# RETRIBUTION IN A CHEAP-TALK EXPERIMENT*

Jordi Brandts
Instituto de Análisis Económico (CSIC), Barcelona
(brandts@cc.uab.es)

Gary Charness
Universitat Pompeu Fabra, Barcelona
and
University of California at Berkeley
(charness@econ.berkeley.edu)

December , 1999

**Abstract:** We use a two-person 3-stage game to investigate whether people choose to punish or reward another player by sacrificing money to increase or decrease the other person's payoff. One player sends a message indicating an intended play, which is either favorable or unfavorable to the other player in the game. After the message, the sender and the receiver play a simultaneous 2x2 game. A deceptive message may be made, in an effort to induce the receiver to make a play favorable to the sender. Our focus is on whether receivers' rates of monetary sacrifice depend on the process and the perceived sender's intention, as is suggested by the literature on deception and procedural satisfaction. Models such as Rabin (1993), Sen (1997), and Charness and Rabin (1999) also permit rates of sacrifice to be sensitive to the sender's perceived intention, while outcome-based models such as Fehr and Schmidt (1999) and Bolton and Ockenfels (1997) predict otherwise. We find that deception substantially increases the punishment rate as a response to an action that is unfavorable to the receiver. We also find that a small but significant percentage of subjects choose to reward a favorable action choice made by the sender.

* We thank Antonio Cabrales, Rachel Croson, Maurice Schweitzer, and seminar participants at the University of Amsterdam for helpful comments. Brandts gratefully acknowledges support from the Spanish DGCICYT (PB93-0679 and PB94-0663-C03-01). This project was started while he was visiting the Department of Economics of the University of California at Berkeley. He thanks the members of the Department at Berkeley for their hospitality. Charness gratefully acknowledges support from the MacArthur Foundation. This paper is part of the EU-TMR Research Network ENDEAR (FMRX-CT98—0238).

## 1. INTRODUCTION

Notions of fair process and honorable behavior have potentially important implications for social and economic interaction. For example, managers and negotiators often have incentives to mislead others and private information may present the opportunity for doing so. Schweitzer and Croson (1999) suggest that "deception in organizations represents a significant managerial challenge across a broad range of functional areas". In many cases, deception could be used to induce an immediate desired response from a group of employees. However, there may be significant limitations to this kind of behavior: Where the character of interaction is highly interpersonal in nature, it is important to take into account the potential impact of social considerations, if only because these may impose bounds on selfish or dishonest behavior.

These issues may be relevant in the context of the current process of formulating more accurate theoretical models of human motivation. The most common assumption in economics is that people only care about maximizing their own income. But a large body of research has shown that many people choose to sacrifice money in laboratory experiments. A number of recent formal models presume that people are also motivated by considerations of altruism, inequality, or reciprocity; while people may nevertheless maximize their utility, one's own money is not the sole determinant of utility.

Pure distributional models, such as Bolton and Ockenfels (1997) and Fehr and Schmidt (1999), assert that while people may sacrifice money to reduce disparities in material payoffs, they are unconcerned with the process leading to these payoffs. On the other hand, models such as Rabin (1993), Falk and Fischbacher (1998), and Charness and Rabin (1999) consider the decisions and motives of other agents to be important determinants of behavior. In a more general vein, Sen (1997) presents a rigorous discussion of process significance and preference and proposes that "a person's preferences over *comprehensive* outcomes (including the choice process) have to be distinguished from the conditional preferences over

*culmination* outcomes given the acts of choice," where the expression 'culmination outcomes' refers to material outcomes.

People may use different kinds of information to infer others' intentions in a given situation. Several experimental studies present data that show how some specific features of process matter to people. Subjects sacrifice money to punish others who they feel have acted in an inappropriate manner and, to a lesser extent, reward generous or favorable actions. Blount (1995), Charness (1996) and Offerman (1998) find that, in sequential games, second-mover responses differ according to whether the choice set is believed to be determined by a self-interested player or by a random mechanism. Brandts and Solà (1998), Charness and Rabin (1999) and Falk, Fehr and Fischbacher (1999) have shown that the nature of foregone alternatives by the first player significantly affects binary choices by the second player.

These experimental results document the existence of two different types of process effects. In this paper we investigate the effects on behavior of another type of process variable: people's deliberate misrepresentation of their intended actions. Our experimental set-up involved communication between players, where message senders could lie in a self-serving manner. The possibility of making deceptive statements and the possibility of people reacting to the deception (once revealed) in an emotional way are important aspects of social interaction in many different contexts. We feel that it is a natural intuition that deception will be seen by others as a very direct indication of (perhaps overly) selfish intentions and may lead to substantial process-driven punishment behavior.

We examine retribution in a sequential two-player experimental game with communication.[1] In our game, after sending a non-binding message about her intended play, the sender chooses between an action which is clearly favorable and one which is clearly unfavorable for the receiver. One of the outcomes gives the sender a much larger payoff than the receiver. This outcome can be reached via two different paths - the sender may play in accordance with her message or she may not. If this outcome is reached, we allow the

receiver to "punish" by reducing both players' payoffs to a lower, but equal, level. If there is a negative reaction to deception, we should expect (at an unfavorable outcome for the receiver) a higher punishment rate when the sender has made a misleading message than when the outcome has been reached after a truthful message. At another culmination outcome of our game a receiver can choose to reward a favorable play by the sender by awarding her money from his own payoff.

