (G - F)/(G + F)$.

To prove this result, let $\bar{\alpha}$ denote the highest degree of inferiority aversion that any investing seller has in equilibrium. Observe that in any equilibrium with investment by all sellers with $\alpha_S \leq \bar{\alpha}$, a selfish buyer will offer the seller at most

$$\underline{x}_S(\bar{\alpha}) = \frac{\bar{\alpha}G}{1 + 2\bar{\alpha}}.$$

Inequity averse buyers, as we have seen above, offer the seller at most $(F + G)/2$. Hence, an $\bar{\alpha}$-seller is only willing to invest if

$$b_0 \frac{\bar{\alpha}G}{1 + 2\bar{\alpha}} + (1 - b_0)\frac{F + G}{2} - b_0\bar{\alpha}\left[G - \frac{\bar{\alpha}G}{1 + 2\bar{\alpha}} - \left(\frac{\bar{\alpha}G}{1 + 2\bar{\alpha}} - F\right)\right] - F \geq 0.$$

Simplifying, we have

$$b_0 \leq \frac{G - F}{G + (1 + 2\bar{\alpha})F}.$$

Since the expression is decreasing in $\bar{\alpha}$, the result follows from setting $\bar{\alpha} = 0$.

We are now ready to derive the model's quantitative predictions. Recall that $F = 60$ and $G = 100$ throughout.

### 3.3. *Predictions for the No Communication Treatment*

Can the model explain why anyone would ever invest in the absence of credible communication? Looking at the necessary condition that we derived in Proposition 3, we see that it is violated for the Fehr/Schmidt parameters. If the fraction of selfish buyers is above 1/4, there should never be investment in equilibrium. If anything, we think our subject pool of students is more selfish than the Fehr/Schmidt average (this judgement is based on standard ultimatum game experiments that we report elsewhere). However, for the reasons given in Section 3.1, we think that this test of the model is too stringent. Given the relatively modest average loss from investment, we also suspect that some investment in the no-communication condition can be explained either by the 'joy of playing' or by simple decision error.[16] However, this remains an open question.

Conditional on investment, the model makes one clear prediction regarding bargaining outcomes, namely that the buyer offers $(F + G)/2$ to the seller in about 40% of the cases and that this offer is always accepted. (Further conditional predictions concerning buyer behaviour depend on what buyers are assumed to believe about the investing sellers.) This prediction fares quite well. Almost 30% of buyers offer an equal split of 80:20 and offers of 70:30 or better are almost 50%. As expected, these offers are never rejected. In fact, the model makes the fairly strong prediction that a seller never rejects an offer of 50 or above, as can be seen by setting $G = 100$ in (6). This prediction is strong in our view, because 50:50 is a rather unfair outcome. The crux is that the outcome triggered by rejection is just as unfair. Consistent with the prediction, none of the five offers between 50 and 75 are rejected.[17]

More than 30% of the offers are very meagre, at 20:80 or less, but most of these (3/5) are rejected. The buyer behaviour is consistent with at least some 30% of the whole population being rather selfish. On the other hand, the rejection behaviour indicates that the completely selfish fraction is not a whole lot larger than 30%, although these numbers are too small to allow strong inferences.

### 3.4. *Credible Communication*

The model presented above cannot explain why communication matters so much for investment and bargaining behaviour. Moreover, it seems unlikely that brief written messages can be very revealing about the sender's distributional preferences. Hence, we maintain the assumption that preferences cannot be communicated. Instead, following the psychological literature cited in the introduction, we believe that communication creates commitment.

[16] It is well known that 'quantal response equilibrium' (which admits decision error) is a better solution concept than Nash equilibrium in many settings; see Goeree and Holt (2000, 2001).

[17] As pointed out by an anonymous referee, the model predicts that buyers should not make any offers in the open interval (50,80). However, this prediction is an artifact of the linear utility function.

We extend the model in a minimal way by introducing a personal cost of being inconsistent. The parameter is called $l$, since the cost is only incurred if a person is caught lying.

