Hurting hurts more than helping helps: the role of the self-serving bias

Theo Offerman*

Current version: February, 1999

Abstract:
This paper investigates an implication of the self-serving bias for reciprocal responses. It is hypothesized that negative intentionality matters more than positive intentionality for reciprocating individuals with a self-serving attributional style. Experimental evidence obtained in the hot response game supports this prediction. Subjects are 67% more likely to reciprocate an intentional hurtful choice over an unintentional hurtful choice. Subjects are only 25% more likely to reciprocate an intentional helpful choice over an unintentional helpful choice. The evidence on the intermediating role of emotions is consistent with the explanation offered by the self-serving bias.

JEL-codes: C70, C92.
Key words: reciprocity, intentionality, self-serving bias, experiment

Address:
University of Amsterdam
Department of Economics
Roetersstraat 11
1018 WB Amsterdam
Netherlands

e-mail: theoo@fee.uva.nl
Tel: 31-20-5254294
Fax: 31-20-5255283

*I especially thank Joep Sonnemans for his numerous helpful suggestions. I am grateful for useful comments from Ronald Bosman, Gary Charness, Jens Grosser and Arno Riedl, and from participants of the CREED seminar. Financial support from NWO is gratefully acknowledged.

1. Introduction

People entertain unduly positive beliefs about themselves. Social psychological research shows that typically a considerable majority of respondents rate themselves in the top 50% part of the population in various desirable skills and qualities.[1] For example, Svensson (1981) reports that 93% of American drivers and 69% of Swedish drivers think that they belong to the top 50% of most skillful drivers (for a list of similar examples and references, see Babcock and Loewenstein, 1997).

Individuals prefer to have a positive self-image. They attribute good events to internal causes, such as great skill, personal attractiveness or high intelligence. They attribute bad events to external causes, such as uncontrollable circumstances or other persons' failures. This attributional style is called the self-serving bias. It helps people to boost their self-image (*cf.* Sedikides, Campbell, Reeder and Elliot, 1998 or Brown, 1986).

Other factors also contribute to a positively biased self-image. For example, when their self-image is low or when they feel threatened, people tend to judge themselves by comparing their own qualities with those of others who score less on the qualities under consideration. This process of downward social comparison helps them to enhance their self-image (Wills, 1981).

There are some interesting implications of the self-serving bias for economics. The self-serving bias explains how costly impasses in bargaining evolve. Bargainers' perceptions of a fair outcome of the bargaining process are distorted in the direction of their own self interest. Those bargainers who are prepared to give up money for avoiding unfair settlements are likely to get stuck in costly impasses (Babcock, Loewenstein, Issacharoff and Camerer, 1995; Babcock, Wang and Loewenstein, 1996). It has also been suggested that the self-serving bias explains part of the high failure rate of new businesses (Camerer, 1997).

This paper focuses on an implication of the self-serving bias for reciprocity. The principle of reciprocity has been successfully used to explain important economic phenomena. For example, Sugden (1984) develops a model based on reciprocity to explain voluntary contributions to public goods. A series of theoretical papers by Akerlof (1982) and Akerlof and Yellen (1988, 1990) shows that reciprocity provides a potential explanation for involuntary unemployment in the labor market. This explanation is in agreement with experimental data obtained by Fehr, Kirchsteiger and Riedl (1993; 1998). Even in their double auction market experiments gift exchange forces are strong enough to prevent market clearing. Employers offer higher than market clearing wages and employees reciprocate by choosing a higher effort level than the payoff maximizing effort level.

Early work on applications of reciprocity in economics was silent about the nature of reciprocity. The recent debate yields two accounts for reciprocity. The first explanation proposes that

[1]Depressed people seem to form an interesting exception: they tend to have a realistic self-image (cf. Lewinsohn, Mischel, Chaplin and Barton, 1980).

players are concerned with the monetary consequences of their actions. However, they do not only care about their own payoff, but also about the distribution of payoffs over the players. In doing so, they are supposed to dislike unequal divisions of payoffs or they are supposed to be motivated by feelings of altruism. There is experimental evidence supporting this distributional hypothesis (Bolton, Brandts and Katok, 1996; Bolton, Brandts and Ockenfels, 1997). Two theoretical models have been developed on the assumption that reciprocity is best explained by distributional considerations (Bolton and Ockenfels, 1997; Fehr and Schmidt, 1997).