We find that many people do send misleading messages and that people do sacrifice money to both punish and reward. Punishment rates are significantly higher when there has been a deceitful message of an intended favorable play, highlighting the importance of intention and process in one's dissatisfaction with a culmination outcome. Finally, the reward rate is modest but significant.

[1] Webster's Third New International Dictionary defines retribution as "the dispensing or awarding of punishment or reward according to the deserts of the individual."

## 2. PREVIOUS ANALYSIS AND EVIDENCE

Material outcomes and payoffs are certainly a crucial factor in determining preferences and decisions. Yet, numerous studies indicate that the process leading to an outcome (or to one's choice between outcomes) can also be a key influence. For example, there is a large body of work on procedural justice which supports the premise that process satisfaction is an important ingredient of human motivation. Thibaut and Walker (1975) and Tyler (1988, 1990) have argued that relational issues may dominate definitions of justice and that procedural satisfaction may be as important as outcome satisfaction. Disputant satisfaction is highest when the process seems fair and when people feel that they have had an opportunity to be heard (*voice*), and have had a hand in shaping the outcome. Evidence from the field is presented by Kitzmann and Emery (1993), who examine child custody decisions reached using different methods. While the fathers almost invariably lost in either case, they felt much better about the situation when they felt they had been treated reasonably well by the process.

Perceptions of the fairness of the process are also important for resource allocation in markets, negotiations, and labor relations. Kahnemann, Knetsch, and Thaler (1986), Barrett-Howard and Tyler (1986), and Bies, Tripp, and Neale (1993) find that procedural information influences judgments of market exploitations. In their survey on fairness in negotiations, Tripp, Sondak, and Bies (1995) suggest that the allocation of resources may be of less concern to individual agents than procedural and interactional fairness. Charness and Levine (1999) find that the perceived fairness of a layoff is highly dependent on the manner in which the layoff is implemented. There is also some evidence that perceived unfairness can lead to retaliatory behavior. Robinson and Bennett (1995) examine how employees respond to violations of the psychological contract and find behavior such as stealing from the company or co-workers, wasting company resources, lying about hours worked, and wrongfully blaming co-workers for mistakes.

Many studies in business ethics and negotiation address the specific issue of deception and its effects on behavior. While a satisfied party is more likely to maintain a positive and productive relationship with others in the environment, violations in relationships can lead to negative affect or even moral outrage. This is particularly true in the case of lying in negotiations (Anton, 1990; Lewicki, 1983). Although some feel that deception is just part of the negotiation "dance," others (e.g., Shapiro and Bies, 1994) believe that such behavior can destroy trust and cooperation in ongoing organizational relationships. Bies and Tripp (1995) suggest that the harm done to the relationship by lying may be irreversible. Schweitzer, Brodt, and Croson (1999) find 36 of 66 "union negotiators" punish deceptive "city negotiators" when the true state of affairs is revealed and that distinctions between "types" of deception (e.g., lies of omission or commission) are important.

Lewicki and Stark (1996) analyze subjects' evaluations of ethically-questionable negotiation tactics. Naturally, players' perceptions of the "game" being played may be important. If people expect lies and deception, these may not produce much of a negative response. In Roth and Murnighan (1982), disbelief of messages was common. The stakes involved may also affect expectations: Tanbrunsel (1998) finds that increased incentives lead to more misrepresentation and that the greater the incentive one has to engage in misrepresentation, the more that she expects that an opponent will engage in misrepresentation. In addition, her evidence suggests that individuals believe themselves to be more ethical than their opponents.

Romer (1996) discusses the effects of deception in the political economy context. According to his analysis, the U.S. social security system was created as an entitlement program : payroll taxes were bundled with an explicit promise of certain future transfers. The reason why it is politically very difficult to cut back benefits is that "the act of making, then breaking, a promise induces a taste for punishing the offender" (Romer, p. 199). In the

political sphere the punishment would be expressed through people voting against those who proposed a reduction in benefits.

Our study adds to the analysis of the effects of deception by presenting results from a simple environment where there is an incentive for misrepresentation and the punishment mechanism and its monetary consequences are clear. In our game, false messages are explicitly permitted, so that it was not clear whether receivers of false messages would judge these harshly.[2] A sender's willingness to engage in deception reflects her own social values and her perception of the matched receiver's social values and willingness to sacrifice money to punish unethical behavior. As shall be seen, our design rotates the subjects' roles (sender and receiver) and so permits us to examine whether subjects play in a consistent manner across roles. If the decision to punish deception is not influenced by whether one has sent (or intends to send) a false message, punishment could be seen as arbitrary and not based on underlying social values.