If our subjects were selfish, costs of lying would have a straightforward effect: the seller can credibly threaten to reject any offer below $l$ and the buyer can credibly promise to offer any amount up to $l$. Thus, in this case, costs of lying would not affect the relative efficiency of promises and threats: they have identical commitment potential. Moreover, the psychological cost of lying needs to exceed the investment cost, $F$, to induce investment.

Suppose now that agents are inequity averse. In this case, buyer promises and seller threats turn out to play more subtle roles, and they are no longer interchangeable.

3.4.1. *Complete information*

To begin with, we make the assumption that all the parameters of the utility functions are common knowledge. Note that $\underline{x}_S$, as defined in (6) is the maximum offer which the seller would reject if there is no communication. For any buyer with $\beta_B < 1/2$ the offer $\underline{x}_S$ is the optimal offer when $\alpha_S$ is known and there is no communication. The buyer's utility from *breaking* a promise (and offer $\underline{x}_S$) is thus [18]

$$G - \underline{x}_S - l - \beta_B[G - x_S - (\underline{x}_S - F)].$$

Of course, it is never worthwhile offering more than the fair share, $F + (G - F)/2$, so we confine attention to promises that are weakly lower than this. The buyer's utility from *keeping* a promise $x_S^P \leq F + (G - F)/2$ is

$$G - x_S^P - \beta_B[G - x_S^P - (x_S^P - F)].$$

Thus, the largest credible promise is given by

$$\bar{x}_S^P = \min[F + (G - F)/2, \underline{x}_S + l/(1 - 2\beta_B)]. \tag{9}$$

The interesting property of this result is the way that the cost of breaking promises gets magnified by the promisor's inequity aversion. If $\beta_B$ is close enough to $1/2$, even a minuscule $l$ is enough to make a fair offer credible. Intuitively, the reason is that the trading partner's utility from an extra unit of money is dampened by superiority aversion. An unfair dollar is worth less than a fair dollar.

Let us now turn to seller threats. We consider only threats to reject unfair offers, i.e., offers below $F + (G - F)/2$. The maximal offer that the seller can credibly threaten to reject, call it $\bar{x}_S^T$, is then given by the equation

$$\bar{x}_S^T - \alpha_S[G - \bar{x}_S^T - (\bar{x}_S^T - F)] - l = -\alpha_S F. \tag{10}$$

[18] As a referee has pointed out it is not absolutely obvious that the costs of lying should be excluded from the payoff comparison over which agents are inequity averse. However, we believe that agents keep monetary and psychological costs separate and care primarily about monetary equity. This belief is based in part on the findings by Borges and Knetsch (1997), that fairness norms change even when monetary opportunity costs replace realised monetary costs.

This is essentially the same condition as (5), except $l$ is deducted on the left hand side. By accepting an offer lower than $\bar{x}_S^T$, the seller gets the offered amount but suffers because of inequity aversion as well as from breaking his word by not executing the threat. The seller can therefore credibly threaten to reject any offer smaller than

$$\bar{x}_S^T = \min[F + (G - F)/2, \underline{x}_S + l/(1 + 2\alpha_S)], \tag{11}$$

where $\underline{x}_S$ is defined by (6). When the fair division cannot be sustained, the maximum credible threat is hence

$$\bar{x}_S^T(l) = \frac{l + \alpha_S G}{1 + 2\alpha_S}. \tag{12}$$

Observe that the effect of $l$ on the maximal credible threat is decreasing in $\alpha_S$. Thus, the seller's inequity aversion diminishes the potential for threats. Again the intuition is quite clear. The investor's utility of an extra dollar is magnified by inferiority aversion. Thus, while the threat can be used to lift the minimum acceptable utility by $l$ units, the minimum acceptable monetary offer is lifted by less. Even if it is very costly to back down from a threat ($l$ is large), it may be more costly to pursue it. In the limit, as the seller becomes infinitely inequity averse, we have

$$\lim_{\alpha_S \to \infty} \bar{x}_S^T = \lim_{\alpha_S \to \infty} \underline{x}_S = G/2.$$

In other words, the infinitely inequity averse seller finds it impossible to reject an even split of the *gross* surplus. In our example, the seller would like to be able to reject all offers below 80 (the fair offer) but, since the rejection leads to such an uneven split, it may be impossible to pursue a threat to reject more than 50.