The second explanation maintains that players try to assess the intention of the other player(s). They are inclined to reward players that intend to reward them and they are inclined to punish players that intend to punish them. There is experimental evidence supporting this intentionality hypothesis (Blount, 1995; Charness, 1998). Two theoretical models build on the assumption that deviations from selfish behavior are driven by assessments about the negative or positive intention of other players (Rabin, 1993; Dufwenberg and Kirchsteiger, 1998). Levine (1997) develops a more general model that allows for distributional and intentionality considerations simultaneously.

Probably both distributional and intentionality considerations play a role in human behavior. People's self-serving attributional style provides an argument for an asymmetric effect of intentions. A helpful action of another person will trigger a positive emotion in an individual. The intensity of this positive emotion will not be much stronger for intentional helpful actions than for unintentional helpful actions. The self-serving bias allows people to take credit for the helpful action of someone else. After all, other people like to help someone as nice as yourself. Helping you provides them an intrinsic payoff, making an extra gift less necessary. A helpful action fits very well in people's positive view of themselves. Therefore, intentional helpful actions will only marginally more often be reciprocated than unintentional reciprocal actions. On the other hand, the self-serving bias allows people to blame the other for an intentionally hurtful action. Such an action is in sharp conflict with the positive self-image of the individual. It may easily lead to reasoning like: "he should not think that he can get away with it". An intentional hurtful action of another person will thus trigger a strong negative emotion. The strong negative emotion takes care that the intentional hurtful action is sternly reciprocated. An unintentional action will also lead to a negative emotion, but it will be less strong because there is nobody to blame for.[2] For hurtful actions intentionality may matter a lot.

The main hypothesis to be tested in this paper is that negative intentions provoke stronger reciprocal responses than positive intentions. The experimental evidence obtained clearly supports this hypothesis. There is also evidence for the explanation offered by the self-serving bias.

[2]Sonnemans and Frijda (1995) ask subjects to report on recalled emotion instances. For 83% of the cases that a subject reports anger, another person was held responsible for the situation. The intensity of the reported anger increases with the blameworthiness of the other.

Other experimental studies shed some light on the effect of intentions on reciprocity. Blount (1995) investigates the effect of negative intentionality in an ultimatum game. She compares the behavior of responders facing a proposed split made by a random number generator with the behavior of responders facing a proposed split made by an actual player. Rejection rates are substantially smaller in her "random treatment" than in her "interested party treatment".[3] Charness (1998) investigates the effect of intentionality in a simulated labor market. In his experiment employers offer a wage before employees choose their effort level. Employees provide more effort at high wages when the wage is employer-determined than when the wage is randomly determined. Conversely, employees provide less effort at low wages when the wage is set by the employer. Charness concludes that both positive and negative reciprocity play a role in his experiment. Bolton, Brandts and Ockenfels (1997) compare the effects of positive and negative intentionality in a series of simple 2-players dilemma games. In contrast to Charness, the authors find no effect of positive intentionality, and in contrast to both Blount and Charness, they only find an insignificant and negligible effect of negative intentionality. These authors conclude that the distributional hypothesis satisfactorily takes account of their data. A combination of three characteristics in their design may dampen the effects of intentionality. First, players choose simultaneously. Second, the recipocating player formulates a strategy: she indicates what she will choose for each of the other player's two possible choices. These two characteristics may easily alleviate emotions experienced by the reciprocating player: you may experience a stronger emotion when you find out that someone has actually hurt or helped you than when you have to imagine that someone has hurt or helped you.[4] Third, subjects play the game twice with role reversal. Even though they play with a different partner and without information in between the games, subjects might think that they can balance a choice for one player-type with a choice for the other player-type.

Bolton *et al*. (1997) show the limits of the intentionality hypothesis. This paper will focus on a situation that is more favorable for intentionality: players choose sequentially, they do not formulate a strategy and they only play the game once.

The potential asymmetric effect of positive and negative intentionality is assessed in the "hot response game". In this game the first mover makes a choice between a helpful and a hurtful choice. The helpful choice increases the payoff of the first mover by 8 guilders and it increases the payoff of the second mover by 4 guilders. The hurtful choice increases the payoff of the first mover by 11 guilders and it decreases the payoff of the second mover by 4 guilders. The second mover observes the choice of the first mover, before she chooses between a cool, remunerative response and a hot, reciprocal response.