Our specific focus is on whether a person's choice between two outcomes is affected by whether deception has been employed. Blount (1995), Charness (1996), and Offerman (1998) have examined whether Player B's choice to act unfavorably toward Person A depends on whether an unattractive choice set was reached by Player A's choice or by random assignment; other studies cited earlier have examined the effect of alternatives foregone by Player A on binary choices by Player B. However, none of these experiments involve messages or deception. In addition, we are unaware of any study of the effects of deception that directly examines the effect of different "paths" to a choice between payoff pairs on the choice actually selected. This approach seems closest to the conceptualization in Sen (1997).

## 3. EXPERIMENTAL DESIGN AND PROCEDURES

[2] In other environments there may be "rules" against lying, while our environment is very neutral on this issue. In this sense, the punishment rates we observe may represent a sort of lower bound. We were concerned that no one would send false messages if explicitly discouraged by the instructions.

There are 3 stages in our game: the first is the announcement stage, the second is the simultaneous choice stage and the third is the retribution stage. The player we will refer to as the sender chooses an announcement in stage 1 and an action in stage 2. The player we will refer to as the receiver observes the sender's announcement in stage 1, chooses an action in stage 2 and makes an additional choice in stage 3, which is described below. Figure 1 shows the payoff table for the central part of our game:

**Figure 1**

| | **B1** | **B2** |
|---|---|---|
| **A1** | **2, 2** | **6, 9** |
| **A2** | **2, 2** | **12, 3** |

In the first stage, each person in the sender role sends a costless announcement about his intended play to an anonymous receiver: the announcement is a non-binding statement about which choice, A1 or A2, the sender will make in the second stage. In the second stage, after the message has been transmitted, both players simultaneously choose actions. The sender chooses between A1 and A2 and the receiver chooses between B1 and B2. In the third stage, the receiver has an option to change the payoffs if she has played B2. If (A1, B2) has been chosen, she can give the sender 2 units (1 unit = $1.50) and so change the payoffs to (8,7). If (A2, B2) is the outcome, she can change the payoffs from (12,3) to

(2,2). It should be clear that two matched players can reach the (A2, B2) cell by two paths - one where the message has been A1 and the other where the message has been A2.[3]

Our design is motivated by two considerations. First, we wanted a simple environment, where the (binary) choices and associated payoffs were transparent to the players. The second issue is related to our main objective of working with an environment in which it is plausible to expect self-serving lies. In our game we expected many senders' preferred outcome to be (12,3) and, therefore, their preferred action in the simultaneous choice stage to be A2. The receiver, however, may be quite averse to the (12,3) outcome, given the unequal distribution of payoffs. Therefore, if the receiver believes that the sender will choose A2 the receiver's choice may be B1. The sender may send the A1 message to encourage a B2 choice, but then actually play A2.

Given that the punishment payoffs in the retribution stage of the game are (2,2), there is really no obvious reason for the receiver to choose B1 in the action stage. If a receiver prefers a payoff of (2,2) to (12,3) she can always choose this, as a response to an A2 play, after playing B2. However, there may exist plausible explanations for receivers choosing B1 and so senders may signal A1 before action A2, in an attempt to circumvent B1 play. The fact that it is not completely transparent why the sender should expect a B1 choice after a A2 message does not interfere with the analysis we wish to perform: As long as we obtain sufficient observations of the (A2, B2) cell being reached after A1 and A2 announcements, we can compare receivers' behavior in the two cases.

Taking as given the punishment payoffs (2,2) in the retribution stage, let us consider some other payoff possibilities that could have been chosen for B1 play. Our conjecture was that combinations like (3,2), (4,2), etc. would have been less attractive for the receiver than (2,2), would have made the choice of B1 quite unlikely in the mind of senders and, hence, would have made false signals rather infrequent. The same argument applies for payoffs

[3] There is a second possibility for false messages in our design. The sender may announce A2 and then choose A1, behaving more favorably to the receiver than announced. A possible rationale for this behavior is

such as (2,1) and (2,0). The possibility that remains are payoffs combinations like (1,2) and (0,2). Our feeling was that payoffs of this kind could have made the B1 choice too attractive for the receiver and kept us from obtaining a sufficient amount of data in the (A2, B2) cell. The payoff combination we chose has a knife-edge feature which allows us to obtain, at the same time, a sizeable amount of false statements by A and of B2 choices by B.

If money is the only element in one's utility function, the only subgame-perfect Nash equilibrium involves actions (A2, B2) without punishment or reward in the third stage; the message is irrelevant.[4] Distributional models of motivation can explain a choice of (2, 2) over (12, 3), but also presume that the message is irrelevant. However, while the choice with either message is between the outcomes of (12, 3) and (2, 2), a false message and a truthful message imply different processes. Some people may resent the tactic of making a "promise" and breaking it. This would be reflected in a differential punishment rate for the two cases.

Following Bolton, Brandts, and Katok (1997), Bolton, Brandts, and Ockenfels (1998), and Brandts and Charness (1998), each game was played twice, so that each person was a sender once and a receiver once. Participants were assured that no two people were ever paired twice and subjects were not informed about the final outcome of the first play of the game when they made their decisions for the second play. This feature allows us to obtain data from one-shot interactions and gives us the opportunity to examine whether people played "consistently" in the two decisions made. Following the two periods, a coin was tossed to determine which period was used for actual payoff purposes.