We summarise the discussion as follows.

PROPOSITION 4 *Inequity aversion should make buyers' promises more credible and sellers' threats less credible.*

Of course, this result depends on the setting that we consider. Promises are credible in our model because they are given by the relatively strong party, who values extra money less. Threats have lower credibility because they are made by a relatively weak party, who values extra money more.[19] An interesting question for further research would be to investigate the credibility of promises and threats when promises are given by the relatively disadvantaged party (say, a hostage), and threats by the advantaged party (say, a kidnapper).

3.4.2. *Incomplete information*

In our experiments interaction was anonymous, so it would be impossible to observe the opponent's inequity aversion. However, the distinction between promises

[19] We are grateful to an anonymous referee for suggesting this interpretation.

and threats remains even if these preference parameters are unobservable. As an illustration, consider the experiment's values of $F = 60$, $G = 100$ and suppose that $\alpha$ is distributed as in Table 3. For computational simplicity let $\beta = 0$ for all. Under buyer promises, there is then a perfect Bayesian equilibrium in which *all* sellers invest for any $l \geq 40$.[20] Under seller threats, it is instead necessary that $l \geq 60$ in order to have *any* investment. (The promise scenario works as follows. Suppose the buyer promises to make a fair offer $x_S = 80$. If the seller believes the promise, there is always investment. The buyer keeps the promise as long as $100 - 80 \geq \max_{\alpha_S}[100 - \underline{x}_S(\alpha_S)]$ Prob $(\alpha \leq \alpha_S) - l$. When $l = 40$, this condition holds for $\alpha_S = l$, $\underline{x}_S(\alpha_S) = 33.33$ and Prob$(\alpha \leq \alpha_S) = 0.9$. The threat scenario is even simpler. No seller will invest if the offer is believed to be below 60. Given the assumption that $\beta_B = 0$, no buyer will offer more than $\bar{x}_S^T = (l + 100\alpha_S)/(1 + 2\alpha_S)$, the maximum credible threat. If $l < 50$, then $\bar{x}_S^T < 50$ for all $\alpha$, and there will not be investment. If $l < 50$, then $\bar{x}_S^T$ is decreasing in $\alpha$, so in that case the buyer's offer will never be above $l$. Thus, there cannot be investment for $l < 60$.)

### 3.5. *Fitting the Evidence?*

It is already quite clear that the model cannot fit the evidence perfectly, because there is an inconsistency between investment rates and the profitability of investment: there should not have been any investment under no communication, and there should (probably) always have been investment following a promise of 80:20. A likely explanation is that expectations about others' behaviour are not entirely correct.[21]

Nonetheless, observed behaviour could potentially falsify the model, because many decisions are predicted to be insensitive to expectations. The seller's acceptance or rejection of bargaining proposals is a case in point. We have noted above that the pattern of acceptance under the no-communication condition is roughly in line with the Fehr/Schmidt distribution of $\alpha$. Likewise, the buyers who offer 80:20 under no communication are predicted to do so regardless of their expectation, and this fraction was consistent with the Fehr/Schmidt distribution of $\beta$.

In order to illustrate a bit more rigorously how well or poorly the model fits the data, we carry out the following estimation exercise. First, we estimate a value for $l$ using the difference in offers between the buyer communication session and the no communication session, maintaining the Fehr/Schmidt assumptions concerning the distribution of the inequity aversion parameters. Using the estimated value of $l$, we then estimate the distribution of $\alpha$ from the sellers' decisions to accept or reject and the distribution of $\beta$ from the buyers' proposals. In order for this procedure to be valid we must make some additional assumptions. These are mentioned along the way.

[20] For some values of $l$ there are also other equilibria.

[21] Manski (2002) emphasises the need for more careful treatment of expectations in experimental work, proposing that elicitation of expectations from subjects would be particularly valuable. We can only agree and pledge to do better on this account in the future.