[3]Blount reports the results of a third treatment, where the proposal has been made by a "third party". This treatment yields similar retaliatory responses as the "interested party" treatment.

[4]Brandts and Charness (1998) evaluate the potential effect of the strategy method: they do not find fewer hot responses with the strategy method. The strategy method alone does not seem to make the difference.

Cool does not affect the payoff of the first mover, but it increases the payoff of the second mover by 10 guilders. Hot responses are reward and punish. Reward increases the payoff of the first mover by 4 guilders and it increases the payoff of the second mover by 9 guilders. Punish decreases the payoff of the first mover by 4 guilders and it increases the payoff of the second mover by 9 guilders. The choices and their consequences for the payoffs are summarized in table 1.

Table 1
The hot response game

	choice	payoff first mover	payoff second mover
first mover	helpful	+ 8	+ 4
	hurtful	+ 11	- 4

	choice	payoff first mover	payoff second mover
second mover	reward	+ 4	+ 9
	cool	+ 0	+ 10
	punish	- 4	+ 9

Notes: first mover makes a choice before second mover. Second mover observes the choice of first mover before she makes a choice. The payoff for a player is the sum of the consequences of both choices. Payoffs are in Dutch guilders.

Note the high degree of symmetry in this game. On the one hand there is symmetry between helpful and hurtful choices: the former gives 4 guilders to the other player, the latter takes 4 guilders from the other player. On the other hand, there is symmetry between rewarding and punishing: in order to punish, a player needs to sacrifice 1 guilder to decrease the other player's payoff by 4 guilders; in order to reward, a player needs to sacrifice 1 guilder to increase the other player's payoff by 4 guilders. Rewarding and punishing are equally costly and equally effective. The game is presented in this decomposed manner to emphasize this symmetry. In the usual extensive form representation the symmetry is somewhat less transparent. This game does not favor positive or negative intentionality.

The experiment consists of two treatments. In "Nature" the choice of first mover is determined by the outcome of a throw with a die. In "Flesh and Blood" first mover chooses voluntarily. The effect of positive intentionality can then be assessed by comparing the probability of a reward given a helpful choice in Flesh and Blood and this probability in Nature. Likewise the effect of negative intentionality appears from the difference in probability of a punishment given a hurtful choice between Flesh and

Blood and Nature. A comparison of both effects will reveal an asymmetry between positive and negative intentionality.

The potential role of negative and positive emotions suggested by the explanation based on the self-serving bias is also evaluated. To that purpose, after the game has been played, second movers are asked to think back to the moment when they were informed about the decision of the first mover. They report whether they experienced some positive and/or negative emotions. If they did experience an emotion, they also report the intensity of the emotion.

Players are also asked to report their beliefs. First movers report their beliefs of the second mover's choice given their own choice. This makes it possible to evaluate whether first movers anticipate the effects of intentionality correctly. Some of the second movers in Flesh and Blood report their beliefs about the choice of the first mover before they are informed about this choice. This makes it possible to see whether surprise affects the intensity of an emotion. Second mover may experience a stronger negative emotion after a hurtful choice if he did not expect such a choice. Likewise, he may experience a stronger positive emotion after a helpful choice if he did not expect it. Players are encouraged to report their beliefs seriously by the payoff of a quadratic scoring rule.

The remainder of this paper is organized as follows. Section 2 describes details of the design and the experimental procedure. Section 3 provides a discussion of the results. Section 4 provides some conclusions.

2. Design and procedure

The experiment is run by hand. Half of a group of subjects are allocated to the role of first mover, the other half to the role of second mover. Subjects do not know with whom they are paired. The instructions of the game are distributed and read aloud. The instructions use neutral formulations. Subjects receive a payoff table like table 1, but with neutral labels: player X (instead of first mover) chooses between X1 (instead of helpful) and X2 (instead of hurtful). This information is communicated to player Y (instead of second mover). Player Y chooses between Y1 (instead of reward), Y2 (instead of cool) and Y3 (instead of hurtful). It is emphasized that the game is only played once.

