Can you trust the good guys?

by Sebastian Fehrler* and Michael Kosfeld**

November 2009, Zurich

...we receive a much greater satisfaction from the approbation of those, whom we ourselves esteem and approve of, than those, whom we hate and despise.

David Hume (1739)

Abstract

The strive for social esteem is an important motive for pro-social behavior. Many people want to be seen as nice. Recent theories have suggested that the valuation of such esteem depends on the audience. In this study we look at trust and trustworthiness towards people who do or do not identify themselves with typical altruistic goals. Those who do are the good and those who do not are the bad audience. In a trust game we observe strong discrimination against the bad audience and no positive discrimination at all of the good audience. Moreover, we find that only those second movers who identify themselves with the goals discriminate between the audiences. The last two findings cannot be explained by existing theories.

*University of Zurich, IZA, and Center for Comparative and International Studies (CIS), Email: sebastian.fehrler@pw.uzh.ch.

**University of Frankfurt and IZA.

1 Introduction

In many situations we have to trust people we do not know much about. Can we infer something about a person's trustworthiness from her identification with some altruistic goals, possibly revealed by an Amnesty International T-shirt or a "Save the Whales" badge on the car? Recent theories suggest that this could depend to a large degree on what that other person herself thinks about us.

The strive for social esteem has long been emphasized as an important motive for prosocial behaviour by classical thinkers as David Hume or Adam Smith (see discussion in Ellingsen and Johannesson 2008). Newer theories of prosocial behaviour have taken up their ideas. In the theories by Bénabou and Tirole (2006) and by Ellingsen and Johannesson (2008) the motive of social reputation is central for acting prosocially. The latter theory adds the idea that the valuation of such esteem depends on the audience, an idea which makes it to some extent similar to Levine's theory of altruism and spitefulness in which an agent's utility from another agent's payoff can depend on the (expected) degree of altruism of the other agent (Levine 1998).

In this study we look at trustworthiness towards audiences of different degrees of niceness. As audiences we consider subjects who do or do not identify themselves with typical altruistic goals like the goals of Amnesty International (AI) and the World Wildlife Fund (WWF). In a simple trust (investment) game subjects can make their transfer decisions conditional on this information about the other player, which is elicited with a short survey before the experiment. In a control treatment designed to control for mere in-group effects, subjects can condition their transfers on the art preferences of the other subject.

We are interested in the following questions. Are the good guys (those who identify themselves with the goals of AI or the WWF) really more trustworthy (nicer) than average? Do (all) subjects discriminate between nice (good) and less nice (bad) subjects, and if there is discrimination, is it driven by negative or by positive discrimination? How does trust, i.e. first mover behavior, depend on the type of the trustor and the type of the trustee?

A number of field and laboratory studies have revealed that people show more prosocial activity in public than in private settings. Gächter and Fehr (1999) show that relaxing anonymity can increase contributions in a public good game. Ariely et al. (2009) demonstrate the importance of social approval for charity in a laboratory and in a field experiment. Futher experimental studies pointing in the same direction are Andreoni and Petrie (2004), Dana et al. (2006), Rege and Telle (2004), and Soetevent (2005). In all these studies it appears that agents value social esteem and expect their prosocial behavior to be esteemed of by the

audience.

The idea that the valuation of such esteem depends on the audience is old as the initial quote from David Hume shows. Ellingsen and Johannesson (2008) formalize this idea and use it to explain behavior in economic experiments in which information on the audience can be inferred from behavior earlier in the game. They discuss, for example, an experiment by Falk and Kosfeld (2006) who show the detrimental effect of controlling an agent's actions by restricting her choice set. This signals lacking trust and makes the controlling principal a worse audience for the agent than a trusting principal. As a consequence the outcomes for a controlling principal are worse even though the control mechanism is available at no cost. Strong negative responses to not so nice behavior have been observed in other experiments as well. Fehr and Gächter (2000), for example, show a strong willingness to punish free riders in public good games. Fehr and List (2004) and Fehr and Rockenbach (2003) show the negative effect of choosing contracts with possible sanctions on trustworthiness.

Discrimination which is not influenced by observed behavior against people outside one's own group even if the group is arbitrarily formed is a well known phenomenon in social psychology, called the "minimal group paradigm".[1] Recent studies addressing this issue are Chen and Li (2009) and Charness et al. (2007). We control for the minimal group effect with our control treatment.