Suppose that all buyers can be thought of as having made a promise (we know that this is not strictly true but not much is gained from making the distinction), and that buyers assume that the distribution of $\alpha$ among investing sellers is the same in the buyer communication treatment as in the no communication treatment. If the offer would have been $x_S$ without communication, it should increase by $\min[80 - x_S, l/(1 - 2\beta_B)]$ for all $\beta_B < 1/2$ and be unaffected otherwise. Using the parameter distribution of Table 3, the average increase in offers should be $0.9\, l$. Since we estimated the difference across the two treatments as 21.43, our estimate of $l$ is 23.81 (the 95% confidence interval is 3.66–43.96).[22]

The acceptance/rejection decision of each seller, conditional on the offer $x_S$, gives information about whether his or her $\alpha$ is above (reject) or below (accept) the threshold value, call it $\underline{\alpha}(x_S)$. From (6), the threshold value is $x_S/(G - 2x_S)$ in the no communication case and in the buyer communication case. Under seller communication the threshold value is similarly derived as $(x_S - l)/(G - 2x_S)$. To be able to use observations from all three treatments in the same regression, we therefore adjust offers in the interval [0, 50) from the seller communication session, coding them as $\max[0,(x_S - l)G/(G - 2l)]$.[23] We then estimate a logistic regression for the acceptance probability as a function of the proposal. The implied estimate of the cumulative distribution of $\alpha$ and the associated 95% confidence interval are shown in Figure 4.

Evidently, the estimate is quite well in line with the Fehr/Schmidt distribution in Table 3. However, the precision is low. In particular, the theory fails to explain why offers above 50 were rejected in the seller communication treatment. The model only admits such rejections if $l$ is larger than 50. If $l$ is very heterogeneous, our assumption that it is homogeneous leads to inefficient estimates of the $\alpha$-distribution. (If we eliminate $l$ by using only the no communication and the buyer communication treatments, the tails of the estimated distribution become thinner; only 2% of subjects are estimated to have $\alpha \geq 4$ compared to 20% in Figure 4. Precision gets very low.)

The buyers' proposals in the no communication and the seller communication treatment can now be used to estimate the distribution of $\beta$. The distribution is truncated at $\beta = 0.5$, because anyone with a higher degree of superiority aversion will offer 80. (By the same token, the proportion of offers that are weakly above 80 is a good estimate of the fraction of subjects with $\beta \geq 0.5$.)

Assuming that buyers' have rational expectations about seller behaviour, we can derive optimal demands, $x_B(\beta)$. Taking the first-order condition associated with (7), we see that the optimal demand if $\beta = 0$ solves $p(x_B) + p'(x_B)x_B = 0$. Any demand at this level or greedier is assigned $\beta = 0$. Less greedy demands translate into $\beta$ values according to the formula

[22] We could also estimate $l$ from the difference in offers between the the seller communication treatment and the no communication treatment but the procedure requires even stronger assumptions.

[23] This is the solution to the equation $y/(G - 2y) = (x_S - 1)/(G - 2x_S)$ (subject to the constraint of non-negative offers). Offers above 50 are not adjusted, because with $l = 24$ all such offers should be accepted regardless of $\alpha$.