In Nature the choice of first mover is determined by her throw with a die: if she throws 1, 2 or 3, she will choose helpful and otherwise she will choose hurtful. When the rules of the game have been explained, subjects learn whether they have the role of first mover or the role of second mover. Each first mover throws the die. The experimenter records choices of players X and communicates each choice to the paired second mover. Second movers make their choice. Then all players fill out a questionnaire.

Amongst other questions second movers are asked to think back to the moment that the decision of first mover was communicated to them. For each of the negative emotions anger, annoyance, disappointment and contempt and each of the positive emotions gladness, relief, appreciation and feeling lucky second movers report whether they experienced it. If they answer 'yes', they also indicate the intensity of the emotion on a 1 (=little intensity) to 7 (=much intensity) scale.

In their questionnaire first movers report their beliefs about the choice of second mover given their own choice. They do this before the actual choice of second mover is communicated to them. Let p_1 denote the reported probability of second mover playing reward (Y1), p_2 the reported probability of second mover playing cool (Y2) and p_3 the reported probability of second mover playing punish (Y3) (thus, $p_1+p_2+p_3=1$). If second mover actually chooses Yj ($1 \leq j \leq 3$), first mover gets a Quadratic Scoring rule payoff in Dutch cents of QS(j):

$$QS(j) = 100 + 200 * p_j - 100 * \sum_{j=1}^{3} p_j^2 \quad .$$

Subjects are told that their expected payoff is highest if they report their probabilities truthfully. It is emphasized that it is not important to have further mathematical insight in the formula. The Quadratic Scoring rule has been used in McKelvey and Page (1990) in an information aggregation experiment and in Offerman, Sonnemans and Schram (1996) to elicit subjects' beliefs in public good games. It is discussed in Offerman (1997). Sonnemans and Offerman (1998) show that on average subjects' reported beliefs are not affected by their risk attitudes.

When subjects have completed their questionnaire, they are privately paid. Only at that time first mover learns the choice of second mover.

In Flesh and Blood first mover chooses voluntarily. The experimental procedure is the same as in Nature with one exception. Some second movers report their belief about the choice of first mover before this choice is communicated to them. Let q_1 denote second mover's reported probability that first mover chooses helpful. If first mover actually chooses helpful, then second mover receives the Quadratic Scoring Rule payoff of $400*q_1-200*q_1^2$. If first mover actually chooses hurtful, then second mover receives the Quadratic Scoring Rule payoff of $200-200*q_1$. These second movers are provided with a table instead of the formula of the scoring rule. The table displays the payoff of each (integer) reported probability q_1 between 0% and 100% when first mover chooses helpful and when she chooses hurtful. Again, it is explained that truthful reporting yields the highest expected payoff and that further mathematical insight in the table is not needed.

Only about half of the second movers report their belief. By comparing the choices of second movers reporting beliefs with the choices of second movers not reporting beliefs it can be assessed

whether an undesired effect exists of formulating beliefs on choices. For example, it could be that subjects who formulate beliefs rationalize their choices.[5]

Subjects

Subjects are recruited at the University of Amsterdam. In total 112 subjects participate in this experiment. The experiment is carried out in 8 sessions of 12 or 16 subjects each: 56 subjects participate in Nature and 56 participate in Flesh and Blood; 16 of the 28 second movers in Flesh and Blood report their belief about the choice of their paired first mover.[6] Reading the instructions, playing the game and filling out the questionnaire takes about 15-20 minutes. For their decisions in the hot response game subjects earn on average 10 guilders and for reporting their beliefs they earn 1.30 guilders.[7]

3. Results

This section is organized as follows: first, the effects of positive and negative intentionality on reciprocal behavior are compared. Then it is investigated whether emotions play the intermediate role suggested by the self-serving bias. Finally, the focus will be on subjects' beliefs.

There exists only a mild effect of positive intentionality in the hot response game. As can be inferred from table 2, second movers are only 25% more likely to reciprocate an intentional helpful choice over an unintentional helpful choice. The difference misses conventional levels of significance ($p=0.15$ for the Mann-Whitney test). In contrast, the effect of negative intentionality appears to be pretty strong. Second movers are 67% more likely to reciprocate an intentional hurtful choice over an unintentional hurtful choice. This difference is significant ($p<0.01$ for the Mann-Whitney test). The effect of negative intentionality is further underlined by the result that hurtful choices in Flesh and Blood never trigger rewards, while still 25% of the hurtful choices in Nature are followed by a reward.[8] Negative intentionality matters more than positive intentionality.