In Ellingsen and Johannesson's theory a good audience is one with altruistic or other prosocial preferences (e.g. for fair outcomes). Similarly, in Levine's theory a player's utility in a two player game depends on own income and the other player's income multiplied by a factor which depends on one's own degree of altruism and (one's belief about) the other player's degree of altruism. It is reasonable to assume that someone who identifies herself with altruistic goals is more likely to have preferences for positive reciprocity, altruism or outcome based fairness concerns (all motives for trustworthy behavior) and is, therefore, a better audience.

There is evidence from another experiment that people are nicer to nice people and in which the information about a person's niceness does not come from observed behavior in the same experiment. Albert et al. (2007) study cooperative behavior in a prisoners' dilemma and a dichotomous trust game of subjects who could donate money to an NGO of their choice in an earlier experiment. They observe that subjects cooperate more often with subjects who donated more money. This result is in line with the theories. However, behavior of both

[1] A classic study is Tajfel et al. (1971). See Crisp and Turner (2007) for a textbook presentation, or Brewer (1979) and Mullen et al. (1992) for reviews of empirical studies on the minimal group paradigm.

players in the prisoners' dilemma game and of first movers in the trust game does not only depend on their preferences about outcomes for different audiences but also on their beliefs about the other player's behavior. Both can be influenced by an information about the other player's niceness and the two channels are not not studied separately.

In our experiment we study behavior of different types of second movers for whom beliefs about how their actions might affect the behavior of the other player do not play a role. We compare their backtransfers to different types of first movers conditional on all possible transfers from these players. The differences of the niceness of the audience solely comes from the survey information and not from observed behavior. We also study first mover behavior and their beliefs to see in how far they anticipate second mover behavior, and in how far trust depends on the type of the trustor and the trustee.

We report the following main results. The good guys, i.e. the second movers who identify themselves with the goals of either NGO, are indeed on average significantly more trustworthy. However, they strongly discriminate between the audiences and transfer back substantially more to first movers who identify themselves with either NGO, as well, i.e. to the good audience. The comparison to our control treatment shows, that the difference in the backtransfer levels entirely stems from negative discrimination against the bad audience and not from positive discrimination of the good audience. Moreover, we find that second movers who do not identify themselves with the NGO goals do not discriminate at all. The last two findings contradict Ellingsen and Johannesson's and Levine's theories, which predict to see positive as well as negative discrimination of different audiences and that all groups of subjects discriminate.

The paper proceeds with the experimental design in Section 2, a detailed presentation of the results in Section 3, and the conclusion in Section 4.

2 Experimental Design

We now turn to the experimental design and make a number of predictions of the outcomes of our experiment.

2.1 Trust Game

The subjects play a standard trust (investment) game. Half of the subjects are first the other half second movers. All recipients receive an initial endowment of 12 points. First movers can transfer 0, 4, 8 or 12 points to the second mover. The transfers are tripled. The second

movers can then send back any integer amount of points from the points they have back to the first mover. Backtransfers are not tripled. After the backtransfers the experiment ends and the subjects are paid out. The experiment consists of only one round.

In the beginning, before distributing instructions for the trust game, the subjects are asked to fill out a short questionnaire on their computer screens. The questionnaire includes questions like "Do you do sports?", "Do you play an instrument?" and the question "Do you strongly identify yourself with the goals of one of the NGOs, Amnesty International or the WWF?". The last question is the one we are interested in in our main treatment. It has the following answer options: "WWF", "Amnesty International" and "None of the two". One answer option has to be checked and multiple answers are ruled out.

In the control group setting we use a different question from the same questionnaire to form groups: "Do you very much like one of the painters: Paul Klee or Wassily Kandinski?" with answer options, "Klee", "Kandinsky" and "None of the two". This setting is designed to control for mere in-group effects. We relate to the classic social psychology study in this field by Tajfel et al. (1971) in which preferences about Klee and Kandinski are used as well to form "minimal" groups. As these art preferences do not carry any information about prosociality, no differences between the transfer and backtransfer levels to different subjects should be observed above a possible minimal group effect. The questionnaire is designed to give the subjects the impression that they take part in a small socioeconomic survey. This makes it unlikely that they expect their answers to play a role in the experiment.