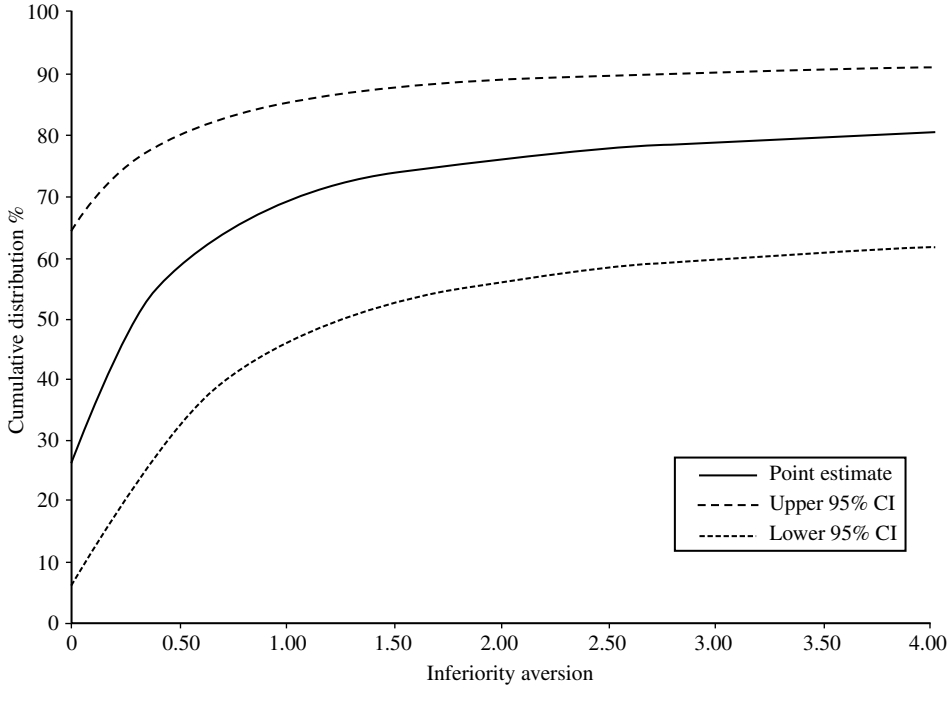


Fig. 4. *The Distribution of Inferiority Aversion* ($\alpha$)

$$\beta(x_B, t) = \frac{p_t(x_B) + p_t'(x_B)x_B}{2p_t(x_B) + p_t'(x_B)(2x_B - G)}, \tag{13}$$

where the subscript $t$ indexes treatments. Estimating the acceptance probabilities $p_t(x_B)$ for each treatment $t$, the distribution of $\beta$ follows directly and is depicted in Figure 5.[24] About 20% of our subjects appear not to have any superiority aversion ($\beta = 0$) and about 30% are strongly superiority averse ($\beta \geq 0.5$). While the precision of these estimates are better than for $\alpha$, our strong assumptions imply that all numbers should be interpreted cautiously.

In conclusion, we think that the model explains some effects of promises and threats on bargaining and investment behaviour. However, the discussion is quite speculative. Much more data, preferably with different stakes, would be needed to estimate a distribution of the preference parameter $l$, and to see whether this simple model is (or can be made) 'scalable'; is there a fixed cost of lying, or is the cost related to the monetary stakes? Also, different experiments would have to be carried out to check whether a simple desire for consistency is all that is at work. Perhaps sellers would be angry if promises were broken, or people feel a stronger urge to keep promises than to pursue threats?

[24] Due to the low number of observations, we pooled the no communication and the buyer communication treatments when estimating acceptance probabilities. (According to our theory, seller behaviour in these two treatments should be the same.)

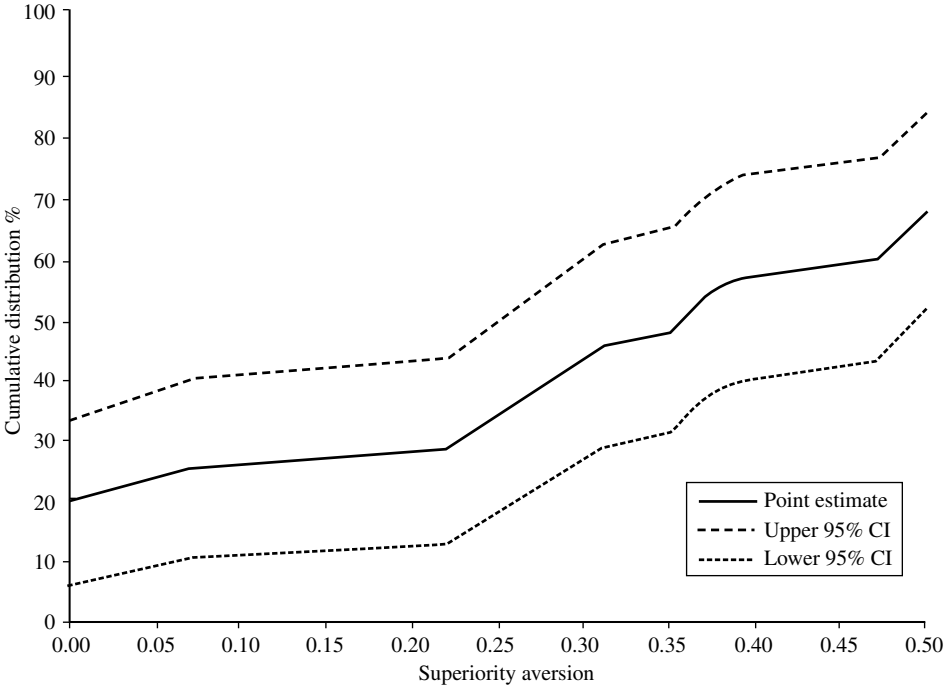


Fig. 5. *The Distribution of Superiority Aversion* ($\beta$)