An important procedural point is that receivers were not told the decisions actually made by the senders before they were asked for their choices of whether to punish and reward. Instead each receiver (who knew the message she had received) was asked to designate (after his B2 play) a contingent choice if the sender actually played A1 and a

---

to surprise the receiver in order to elicit the reward. As will be shown below, we actually observed instances of this behavior.

contingent choice if the sender actually played A2. One obvious advantage of this approach is that we can obtain a full set of two responses regardless of the sender's play.

This *strategy method* (Selten, 1967) plausibly induces different behavior than does the standard "direct-response" method. Roth (1995) mentions on pg. 323 that "having to submit entire strategies forces participants to think about each information set in a different way than if they could primarily concentrate on those information sets that may arise in the course of the game." This statement certainly makes sense, but seems most applicable in complex or unfamiliar environments. Cason and Mui (1998) and Brandts and Charness (1998) find that the strategy method does not significantly alter choices. We are unaware of any empirical evidence for a significant difference in behavior due to this method in simple games.[5]

We conducted our experimental sessions at UC-Berkeley in February and March, 1999. A total of 118 people participated in exactly one of the 5 sessions. Average earnings were around $15, including a $5 show-up fee. Recruiting was conducted primarily through the use of campus e-mail lists. An e-mail message that was sent to randomly-selected people through the Colleges of Letters, Arts, and Sciences provided the bulk of the participants, so our sessions typically included individuals from a broader range of academic disciplines than is common in economics experiments. Since e-mail is used regularly by the vast majority of Berkeley students, selection bias from this recruiting method should be minimal, at least with respect to other laboratory experiments. Instructions are provided in the Appendix.

People met in a large classroom which was divided into two sides. Individuals sat where instruction packages had earlier been placed and were well-spaced. During the first decision task the subjects on one side of the room had the role of senders while the subjects on the other side were receivers. For the second decision task these roles were reversed, and

---

[4] Two other Nash equilibrium strategy combinations, both of which presume there is nonpecuniary utility, consists in the sender playing A1 and B punishing A2 and 1) not rewarding A1 or 2) rewarding A2.

[5] Note that the fact that we elicit contingent choices in stage 3 ensures that no player knows the outcome of the first play before making a decision in the second play.

subjects were assured that they were not matched with the same person as in the first decision.

## 4. RESULTS

Table 1 presents a summary of the data, where the notation s(A1) and s(A2) refers to the A1 and A2 signals. We find that about 1/3 of all A1 messages are false, as they are followed by A2 play. By punishment we refer to the receiver choosing payoffs (2,2) after an (A2,B2) realization and by reward to the choice of (8,7) after (A1,B2). We remind the reader that each receiver was asked to designate separate choices for the case where the sender actually played A1 and the case where the sender actually played A2.[6]

There is a substantial amount of punishment and a modest, but non-zero, reward rate. The rate of punishment of action A2 is much higher given a false signal. We also find that overall about 1/6 of the subjects choose to reward a play of A1. A surprising phenomenon is that about 1/4 of all people who signal A2 actually play A1; in addition, this sequence is rewarded at a slightly higher rate than is signaling A1 and playing A1. Somewhat reassuringly, 95% of the subjects played B2.[7]

**Table 1 – Data Summary**

| Decision | Proportion | Percentage |
|---|---|---|
| s(A1) | 76/118 | 64% |
| s(A2) | 42/118 | 36% |
| A1 | 62/118 | 53% |
| A2 | 56/118 | 47% |
| s(A1), play A2 | 25/76 | 33% |

[6] Thus, there should be 112 total responses to A1 messages and to A2 messages (118 subjects less those 6 who played B1). However, 3 subjects failed to indicate a response to a hypothetical A2 play and one subject failed to indicate a response to a hypothetical A1 play. Thus, there are only 109 total responses to A2 play and 111 total responses to A1 play.

[7] 95% of B players played B2 in each case - 72/76 played B2 after an A1 message, while 40/42 played B2 after an A2 message. If one interprets the choice of B1 as a pure mistake, note that a 5% "error rate" is not at all unusual in laboratory experiments. Of the 6 people who played B1, 4 had received messages of A1. This behavior is not easy to interpret in terms of purposeful actions, although a possible explanation is that these receivers were very sure that an A1 statement was a lie. In the presentation of our statistical tests below we follow up on this interpretation.

| | | |
|---|---|---|
| s(A2), play A1 | 11/42 | 26% |
| B1 | 6/118 | 5% |
| B2 | 112/118 | 95% |
| Punish(s(A1), play A2) | 19/69 | 28% |
| Punish(s(A2), play A2) | 5/40 | 12% |
| Punish (overall) | 24/109 | 22% |
| Reward (s(A1), play A1) | 10/72 | 14% |
| Reward (s(A2), play A1) | 9/39 | 23% |
| Reward (overall) | 19/111 | 17% |

We can use the test of the equality of proportions (Glasnapp and Poggio, 1985) to see if the observed behavior differs across conditions. The Z-statistic reported is the normal approximation to the binary distribution, defined by the difference between the proportions divided by the standard error of the difference.