In the trust game first movers and second movers can make their transfer decision conditional on the type of the recipient. In the main treatment they can make their decisions dependent on the answer of their partner to the NGO question in the control treatment on the answer to the art question.

2.2 Procedural Details

The trust game is played with the strategy method. First movers make three transfer decisions, one for each potential type of second mover. Second movers make twelve decisions, one for each possible first mover type and received transfer.

One point in the trust game is worth 0.8 Swiss Francs. Overall, 190 subjects participated in the experiment in the laboratory of the Institut für empirische Wirtschaftsforschung (IEW) at the University of Zurich.[2]

[2]The treatments were programmed with zTree (see Fischbacher 2007).

2.3 Theoretical Predictions

From the models by Ellingsen and Johannesson (2008) and Levine (1998) we derive a number of prediction for our experiment. In Ellingsen and Johannesson's model agents derive utility from their own pay-off, from the other player's payoff, and from the pride they take in the other player's esteem of their prosociality. They specifically consider altruism but other forms of prosociality can be modeled similarly (as demonstrated in an earlier version of their paper, Ellingsen and Johannesson 2006).

In their model, which we slightly simplify here, agents maximize the following utility function

$$u_i = m_i + \theta_i m_j + \widehat{\theta}_{ji}$$

with m_i denoting player i's material payoff, θ_i the degree of altruism, with $1 > \theta_i > 0$, and $\widehat{\theta}_{ji}$ player i's pride, defined as

$$\widehat{\theta}_{ji} = E_{\theta_j}[\sigma(\theta_j)\theta_{ji}]$$

with $\sigma(\theta_j)$ being the salience of the opponent's esteem and θ_{ji} the esteem of player j of player i, defined as

$$\theta_{ji} = E[\theta_i|h]$$

with h denoting the history of the game.

The valuation of esteem, $\sigma(\theta_j)$, depends on the expected degree of altruism of the other player. It is assumed that σ is increasing in θ_j, that is, esteem from more altruistic agents is valued higher. In the original version Ellingsen and Johannesson allow for biased beliefs, in the sense that agents expect other agents to be similar to them.

In Levine's (1998) theory the utility of an agent in a two player game is given by the following (slightly simplified) function

$$u_i = m_i + (\theta_i + \lambda\theta_j)m_j$$

where m and θ have the same meanings as above and $\lambda \geq 0$ is a parameter common to all players.

Both theories predict that nice people are treated better (if $\widehat{\theta}$ or λ are greater than zero). They also predict that all subjects discriminate between good and bad audiences (neither the valuation of the esteem nor the valuation of the other player's altruism depends on the

player's own type). Moreover, both theories predict that nice people are treated better and less nice people treated worse than average nice people.

In our setting niceness takes the form of trustworthiness which can be motivated by altruism, positive reciprocity or fairness concerns.

As we compare backtransfers in the trust game for each transfer level, the history h of the game does not play a role. Another factor possibly influencing the expectation of the other player's type is the information on her answer to the NGO question. We can think of this information as also being represented by the term h.

We predict that subjects show higher trustworthiness towards a good audience, i.e. towards subjects who identify themselves with either NGO. We predict to see this pattern for all second movers. Moreover, as we expect the information on identification to increase the expected θ of the other player, as well as, the contrary information decreases it, we predict to see higher (lower) trustworthiness towards the good (bad) audience as compared to trustworthiness in the control treatment.

We also expect first movers to trust nice second movers more than not so nice second movers and, therefore, transfer higher amounts to them. This is crucial as it is also a test of the assumption that the good guys are indeed perceived as a better audience.