---

[5]A similar problem cannot occur for first movers. First movers only report their beliefs after they have made a choice. At the time of choosing they are not aware of the fact that they will be asked to report their beliefs.

[6]Prior to this experiment subjects participate in a pyramid game. There is no connection between the two experiments: in the pyramid game, there is no potential role for reciprocity and subjects only receive information about their own payoffs.

[7]One US dollar is worth about two Dutch guilders.

[8]There does not appear to be a systematic difference between choices of second movers who report beliefs and those who do not report beliefs. Therefore, all choices of second movers are pooled in the analysis.

Table 2
Asymmetry of punishments and rewards

	punish			cool			reward		
	Nature	Flesh and Blood	Mann-Whit-ney Z	Nature	Flesh and Blood	Mann-Whit-ney Z	Nature	Flesh and Blood	Mann-Whitney Z
helpful	0% (0/16)	0% (0/16)	0.00	50% (8/16)	25% (4/16)	-1.44	50% (8/16)	75% (12/16)	-1.44
hurtful	16.7% (2/12)	83.3% (10/12)	-3.20**	58.3% (7/12)	16.7% (2/12)	-2.06*	25% (3/12)	0% (0/12)	-1.81

*Notes: * indicates significance at 5% level; ** at 1% level.*

This does certainly not imply that intentionality is the only cause for deviations from selfishness. If that were true, no reciprocal choices should be observed in Nature. Even in Nature in 50% of the cases second mover does not play cool. Distributional considerations matter. It is less clear how exactly distributional concerns matter. The 50% rewards after a helpful choice in Nature are consistent both with altruism and with dislike of unequal payoffs. The 16.7% punishments after a hurtful choice in Nature are consistent with dislike of unequal payoffs but not with altruism. The 25% rewards in Nature after a hurtful choice are consistent with altruism but not with dislike of unequal payoffs.

How should the asymmetric effect of positive and negative intentionality be explained? In the following analysis a positive emotion is calculated as the average of the emotions reported for gladness, relief, appreciation and feeling lucky; a negative emotion is calculated as the average of the emotions reported for anger, annoyance, disappointment and contempt (*cf.* Sonnemans, 1991, p.149 and p.189-190). Emotions are rated on a 0,..,7 scale (0 denotes no emotion, 1 denotes emotion with little intensity and 7 with much intensity). The following classification is used: if the average emotion equals 0, there is 'no emotion', if the average emotion is larger than 0 but smaller than or equal to 3.5, it is classified as a 'weak emotion', and if the average emotion is larger than 3.5, it is classified as a 'strong emotion'.

Naturally, helpful choices tend to lead to positive emotions and hurtful choices tend to lead to negative emotions in Nature as well as in Flesh and Blood (see table 3). If the self-serving bias forms the foundation of an asymmetry between positive and negative intentionality, one would expect that the intentionality of a helpful choice matters less for the (intensity of the) positive emotion experienced than that the intentionality of a hurtful choice matters for the (intensity of the) negative emotion. This is supported by the data. A helpful choice always leads to a positive emotion. There is no systematic difference between the intensity of a positive emotion in Flesh and Blood and in Nature (Mann-Whitney rank test: m=16, n=15; p=0.89). A hurtful choice triggers more often a negative emotion in Flesh and

Blood. If an emotion is triggered, it is always a weak emotion in Nature, while the larger part is a strong emotion in Flesh and Blood. In fact, after a hurtful choice a significantly stronger negative emotion is triggered in Flesh and Blood than in Nature (Mann-Whitney rank test: m=12, n=11; $p<0.01$).

Table 3
Effect of choice first mover on emotion second mover

		helpful choice		hurtful choice	
		Nature	Flesh and Blood	Nature	Flesh and Blood
positive emotion	no	0%	0%	90%	75%
	weak	60%	56.3%	10%	25%
	strong	40%	43.8%	0%	0%
	total	n=15	n=16	n=10	n=12
negative emotion	no	100%	100%	18.2%	8.3%
	weak	0%	0%	81.8%	33.3%
	strong	0%	0%	0%	58.3%
	total	n=15	n=11	n=16	n=12

Table 4 shows the consequences of the second movers' experienced emotions for their choices. A higher intensity of a positive emotion increases the likelihood of a reward both in Nature and in Flesh and Blood. Similarly, a higher intensity of a negative emotion increases the likelihood of a punishment both in Nature and Flesh and Blood. Nevertheless, subjects show more restraint in responding to their emotions when the choice of first mover is unintentional. Given an emotion, they reciprocate less in Nature than in Flesh and Blood.