The first row of table 2 shows that, despite the modest number of observations, the rates of punishment are significantly different depending on whether the message had been A1 or A2.[8] This is support for our primary hypothesis that the process, here gauged

**Table 2 – Statistical Tests**

| *Comparison* | *Proportions* | *Z* | *p-value*[9] |
|---|---|---|---|
| Punish (L)<br>vs.<br>Punish (NL) | 19/69<br>vs.<br>5/40 | 1.83 | .03 |
| R(1)<br>vs.<br>R(2) | 10/72<br>vs.<br>9/39 | 1.23 | .22 |
| Actual R(1+2)<br>vs.<br>Presumed R(1+2) = 0 | 19/111<br>vs.<br>0/111 | 4.56 | .00 |
| Prob(PL\|L)<br>vs. | 3/19<br>vs. | 2.04 | .02 |

[8] If we interpret the 6 instances of B1 choices in terms of purposeful behavior the existence of a statistical difference between the punishment rates goes through. If B1 after A1 is not an error, a reasonable interpretation of this choice is that B is expecting a lie. In this case the punishment rate after a lie changes to 23/73=31% by adding 4 in the numerator and the denominator. Analogously the punishment rate after truthful messages changes to 7/42=16%. The test for differences yields Z=1.75 and p=.04 (one-tailed test).

[9] We had directional hypotheses for comparisons 1, 3, and 4, so these p-values reflect one-tailed tests. However, there was no directional hypothesis regarding reward rates across messages, so this p-value reflects the two-tailed test.

| | | | |
|---|---|---|---|
| Prob(PL\|NL) | 21/50 | | |

Here L means "lie" or (s(A1), play A2), NL means "not lie" or (s(A2), play A2), PL|L means a receiver punished a lie if he himself lied as a sender, and PL|NL means a receiver punished a lie if he himself didn't lie as a sender. R(1) or R(2) means reward after signal A1 or A2, respectively, followed by an A1 play.

by the accuracy of the message, affects the choice between culmination outcomes.[10] A deceptive message is not considered to be appropriate by many participants and a substantial number of these are even willing to sacrifice money to express their displeasure; this suggests that receivers felt that senders were intending to use a false message in a manipulative way. The punishment of deceptive statements can also be interpreted in terms of aspiration levels: When player B gets the message A1 his aspiration level moves to a payoff level of 9. A subsequent action A2 gives him a payoff of 3, below his aspiration level, and induces punishment.

The second row of table 2 shows that the reward rate is more than 50% higher when A1 is played after an A2 message, suggesting that people respond more favorably when the sender shows a favorable change of heart. While this "surprise effect" seems plausible *ex post*, we had no hypothesis to this effect. The difference in rates is not significant at standard levels, but is suggestive. Possibly any such effect could be explained by a feeling of entitlement once one has received an A1 signal, so that the actual A1 play doesn't trigger a reward.

The test result shown in the third row of table 2 does not pertain to the analysis of the effects of deception. Our focus here is on studying the reward of senders taking the favorable action A1. We feel that the 17% reward rate is much higher than what would have been obtained if we simply gave B players unilateral power to choose (6,9) or (8,7) and had no messages, although we do not run this control. It is rare for subjects in experiments to

[10] An argument could be made that the true difference in rates is higher, as the Punish (NL) data may include some B's who may have been indifferent to playing B1, except for the small chance of an A1 play.

choose to receive less than the other player when a choice to receive more is available.[11] Here the total social payoff doesn't even increase.

Although this is only a hypothetical test, the difference is highly significant (even a control rate of 10/111 would be significantly less than the observed combined reward rate of 19/111). We take this result to be evidence of some kind of positive reciprocity. By that we mean that Player B is more willing to sacrifice money to increase the payoff of a Player A who is perceived to have intentionally chosen an action favorable to Player B than to increase the payoff of a Player A who is perceived to either have had no choice in the allocation or to have intentionally chosen an action *un*favorable to Player B.

In the last row of table 2 we can also see that subjects exhibit a fairly consistent attitude towards punishment behavior, in the sense that far fewer receivers punish a false message if they would send it themselves than if not. Over 40% punish deception, denoted in table 2 by PL/L, when they don't use it themselves, compared to 16% otherwise (PL/NL). This "internal consistency" is an indication that subjects understand the situation and parallels results such as Kahneman, Knetsch, and Thaler (1986), where 88% chose to punish a selfish chooser if one had not been a selfish chooser, but only 31% punished behavior they would engage in themselves. Brandts and Charness (1998) and Charness and Rabin (1999) find the same kind of consistency in behavior across rotated roles.

How do the effective levels of retribution affect players' material payoffs? Table 3 shows expected material payoffs for all combinations of message and action.[12] If a sender

[11] It is easy to prove that, given the parameter specifications in Fehr and Schmidt (1999), no B should prefer (8,7) to (6,9). The Bolton and Ockenfels model does not provide a functional form, so we only know that an individual would compare a mild improvement in the equality of the payoff ratio [from (6/15, 9/15) to (8/15, 7/15)] with the sacrifice of the 2 payoff units.