3 Results

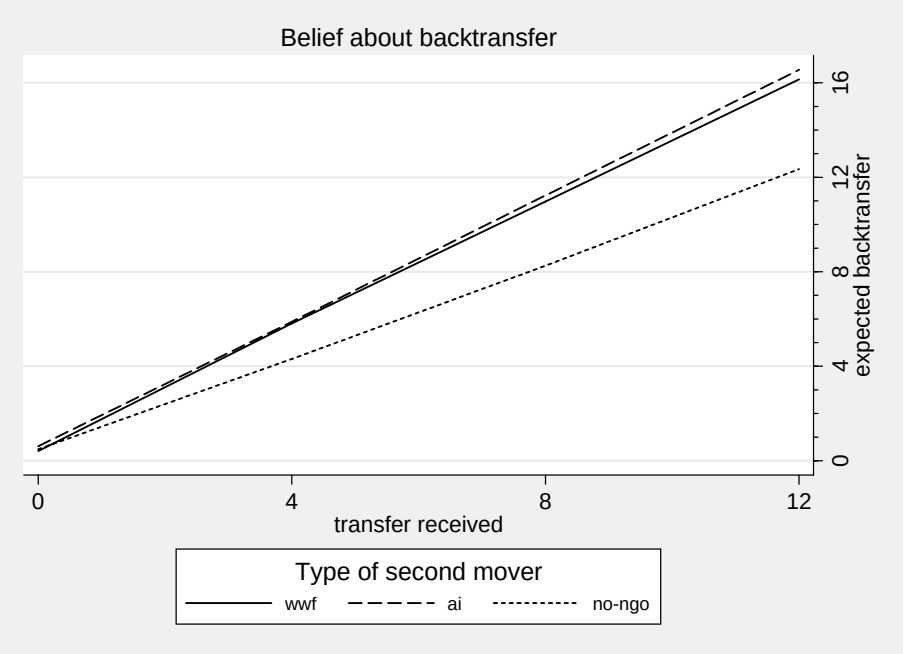
We now turn to the presentation of the results. Overall, there are 32% subjects who identify themselves strongly with the goals of the WWF, 26% with the goals of Amnesty and 42% with neither NGO's goals. There are 29% subjects who state to like Klee, 22% to like Kandinski and 49% who indicate to like neither painter.

3.1 First Mover Behaviour

We asked the first movers about their beliefs regarding backtransfers for all possible transfer levels and types of second movers. In Figure 1 we see that first movers expect lower backtransfers from subjects who do not strongly identify themselves with the goals of either NGO, henceforth called No-NGO types. Looking at different types of first movers separately, shows that this is true for all types of first movers (see Figures 5 in the Appendix).

Moreover, we see that the beliefs about backtransfers from AI or WWF types are almost the same. Table 1 shows the transfer levels to the different types of second movers by the different types of first movers. The differences between the transfer levels reflect the beliefs

Figure 1: Beliefs about trustworthiness from different second movers (who make the backtransfer).



about the backtransfers. Even the No-NGO types transfer less to other No-NGO types, than to AI or WWF types. For the No-NGO types the differences of the transfer levels to the three second mover types are pairwise statistically different at the 5% level (Wilcoxon rank sum test).[3] For the other two groups the transfer level to No-NGO types is statistically different at the 5% level from the other two groups which themselves are not significantly different from each other. Transfers to No-NGO types are lower than to any other group. The NGO types receive, on average, 47% higher transfers than No-NGO types. This shows that NGO types are believed to be more trustworthy than No-NGO types and NGO types are, therefore, a better audience for second movers seeking esteem for their trustworthiness.

Table 2 shows the transfer levels in the control group. Here, each type of first mover favors second movers with the same art preferences but there is no group nobody trusts less than all other groups.

[3]All test results we report are for undirected hypotheses.

Table 1: Transfer levels from different NGO types to different NGO types.[a]

Transfer	to WWF	to AI	to No-NGO	N
from WWF	8.3 (0.8)	7.9 (0.8)	4.6 (0.9)	28
from AI	7.1 (1.2)	8.3 (1.2)	4.6 (1.8)	14
from No-NGO	6.8 (0.7)	7.6 (0.7)	5.9 (0.8)	36

[a] We use NGO type and then just the NGO name as abbreviations for subjects strongly identifying themselves with the goals of that NGO. Standard errors in parentheses.

Table 2: Transfer levels from different artist types to different artist types.[a]

Transfer	to Klee	to Kandinski	to No-Artist	N
from Klee	8.6 (1.0)	6.6 (1.1)	6.4 (1.1)	22
from Kandinski	7.6 (0.9)	8.8 (0.8)	6.7 (1.1)	19
from No-Artist	6.3 (1.0)	6.2 (1.0)	8.3 (0.9)	26

[a] We use artist type and then just the artist's name as abbreviations for subjects liking the work of the artist a lot. Standard errors in parentheses.