Table 4
Effect of emotion second mover on choice second mover

		Nature				Flesh and Blood			
		reward	cool	punish	total	reward	cool	punish	total
positive emotion	no	22.2%	66.7%	11.1%	n=9	0%	22.2%	77.8%	n=9
	weak	40%	50%	10%	n=10	58.3%	16.7%	25%	n=12
	strong	50%	50%	0%	n=6	71.4%	28.6%	0%	n=7
negative emotion	no	41.2%	52.9%	5.9%	n=17	70.6%	23.5%	5.9%	n=17
	weak	33.3%	55.6%	11.1%	n=9	0%	25%	75%	n=4
	strong	--	--	--	n=0	0%	14.3%	85.7%	n=7

Subjects' reported beliefs are also useful to explain their behavior. The beliefs of second mover help to evaluate a potential effect of surprise on the intensity of the emotion experienced. A stronger negative emotion may be experienced if the second mover expects a helpful choice and then learns that first mover has actually chosen hurtful (*cf.* Frijda, 1986, p.292). The data for hurtful choices suggest that subjects experience stronger negative emotions when they estimate the probability of a helpful choice to be higher (Spearman rank correlation coefficient between reported probability of a helpful choice and negative emotion is 0.65; n=6; p=0.16). Likewise, a helpful choice may trigger a stronger positive emotion if it is not expected. For helpful choices the reported positive emotion scarcely decreases when the probability of a helpful choice increases (Spearman rank correlation coefficient between reported probability of a helpful choice and positive emotion is only -0.17; n=10; p=0.64). Surprise seems to matter more for negative emotions than for positive emotions. On average second movers are remarkably accurate when predicting the choice of first mover. The average reported probability of a helpful choice is 56.8%, while in Flesh and Blood 57.1% of the first movers actually choose the helpful choice.

The beliefs of first movers help to evaluate whether they anticipate the effects of intentionality. Table 5 reports the beliefs, expected and realized payoffs of first movers. In Nature first movers have on average unbiased beliefs about the choices of second movers. The accurateness of first movers in Nature contrasts sharply with the biases in first movers' beliefs in Flesh and Blood. There first movers substantially underestimate the probability of a hot, reciprocal response. It is nevertheless striking that first movers seem to anticipate some degree of asymmetry between positive and negative intentions. First movers provide similar estimates for the probability of a reward after a helpful choice in both treatments. On the other hand, first movers in Flesh and Blood estimate the probability of a punishment after a hurtful choice to be higher than first movers in Nature.

Table 5
Realized and expected payoffs and beliefs first mover

	choice first mover	payoff first mover		choice second mover	probability first mover		Wilcoxon Z
		expected	realized		reported	actual	
Nature	helpful	9.64	10.00	reward	46.4%	50.0%	-0.26
				cool	48.3%	50.0%	-0.26
				punish	5.4%	0.0%	-2.37*
	hurtful	11.58	11.33	reward	30.3%	25.0%	-0.13
				cool	54.0%	58.3%	-0.39
				punish	15.8%	16.7%	-0.31
Flesh and Blood	helpful	9.28	11.00	reward	46.6%	75.0%	-3.36**
				cool	39.1%	25.0%	-2.78**
				punish	14.4%	0.0%	-2.93**
	hurtful	9.94	7.67	reward	10.2%	0.0%	-2.52*
				cool	53.2%	16.7%	-2.98**
				punish	36.7%	83.3%	-2.98**

*Notes: the expected payoff for first mover is calculated on the basis of her choice and her reported probabilities. The Wilcoxon rank test compares reported probabilities with actual probabilities. Actual probabilities are computed on the basis of second movers' choices. ** indicates significance at 1% level; * at 5% level.*

In Flesh and Blood first movers expect on average a slightly higher payoff from the hurtful choice than from the helpful choice. The actual responses of second movers are not in accordance with this belief. Helpful choices are more remunerative than hurtful choices. The bias in judging the effect of intentionality is costly for first movers.