## 4. Final Remarks

We have documented anew that people are less opportunistic than economists regularly assume. As found by many researchers before us, there are indications that people care both about fairness and consistency. A novel finding is that promises appear to be more credible than threats in the setting that we consider.

Building on existing theories, we have proposed a simple model that can fit many of the regularities that we observe. An implication of the model is that concerns for fairness and consistency are not orthogonal. Fairmindedness strengthens the credibility of promises and weakens the credibility of threats.

The model offers an explanation for why it is that communication is often found to mitigate social dilemmas. The meta-analysis of Sally (1995) considers 130 distinct experiments from 37 different studies. Of all the factors which according to the 'economic' model should not matter, communication is by far the most important, vastly raising cooperation rates. Almost all these studies concern dilemmas in which the relevant commitments are promises rather than threats. There is no scope for statements of the form: 'if you do not cooperate, I will punish, even if punishment is costly to myself.' Because an agent benefits from withholding the own private contribution to the public good, he can only give a promise. In contrast, we have studied an asymmetric situation, in which the seller can only threaten and the buyer can only promise.

Many open questions remain. How sensitive are our findings to the stakes involved and to the choice of subject pool? Are there large cultural differences? How

would behaviour be affected if interaction were oral and face-to-face rather than written and anonymous? What is the impact of information conditions (complete information about all benefits and costs) and of the nature of the investment decision (monetary)?

One reason for pursuing this line of research further is that the hold-up problem is at the very core of organisation theory and institutional economics. The standard analysis, due to Grossman and Hart (1986), Grout (1984) and Tirole (1986), assumes that agents are selfish and that they are completely unable to commit to future actions. Both assumptions square badly with our findings. Indeed, our work belongs to a long line of experiments that suggest an important role for fairness in mitigating hold-up problems, including Gantner *et al.* (1998), Hackett (1994), Oosterbeek *et al.* (1999) and Sonnemans *et al.* (2001).

Theoretical work by Ellingsen and Robles (2002) and Tröger (2002) has attempted to explain the limited opportunism in hold-up experiments without resorting to social preferences. Instead, they use evolutionary game theory and exploit the fact that many bargaining games have multiple Nash equilibria. They find that evolutionary selection criteria can both make very specific predictions when there are multiple subgame perfect equilibria and admit a whole host of outcomes when there is only one subgame perfect equilibrium. However, evolutionary models fail to predict the prevalence of equal splits in hold-up experiments.

*Stockholm School of Economics*

*Date of receipt of first submission: July 2001*
*Date of receipt of final typescript: May 2003*

## References

D'Agostino, R.B., Chase W. and Belanger, A. (1988). 'The appropriateness of some common procedures for testing the equality of two independent binomial populations', *American Statistician*, vol. 42 (3), pp. 198–202.

Bolton, G.E. and Ockenfels, A. (2000). 'ERC: a theory of equity, reciprocity, and competition', *American Economic Review*, vol. 90(1), pp. 166–93.

Borges, B.F.J. and Knetsch, J.L. (1997). 'Valuation of gains and losses, fairness and negotiation outcomes', *International Journal of Social Economics*, vol. 24 (1–3), pp. 265–81.

Braver, S.L. (1995). 'Social contracts and the provision of public goods', in (D. Schroeder, ed.), *Social Dilemmas: Perspectives on Individuals and Groups*, New York: Praeger.

Camerer, C.F. (2003). *Behavioral Game Theory: Experiments in Strategic Interaction*, Princeton: Princeton University Press.