3.2 Second Mover Behavior

In the analysis of second mover behavior we start by looking at the trustworthiness of the different NGO types in the control group setting where they cannot condition their backtransfer on the NGO type of the first mover. This allows us to see whether the NGO types are more trustworthy in general. In the control treatment the transfers have to be conditioned on the art preferences of the first mover. As we used the same questionnaire for both control and treatment group we can group the control group results by the answers to the NGO question. Figure 2 presents the backtransfers for the different potential transfers averaged over the three potential recipient types ("Klee", "Kandinski" and "No-Artist").

We see that people who identify themselves with one of the NGOs are more trustworthy than people who do not, just as first movers expect. Regressing backtransfer on transfer gives significantly different slopes for the AI group when compared to the No-NGO group (at the 5% level). In this regression there are four observations from every first mover, one for each

Figure 2: Trustworthiness of different NGO types.[a]

[a] Average backtransfers from the control treatment, in which transfers could not be made conditional on the NGO type of the receiver. N=67.

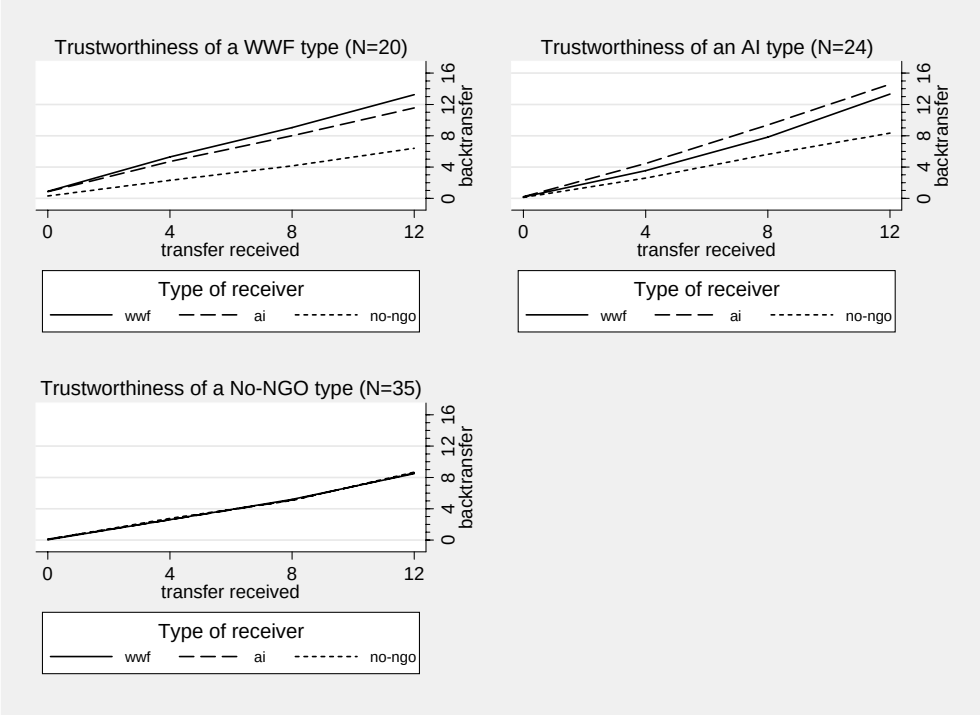
possible transfer level. This is taken into account in the estimation of the standard errors by treating these four observations as one cluster each. The difference between the slopes is tested using an adjusted Wald test. Pooling the AI and the WWF group in the regression gives a significantly different slope of this NGO group to the No-NGO group slope (at the 5% level). The backtransfer after a transfer of 12 is also significantly different between the WWF and the No-NGO group (at the 10% level, Wilcoxon rank sum test). The slope of the WWF group alone is not significantly different to the slope of the No-NGO group.

This tells us that subjects who identify themselves with an NGO are indeed good in the sense of being more trustworthy. The backtransfers they make are about 1.5 times the backtransfers of the No-NGO types.

Do second movers discriminate when they face different audiences? An answer to this question is given by the three graphs in Figure 3.

We see that NGO types, the good guys, do indeed strongly discriminate against No-NGO types. The slopes of the regression lines when regressing backtransfer on transfer by first mover type are significantly different from each other in case of an Amnesty and an WWF

Figure 3: Trustworthiness of different NGO types towards different first movers.