4. Conclusions

Previous experimental work on the self-serving bias has been carried out in a natural rich context. It was believed that the self-serving bias was less likely to manifest itself in a pronounced way in an experiment designed along the clean, context free lines of the experimental methodology typically used in economics (*e.g.*, Babcock *et al.*, 1995). This belief is perhaps a bit pessimistic: even in the abstract setting of the present experiment a pronounced effect of the self-serving bias is observed.

This paper evaluates an implication of the self-serving bias for reciprocity. People with a positive self-image may not be so impressed by an intentional helpful choice. After all, someone as nice as yourself deserves to be treated well. In principle this is nothing special. In accordance with this

conjecture, subjects experience equally positive emotions after a helpful choice in Nature and in Flesh and Blood. Nevertheless, subjects show more restraint in Nature than in Flesh and Blood when responding to a positive emotion. This explains the mild effect of positive intentionality on the probability of reciprocation. The story is different for an intentional hurtful choice. A hurtful choice contrasts sharply with a positive self-image. People easily feel insulted after a hurtful choice. In accordance with this conjecture, subjects experience stronger negative emotions when a hurtful choice is intentional than when it is unintentional. In addition, they show more restraint in responding to their emotion when the hurtful choice was unintentional. Both an increase in the negative emotion experienced and a decrease in the restraint of responding to the emotion explain the strong effect of negative intentionality on the probability of reciprocation. Negative intentionality matters considerably more than positive intentionality.

By and large the evidence of this experiment turns the balance in favor of a nice view of people when intentionality does not play a role. In Nature forced helpful choices are often rewarded but never punished. Hurtful choices are sometimes punished but also sometimes rewarded. Overall, people are more often nice than mean in this treatment. This effect is more than offset when intentionality is introduced: in Flesh and Blood intended helpful choices are often rewarded and never punished; on the other hand, hurtful choices are more often punished and never rewarded. When intentionality plays a role, an uglier view of mankind results. This finding may shed some light on results for other games that may seem paradoxical at first sight. For example, it may help explain why public good games elicit nicer behavior than ultimatum games. The simultaneous choice structure of most public good games diminishes the role of intentionality, whereas the sequential choice structure of ultimatum games encourages the impact of intentionality.

A remarkable result is that first movers do not sufficiently anticipate the magnitude of the effect of the intentionality of their choice on the probability of reciprocation. This especially holds for negative intentionality. This finding has some relevance for people in bargaining type of situations. When forming beliefs about the behavior of their opponent, they should be careful not to overestimate the probability of a cool, self-interested response.

An important question is whether it pays to have a self-serving bias. The self-serving bias can be useful in some situations. For example, it has been argued that the self-serving bias has survival value for people coping with serious illnesses (Brown, 1986, p. 161-167). On the other hand, in bargaining games the self-serving bias may lead to costly impasses (Babcock and Loewenstein, 1997), and in the hot response game it can lead to costly punishments of hurtful actions. These short term costs of the self-serving bias may be offset by long term benefits. A hot response after a hurtful choice may give the reciprocator a reputation of being tough. This may prevent future hurtful choices. Particularly interesting in this respect is that first movers seem to anticipate some effect of negative intentionality: the probability of a punishment is estimated to be higher after an intentional hurtful

choice than after an unintentional hurtful choice. A similar effect of positive intentionality is not expected by first movers.

References

Akerlof, G.A., 1982: 'Labor Contracts as a Partial Gift Exchange', *Quarterly Journal of Economics* 97, 543-569.

Akerlof, G.A., and J.L. Yellen, 1988: 'Fairness and Unemployment', *American Economic Review* 78, 44-49.

Akerlof, G.A., and J.L. Yellen, 1990: 'The Fair-Wage Effort Hypothesis and Unemployment', *Quarterly Journal of Economics* 195, 255-284.

Babcock, L., and G. Loewenstein, 1997: 'Explaining Bargaining Impasse: The Role of Self-Serving Biases', *Journal of Economic Perspectives* 11, no 1, 109-126.