[12] Given receivers' behavior, it appears that material payoffs account only for part of their motivation. The other part stems from the satisfaction of punishing a self-serving lie and of rewarding a pleasant surprise. It is conceivable that senders' payoff are also affected by a non-material component, although this force is more difficult to gauge; senders' may derive some satisfaction from sending a truthful message.

only cares about her own material payoff she should signal A2 and then play A2; it does not pay to send an A1 message prior to an A2 play. [13]

**Table 3 – Ex post expected material payoffs**

| Message, action | Sender expected payoff | Receiver expected payoff | Combined expected payoff |
|---|---|---|---|
| s(A2), A2 | 10.36 | 2.84 | 13.20 |
| s(A1), A2 | 8.86 | 2.69 | 11.55 |
| s(A2), A1 | 6.23 | 8.25 | 14.48 |
| s(A1), A1 | 6.05 | 8.38 | 14.43 |

On the other hand, a sender with a sufficient positive weight on the receiver's material payoff should play A1, as this leads to higher social benefits than A2, regardless of the signal. We see (Table 1) that 53% of senders play A1, giving them lower material payoffs *ex post* than would have expected from an A2 play. While we cannot know precisely why this choice was made (perhaps an exaggerated fear of punishment), the results suggest this may reflect pro-social behavior.[14]

## 5. DISCUSSION

Previous studies indicate that process satisfaction is a very salient consideration in many situations. The negotiation and business ethics literature suggests that deception, while a fairly common practice, often induces negative responses. Generally, people react in an adverse manner when they feel their right to fair treatment has been violated.

Our study examines the effect of deception by comparing responses when messages are honest and when they are not. In accordance with this view, we find that false messages

[13] This result suggests that in a dynamic relationship, false messages would be driven out, since they ultimately are not in the sender's interest. However, our focus here is on the study of preferences. For this purpose what matters are the reactions to false statements, when they are made.

[14] A comparison of the expected payoffs shown in the first and second rows of table 3 reveal the effects, in economic terms, of the punishment of false messages. If one relates the reductions of sender and receiver payoffs to the largest possible reduction from the payoffs shown in the first row of table 3 down to a payoff of 2, one obtains a fraction of 1.5/8.36 for the sender and of .15/.84 for the receiver. Both fractions are close to .18.

lead to punishment much more frequently than do accurate messages, even though the choices between culmination outcomes is the same in both cases. This may be a bit surprising, since false messages are explicitly permitted in the instructions and there is no obvious deterrent value for punishment. We also find that a receiver's decision regarding punishing deception is correlated with his behavior as a sender, supporting the view that the decision to punish is part of a consistent value orientation. While responses to favorable play are not our primary focus, we do find that many people donate money in response, choosing to come out behind; this suggests the presence of positive reciprocity.

This experiment adds to the literature on process satisfaction/outcome satisfaction by examining a specific choice between outcomes, when this choice can eventuate through two different processes. This simple design highlights the relevance of a person's beliefs (influenced here by the message choice) about another individual and provides a clean illustration (with clear monetary incentives) of the consequences of deception. The data from the role rotation indicates that punishment is not simply arbitrary or random.

Punishment has been defined (Kadzin, 1975) as the presentation of an adverse event or the removal of a positive event following a response. Romer (1996) suggests that people can threaten (and impose) punishment for at least two reasons: 1) threats of punishment may have strategic value (even in the static case) or actual punishment may have future deterrent value (dynamic case); 2) a person can have a taste of desire for punishing others that is triggered by a particular sequence of events. In this case, a person punishes because it is satisfying to do so. Given the non-repeated nature of our design, there should be no deterrence motivations or future financial considerations. It seems that people simply do not like being misled and that this triggers a "taste for punishment", which may be related to self-respect issues. However, it is also possible that people are willing to provide a public good - a socially-beneficial "object lesson" to deter further anti-social behavior outside the laboratory.

Models such as those in Rabin (1993), Sen (1997), and Charness and Rabin (1998) offer a role for the perceived intention of an agent and the extent to which an action is considered socially appropriate. While pure distributional models such as Fehr and Schmidt (1999) and Bolton and Ockenfels (1997) can explain a B preference for (2,2) over (12,3), these models consider process and intention irrelevant and so offer no explanation for the differening punishment rates. In addition, the non-zero reward rate suggests positive reciprocity, as evidence and models indicate very few people would otherwise sacrifice money to come out behind rather than ahead.

We would like to mention some limitations of our study. Punishment and reward decisions are made using the strategy method and may not be the same as would be made using the direct-response method. Some may be concerned that the rotation of roles could affect choices, even though pairings differ across periods. As is usual in such experiments, students are the subjects and the stakes are modest. Yet we provide a comparative test of punishment rates and so try to compensate for any bias introduced by our methodology.

Further research should focus on identifying the underlying determinants of why people choose to sacrifice money to punish or reward a person's behavior. Given that non-proscribed misleading messages lead to such a substantial increase in punishment in a one-shot interaction, one should expect that effects would be larger in an ongoing relationship. Firms and managers should take heed and avoid deception and similar practices.

## REFERENCES


Anton, R., "Drawing the Line: An Exploratory Test of Ethical Behavior in Negotiations," 1990, The International Journal of Conflict Management, **1**, 265-280.