Carrillo, J.D. and Dewatripont, M. (2000). 'Promises, promises,…', manuscript, ECARES, Bruxelles.

Charness, G. and Rabin, M. (2002). 'Understanding social preferences with simple tests', *Quarterly Journal of Economics*, vol. 117 (3), pp. 817–69.

Cialdini, R.B. (1993). *Influence: Science and Practice,* 3rd edition, New York: Harper Collins.

Davidson, R. and MacKinnon, J.G. (1999). 'The size distortion of bootstrap tests', *Econometric Theory*, vol. 15 (3), pp. 361–76.

Davidson, R. and MacKinnon, J.G. (2000). 'Bootstrap tests: how many bootstraps?', *Econometric Reviews*, vol. 19 (1), pp. 55–68.

Dawes, R.M., Orbell, J. and van de Kragt, A.J. (1988). 'Not me or thee but we: the importance of group identity in eliciting cooperation in dilemma situations', *Acta Psychologica*, vol. 68 (1), pp. 83–97.

Diekmann, K., *et al.* (1996). 'The descriptive and prescriptive use of previous purchase price in negotiations', *Organizational Behavior and Human Decision Processes*, vol. 66 (2), pp. 179–91.

Efron, B. and Tibshirani, R.J. (1993). *An Introduction to the Bootstrap, Monographs on Statistics and Applied Probability*, No 57, New York: Chapman and Hall.
Ellingsen, T. (1997). 'The evolution of bargaining behavior', *Quarterly Journal of Economics*, vol. 112 (2), pp. 581–602.
Ellingsen, T. and Robles, J. (2002). 'Does evolution solve the hold-up problem?', *Games and Economic Behavior*, vol. 39 (1), pp. 28–53.
Fehr, E. and Schmidt, K.M. (1999). 'A theory of fairness, competition, and cooperation', *Quarterly Journal of Economics*, vol. 114 (3), pp. 817–68.
Fehr, E. and Schmidt, K.M. (2002). 'Theories of fairness and reciprocity – evidence and economic applications', in (M. Dewatripont, L.P. Hansen and S. Turnovski, eds.), *Advances in Economic Theory, 8th World Congress of the Econometric Society*, Cambridge: Cambridge University Press.
Festinger, L. (1957). *A Theory of Cognitive Dissonance*, Stanford: Stanford University Press.
Frank, R.H. (1987). 'If homo economicus could choose his own utility function, would he want one with a conscience?', *American Economic Review*, vol. 77 (3), pp. 593–604.
Frank, R.H. (1988). *Passions within Reason*, New York: W.W. Norton.
Gantner, A., Güth, W. and Königstein, M. (1998). 'Equity anchoring in simple bargaining games with production', Discussion paper 128, Department of Economics, Humboldt University.
Goeree, J.K. and Holt, C.A. (2000). 'Asymmetric inequality aversion and noisy behaviour in alternating–offer bargaining games', *European Economic Review*, vol. 44 (4–6), pp. 1079–89.
Goeree, J.K. and Holt, C.A. (2001). 'Ten little treasures of game theory and ten intuitive contradictions', *American Economic Review*, vol. 91 (5), pp. 1402–22.
Grossman, S.J. and Hart, O.D. (1986). 'The costs and benefits of ownership: a theory of vertical and lateral integration', *Journal of Political Economy*, vol. 94 (4), pp. 691–719.
Grout, P. (1984). 'Investment and wages in the absence of a binding contract', *Econometrica*, vol. 52 (2), pp. 449–60.
Güth, W., Schmittberger, R. and Schwarze, B. (1982). 'An experimental analysis of ultimatum bargaining', *Journal of Economic Behaviour and Organization*, vol. 3 (4), pp. 367–88.
Hackett, S.C. (1994). 'Is relational exchange possible in the absence of reputations and repeated contact?', *Journal of Law, Economics and Organization*, vol. 10 (4), pp. 360–89.
Heider, F. (1946). 'Attitudes and cognitive organization', *Journal of Psychology*, vol. 21, pp. 107–12.
Hirshleifer, J. (1987). 'On the emotions as guarantors of threats and promises', in (J. Dupré, ed.), *The Latest on the Best: Essays in Evolution and Optimality*, Cambridge MA: MIT Press.
Huck, S. and Oechssler, J. (1999). 'The indirect evolutionary approach to explaining fair allocations', *Games and Economic Behavior*, vol. 28 (1), pp. 13–24.
Kerr, N.L. (1995). 'Norms in social dilemmas', in (D. Schroeder, ed.), *Social Dilemmas: Perspectives on Individuals and Groups*, New York: Praeger.
Kerr, N.L. and Kaufmann-Gilliland, C.M. (1994). 'Communication, commitment, and cooperation in social dilemmas', *Journal of Personality and Social Psychology*, vol. 66 (3), pp. 513–29.
Klein, D.B. and O'Flaherty, B. (1993). 'A game-theoretic rendering of promises and threats', *Journal of Economic Behavior and Organization*, vol. 21 (3), pp. 295–314.
Kramer, R.M. and Brewer, M.B. (1984). 'Effects of group identity on resource use in a simulated commons dilemma', *Journal of Personality and Social Psychology*, vol. 46, pp. 1044–57.
Königstein, M. and Tietz, R. (2000). 'Profit sharing in an asymmetric bargaining game', in (M. Königstein, ed.), *Equity, Efficiency and Evolutionary Stability in Bargaining Games with Joint Production*, pp. 5–32. Lecture Notes in Economics and Mathematical Systems, vol. 483, Berlin, Heidelberg: Springer-Verlag.
Loomis, J.L. (1959). 'Communication, the development of trust, and co-operative behaviour', *Human Relations*, vol. 12, pp. 305–15.
Manski, C.F. (2002). 'Identification of decision rules in experiments on simple games of proposal and response', *European Economic Review*, vol. 46 (4/5), pp. 880–91.
Newcomb, T. (1953). 'An approach to the study of communicative acts', *Psychological Review*, vol. 60, pp. 393–404.
Nietzsche, F. (1887). *On the Genealogy of Morals*, Oxford: Oxford University Press.
Orbell, J.M., van de Kragt, A.J. and Dawes, R.M. (1991). 'Covenants without the sword: the role of promises in social dilemma situations', in (K. Koford and J. Miller, eds.), *Social Norms and Economic Institutions*, Ann Arbor: University of Michigan Press.
Oosterbeek, H., Sonnemans, J. and van Velzen, S. (1999). 'Bargaining with endogenous pie size and disagreement points: a holdup experiment', mimeo., University of Amsterdam.
Ostrom, E., Walker J.M. and Gardner, R. (1992). 'Covenants with and without the sword: self-governance is possible', *American Political Science Review*, vol. 86 (2), pp. 404–17.
Ostrom, E. (2000). 'Collective action and the evolution of social norms', *Journal of Economic Perspectives*, vol. 14 (3), pp. 137–58.