Babcock, L., G. Loewenstein, S. Issacharoff and C. Camerer, 1995: 'Biased Judgments of Fairness in Bargaining', *American Economic Review* 85, 1337-1343.

Babcock, L., X. Wang, and G. Loewenstein, 1996: 'Choosing the Wrong Pond: Social Comparisons that Reflect a Self-Serving Bias', *Quarterly Journal of Economics* 111, 1-19.

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

Bolton, G.E., J. Brandts and E. Katok, 1996: 'A Simple Test of Explanations for Contributions in Dilemma Games', *Working Paper*, Penn State University.

Bolton, G.E., J. Brandts and A. Ockenfels, 1997: 'Measuring Motivations for the Reciprocal Responses Observed in a Simple Dilemma Game', *Working Paper* Nr. 34, Otto-von-Guericke University of Magdeburg.

Bolton, G.E., and A. Ockenfels, 1997: 'ERC- A Theory of Equity, Reciprocity and Competition', *Working Paper*, Penn State University.

Brandts, J., and G. Charness, 1998: 'Hot vs. Cold: Sequential Responses and Preference Stability in Experimental Games', *Working Paper* Nr. 424.98, Universitat Autonoma de Barcelona.

Brown, R., 1986: *Social Psychology*, 2nd edition. The Free Press: New York.

Camerer, C., 1997: 'Progress in Behavioral Game Theory', *Journal of Economic Perspectives* 11 (4), 167-188.

Charness, G., 1998: "Attribution and Reciprocity in a Simulated Labor Market: An Experimental Investigation', *Working Paper*, Universitat Pompeu Fabra.

Dufwenberg, M., and G. Kirchsteiger, 1998: 'A Theory of Sequential Reciprocity', *Working Paper*, Tilburg University.

Fehr, E., G. Kirchsteiger and A. Riedl, 1993: 'Does Fairness Prevent Market Clearing? An Experimental Investigation', *Quarterly Journal of Economics* 108, 437-459.

Fehr, E., G. Kirchsteiger and A. Riedl, 1998: 'Gift Exchange and Reciprocity in Competitive Experimental Markets', *European Economic Review* 42, 1-34.

Fehr, E., and K. Schmidt, 1997: 'A Theory of Fairness, Cooperation and Competition', *Working Paper*, University of Zurich.

Frijda, N.H., 1986: *The Emotions*. Cambridge University Press: Cambridge.

Levine, D.K., 1997: 'Modeling Altruism and Spitefulness in Experiments', *Working Paper*, UCLA.

Lewinsohn, P.M., W. Mischel, W. Chaplin and R. Barton, 1980: 'Social Competence and Depression: The Role of Illusory Self-Perceptions', *Journal of Abnormal Psychology*, 89, 203-212.

McKelvey, R.D., and T. Page, 1990: 'Public and Private Information: An Experimental Study of Information Pooling', *Econometrica* 58, 1321-1339.

Offerman, T., J. Sonnemans and A. Schram, 1996: 'Value Orientations, Expectations, and Voluntary Contributions in Public Goods', *Economic Journal* 106, 817-845.

Offerman, T, 1997: *Beliefs and Decision Rules in Public Good Games -Theory and Experiments-*. Kluwer: Dordrecht.

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

Sedikides, C., W.K. Campbell, G.D. Reeder and A.J. Elliot, 1998: 'The Self-Serving Bias in Relational Context', *Journal of Personality and Social Psychology* 74, 378-386.

Sonnemans, J., 1991: 'Structure and Determinants of Emotional Intensity', *PhD dissertation*, University of Amsterdam.

Sonnemans, J., and N.H. Frijda, 1995: 'The Determinants of Subjective Emotional Intensity', *Cognition and Emotion* 9, 483-506.

Sonnemans, J., and T. Offerman, 1998: 'Is the Quadratic Scoring Rule Behaviorally Incentive Compatible?', *Working Paper*, University of Amsterdam.

Sugden, R., 1984: 'Reciprocity: The Supply of Public Goods Through Voluntary Contributions', *Economic Journal* 94, 772-787.

Svensson, O., 1981: 'Are We All Less Risky and More Skillful than our Fellow Drivers?', *Acta Psychologica* 47, 143-148.

Wills, T.A., 1981: 'Downward Comparison Principles in Social Psychology', *Psychological Bulletin* 90, 245-271.