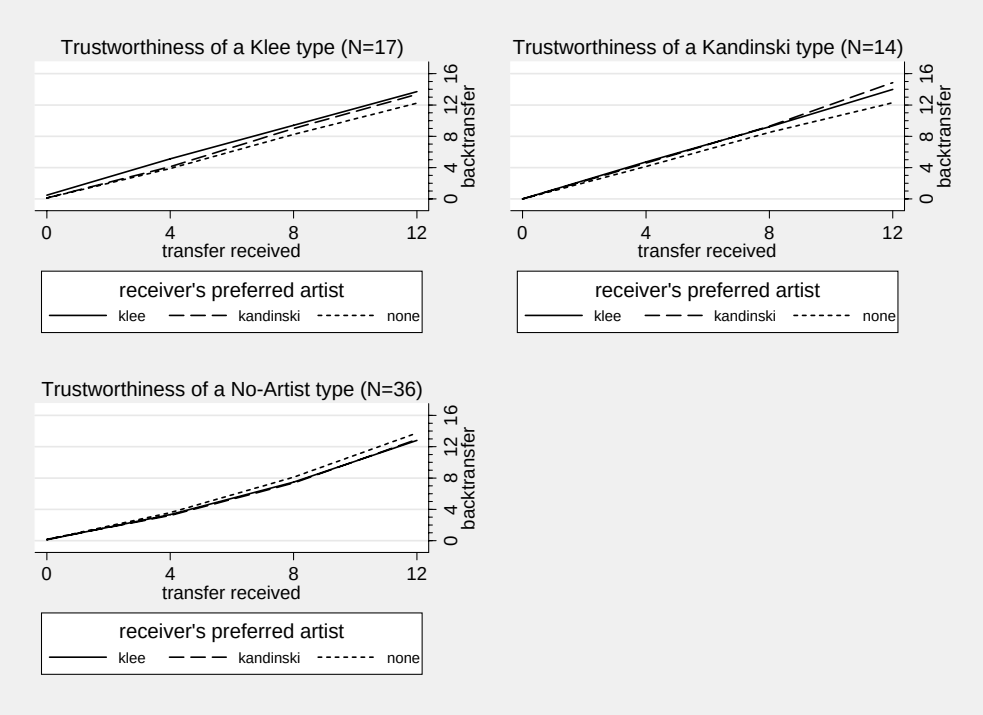


second mover (at the 5% level, adjusted Wald tests).[4] It is also the case that AI types favour other AI types over WWF types, and WWF types favour other WWF types over AI types. This can be explained by the mere in-group effect which is present even when group formation is arbitrary as in our control treatment. Figure 4 shows that there are small differences between the artist types at high transfer levels. These differences are tiny, though, compared to the difference between the good and the bad audience group.

Backtransfers to either NGO type first mover from either NGO type second mover are, on average, 1.8 times higher than backtransfers to No-NGO types. This difference is large and in line with our first theoretical prediction that the different audiences will be treated differently. However, we find that No-NGO types (third picture in Figure 3) do not discriminate between the audiences. This contradicts our prediction to observe the same pattern for all second movers.

[4]In these regressions there are again four observations from every first mover, one for each possible transfer level. This is taken into account in the estimation of the standard errors by treating these four observations as one cluster each.

Figure 4: Trustworthiness of different Artist types towards different first movers.



What drives the discrimination of the different audiences by the NGO types? The comparison of the treatment with the control group reveals a strong negative discrimination against the bad audience and no positive discrimination of the good audience. Regressing backtransfers from either NGO group (which we pool in the regression) to either NGO group on the received transfer gives slopes which are not significantly different to the slope coefficient for the NGO groups (again pooled) in the control treatment setting which is obtained by regressing their average backtransfers to the different artist types on the received transfer (at the 5% level, adjusted Wald tests). The slope coefficient in a regression of backtransfers from the pooled NGO group to No-NGO types on the received transfer, however, is significantly lower than the slope for the pooled NGO group in the control group setting (at the 5% level, adjusted Wald test).[5]

This contradicts our theoretical prediction to observe both negative and positive discrimination, because of the changed expectations about the other person's type. What is driving the differences between the trustworthiness toward different audiences is clearly negative

[5]These result also hold if NGO groups are not pooled and AI and NGO groups are looked at separately.

discrimination of the bad audience.

4 Conclusion

We conducted a simple trust (investment) game experiment to gain some insight into the old idea that the valuation of social esteem depends on the audience. Ellingsen and Johannesson (2008) formalized this idea and we derive predictions from their and Levine's (1998) theories which we test. We form different audience groups on the basis of the subjects' identification with typical altruistic goals.