Rabin, M. (1993). 'Incorporating fairness into game theory and economics', *American Economic Review*, vol. 83 (5), pp. 1281–302.
Rabin, M. (2002). 'A perspective on psychology and economics', *European Economic Review*, vol. 46 (4/5), pp. 657–85.
Roth, A.E., Prasnikar,V., Okuno-Fujiwara, M. and Zamir, S. (1991). 'Bargaining and market behaviour in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: an experimental study', *American Economic Review*, vol. 81 (5), pp. 1068–95.
Sally, D.F. (1995). 'Conversation and cooperation in social dilemmas: experimental evidence from 1958 to 1992', *Rationality and Society*, vol. 7 (1), pp. 58–92.
Schelling, T.C. (1960). *The Strategy of Conflict*, Cambridge MA: Harvard University Press.
Sonnemans, J., Sloof, R. and Oosterbeek, H. (2001). 'On the relation between asset ownership and specific investments', Economic Journal, vol. 111 (474), pp. 791–820.
Tirole, J. (1986). 'Procurement and renegotiation', *Journal of Political Economy*, vol. 94 (2), pp. 235–59.
Tröger, T. (2002). 'Why sunk costs matter for bargaining outcomes: an evolutionary approach', *Journal of Economic Theory*, vol. 102 (2), pp. 375–402.