Barrett-Howard, E. and T. Tyler, "Procedural Justice as a Criterion in Allocation Decisions," 1986, Journal of Personality and Social Psychology, **50**, 296-304.

Bies, R. and T. Tripp, "Beyond Distrust: 'Getting Even' and the Need for Revenge," 1995, in Trust and Organizations, R. Kramer and T. Tyler, eds., Sage: Newbury Park, CA.

Bies, R., T. Tripp, and M. Neale, "Procedural Fairness and Profit Seeking: The Perceived legitimacy of Market Exploitation," 1993, Journal of Behavioral Decision Making, **6**, 243-256.

Blount, S., "When Social Outcomes Aren't Fair: The Effect of Causal Attributions on Preferences," 1995, Organizational Behavior and Human Decision Processes **63**, 131-144.

Bolton, G. and A. Ockenfels, "ERC: A Theory of Equity, Reciprocity, and Competition," 1997, mimeo (forthcoming, American Economic Review).

Bolton, G., J. Brandts, and E. Katok, "How Strategy Sensitive are Contributions? A Test of Six Hypotheses in a Two-Person Dilemma Game," 1997, mimeo (forthcoming, Economic Theory).

Bolton, G., J. Brandts, and A. Ockenfels, "Measuring Motivations for the Reciprocal Responses Observed in a Simple Dilemma Game," 1998, Experimental Economics, **1**, 207-219.

Brandts, J. and G. Charness, "Hot vs. Cold: Sequential Responses in Simple Experimental Games," 1998, mimeo, (forthcoming, Experimental Economics).

Brandts, J. and C. Solà, "Reference Points and Negative Reciprocity in Simple Sequential Games," 1998, mimeo.

Cason, T. and V. Mui, "Social Influence in the Sequential Dictator Game," 1998, Journal of Mathematical Psychology, **42**, 248-265.

Charness, G., "Attribution and Reciprocity in a Simulated Labor Market: An Experimental Investigation," 1996, mimeo.

Charness, G. and D. Levine, "When are Layoffs Acceptable? Evidence from a Quasi-Experiment," 1999, mimeo (forthcoming in Industrial and Labor Relations Review).

Charness, G. and M. Rabin, "Social Preferences: Some Simple Tests and a New Model," 1999, mimeo.

Falk, A. and U. Fischbacher, "A Theory of Reciprocity," 1998, mimeo.

Falk, A., E. Fehr, and U. Fischbacher, "On the Nature of Fair Behavior," 1999, mimeo.

Fehr, E. and K. Schmidt, "A Theory of Fairness, Competition, and Cooperation," Quarterly Journal of Economics, 1999, **CXIV**, 817-868.

Glasnapp, D. and J. Poggio, Essentials of Statistical Analysis for the Behavioral Sciences, 1985, Merrill: Columbus.

Kadzin, A., The Token Economy, 1977, Plenum Press: New York.

Kahnemann, D., J. Knetsch, and R. Thaler, "Fairness as a Constraint on Profit-Seeking: Entitlements in the Market," 1986, American Economic Review, **76**, 728-741.

Kitzmann, K., and R. Emery, "Procedural Justice and Parents' Satisfaction in a Field Study of Child Custody Dispute Resolution,"1993, Law and Human Behavior, **17**, 553-567.

Lewicki, R., "Lying and Deception: A Behavioral Model,"1983, in Negotiating in Organizations, M. Bazerman and R. Lewicki, editors, Sage: Beverly Hills.

Lewicki, R. and N. Stark, "What is Ethically Appropriate in Negotiations: An Empirical Examination of Bargaining," 1996, Social Justice Research, **9**, 69-95.

Offerman, T., "Hurting Hurts More than Helping Helps: The Role of the Self-Serving Bias,"1998, mimeo.

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

Robinson, S., and R. Bennett, "A Typology of Deviant Workplace Behaviors: A Multidimensional Scaling Study, 1995, Academy of Management Journal, **38**, 555-572.

Romer, P., "Preferences, Promises, and the Politics of Entitlement," 1996, in Individual and Social Responsibility: Child Care, Education, Medical Care, and Long-Term Care in America, V. Fuchs, ed., University of Chicago Press: Chicago.

Roth, A., "Bargaining Experiments," in Handbook of Experimental Economics, J. Kagel and A. Roth, eds., 1995, 253-348.

Roth, A. and J. Murnighan, "The Role of Information in Bargaining: An Experimental Study," 1982, Econometrica, **50**, 1123-1142.

Schweitzer, M. and R. Croson, "Curtailing Deception: The Impact of Direct Questions on Lies and Deceit," 1999, mimeo.

Schweitzer, M., S. Brodt, and R. Croson, "Deception, Distance, and Technology: A Comparison of Vidoeconference and Telephone Negotiations," 1999, mimeo.

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

Sen. A., "Maximization and the Act of Choice," 1997, Econometrica, **65**, 745-779.

Shapiro, D. and R. Bies, "Threats, Bluffs, and Disclaimers in Negotiations," 1994, Organizational Behavior and Human Decision Processes, **60**, 14-35.

Tanbrunsel, A., "Misrepresentation and Expectations of Misrepresentation in an Ethical Dilemma: The Role of Incentives and Temptation,"1998, Academy of Management Journal, **41**, 330-339.

Thibaut, J., and L. Walker, Procedural Justice: A Psychological Analysis,1975, Erlbaum: Hillsdale, NY.

Tripp, T., H. Sondak, and R. Bies, "Justice as Rationality: A Relational Perspective on Fairness in Negotiations," 1995, in Research on Negotiation in Organizations, Volume 5, R. Bies, R. Lewicki, and B. Sheppard, eds., 45-64, JAI Press.

Tyler, T., "What is Procedural Justice?: Criteria Used by Citizens to Assess the Fairness of Legal Procedures," 1988, Law Soc. Review, **22**, 301-355.

Tyler, T., Why People Obey the Law: Procedural Justice, Legitimacy, and Compliance, 1990, Yale University: New Haven.

# APPENDIX

## Instructions

*Introduction.* Thank you for participating in this session. The purpose of this session is to study how people make decisions in a particular situation. Feel free to ask us questions as they arise, by raising your hand. From now until the end of the session, unauthorized communication of any nature with other participants is prohibited.

You will receive $5 for participating in this session. You will receive additional money for the decisions that are made in the session. Upon completion of the session the amount you make will be paid to you in cash. Payments are confidential: no other participant will be told the amount of money you make.

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 task*. 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 you make and on the decisions of the person with whom you are paired. The decision task consists of three stages and of two decisions for A and two for B. You will be paired with the same person during the three stages of the decision task. The attached figure, also shown on the blackboard, describes the options available to each person and the associated payoffs. In stage 1, the person in role A will announce an intended play for stage 2. A will announce either A1 or A2; you are not required to play the move you have announced. This announcement will be communicated to the person in role B.

Then you will move to stage 2 of the decision task: the person in role A will choose a move between A1 and A2, and the person in role B will choose a move between B1 and B2. At the end of stage 2 we will communicate to B the choice made by A. The earnings table shown in the figure under stage 2 exhibits the possible payoffs at the end of stage 2.

In stage 3, B's will choose between different options, depending on what has happened in stage 2. The final payoffs thus depend on A's choice between A1 and A2, on B's choice between B1 and B2 and on B's choice in stage 3. In the table shown under stage 3 in the figure, you can see how B's choice in stage 3 affects the payoffs of stage 2.

In stage 3, B's choice will be conditional; that is, if B has chosen B2 in stage 2, B will now indicate a decision for the case where A has chosen A1 and a decision for the case where A has chosen A2. The decision that counts is the one that corresponds to the decision of A.

*Conduct of the sessions.* You will make your decisions by choosing one of the options available to you and recording it on a paper form. In stage 1, each A will record the announcement of A1 or A2 on his or her announcement sheet. Each B will fill out a no announcement form. We will collect all forms for stage 1 and will present each B the announcement of the corresponding A. Next, A's and B's will simultaneously make choices for stage 2 and will record them on their decision sheets. We will collect these forms and communicate to each B what the corresponding A chose in stage 2. B will then make stage 3 decisions.

You will participate in two decision tasks, Task 1 and Task 2. Both tasks are identical to the description in the previous paragraph. For each task, you will be paired with a different person. You will have the role of A in one task, and the role of B in the other. First, you will receive decision forms for the role you have in task 1. You will complete task 1, and all the forms will be collected. You will then receive decision forms for task 2 and complete task 2. The results of task 1 will not be revealed prior to completion of task 2. At the end of the session, you can learn your results for each decision task.

*Payment*. You will actually be paid your earnings for just one of the two tasks. The one for payment will be chosen by a coin flip after both tasks have been completed. We will then call you one by one to receive your payment, which you will receive privately. Once you are paid you may leave.

## TASK 1

## A'S ANNOUNCEMENT FORM FOR STAGE 1

You have role A. Please make an announcement of your intended play. This announcement will be shown to the B player with whom you are paired. After B has received the announcement, you will both simultaneously make your stage 2 choices. B will be informed of your choice of moves and will then make a stage 3 choice.

Please indicate your announcement below.

**I intend to play:**

**A1** **A2**

# TASK 1

B'S NO ANNOUNCEMENT FORM FOR STAGE 1

You have role B. You do not make a decision at this stage. After you receive A's announcement, you will make choices simultaneously.

**I understand I do not make an announcement now**

_____

## TASK 1

## A'S DECISION FORM FOR STAGE 2

You have role A. Please choose to play either A1 or A2 and indicate your choice below.

**I choose to play:**

**A1** **A2**

## TASK 1

## B'S DECISION FORM FOR STAGE 2

You have role B. Please choose to play either B1 or B2 and indicate your choice below.

**I choose to play:**

**B1** **B2**

## TASK 1

## B'S DECISION FORM FOR STAGE 3

You have role B. If you have chosen B2 in stage 2, please indicate your choice below. Your choice is conditional; that is, if you chose B2 before, you will now indicate a decision for the case where A has chosen A1 and a decision for the case where A has chosen A2. Please refer to the tables to see which final payoffs would result from your choice.

**For the case where A chose A1, I choose to play:**

**B3** **B4**

**For the case where A chose A2, I choose to play:**

**B5** **B6**