In our trust game subjects can make their decisions dependent on the type of the other player they have to interact with. They are informed about whether the other player identifies herself with the goals of Amnesty International or the WWF or none of the two, an information elicited in a short survey before the experiment.

Our first finding is that the good audience, that is the subjects who stated to strongly identify themselves with the goals of an NGO, are indeed expected to be more trustworthy by the first movers. This also means that being perceived as good comes with economic benefits in form of higher trust. Prosocial activities, like charity, could therefore, have a role as a signaling device of trustworthiness. Fehrler (2009) shows that the observation of higher transfers remains unchanged if groups are build on the basis of a public voluntary donation to Amnesty International.

The main focus of this study is on second mover behavior and the next question is, therefore, whether the good guys are indeed nicer than the others. We find that subjects identifying themselves with one of the NGOs are more trustworthy than subjects who do not identify themselves with either NGO and on average transfer back substantially more. However, we also find that the good guys strongly discriminate against the bad audience (first movers who do not identify themselves strongly with either NGO). Backtransfers to the good audience recipients are on average 1.8 times higher. This difference in the treatment of the audience groups is what the theories predict.

However, backtransfers to other NGO types are not higher than average backtransfers in the control group setting. This contradicts our prediction to see both positive and negative discrimination as the expectation of the other person's trustworthiness rises or falls with the information on her type. This observation runs against our theoretical prediction.

Another interesting finding is that the not so nice guys, that is those second movers who, like the bad audience, do not identify themselves with either NGO, do not discriminate at

all. This contradicts the prediction to see the same pattern for all second movers. Neither Elligsen and Johannesson's nor Levine's theory explain this. In the first theory the pride term would have to depend on the type of the subject whose pride it describes not only on the audience's type, in Levine's theory the term capturing utility from the other subjects' payoff.

Returning to the initial question, whether one can trust the good guys, we conclude that if oneself needs to decide whether to trust somebody or not, taking into account what the other person probably thinks about oneself, i.e. taking into account one's own type, irrespectively of one's own potential actions, appears to be very important. The subjects in our experiment interestingly do not anticipate any form of discrimination from second movers.

References

Albert, M., W. Güth, E. Kirchler, and B. Maciejovsky (2007): "Are we nice(r) to nice(r) people? An experimental analysis," *Experimental Economics*, 10(1), 53–69.

Andreoni, J., and R. Petrie (2004): "Public goods experiments without confidentiality: a glimpse into fund-raising," *Journal of Public Economics*, 88(7-8), 1605 – 1623.

Ariely, D., A. Bracha, and S. Meier (2009): "Doing Good or Doing Well? Image Motivation and Monetary Incentives in Behaving Prosocially," *American Economic Review*, 99(1), 544–55.

Bénabou, R., and J. Tirole (2006): "Incentives and prosocial behaviour," *American Economic Review*, 96(5), 1652–1678.

Brewer, M. B. (1979): "In-Group Bias in the minimal intergroup situation: A cognitive-motivational analysis," *Psychological Bulletin*, 86(2), 307–324.

Charness, G., L. Rigotti, and A. Rustichini (2007): "Individual Behavior and Group Membership," *American Economic Review*, 97(4), 1340–1352.

Chen, Y., and S. X. Li (2009): "Group Identity and Social Preferences," *American Economic Review*, 99(1), 431–57.

Crisp, R. J., and R. N. Turner (2007): *Essential Social Psychology*. London: Sage.

Dana, J., D. M. Cain, and R. M. Dawes (2006): "What you don't know won't hurt me: Costly (but quiet) exit in dictator games," *Organizational Behavior and Human Decision Processes*, 100(2), 193 – 201.

Ellingsen, T., and M. Johannesson (2006): "Pride and Prejudice: The Human Side of Incentive Theory," CEPR Discussion Papers 5768, C.E.P.R. Discussion Papers.

——— (2008): "Pride and Prejudice: The Human Side of Incentive Theory," *American Economic Review*, 98(3), 990–1008.

Falk, A., and M. Kosfeld (2006): "The Hidden Costs of Control," *American Economic Review*, 96(5), 1611–1630.

Fehr, E., and S. Gächter (2000): "Cooperation and Punishment in Public Goods Experiments," *American Economic Review*, 90(4), 980–994.

Fehr, E., and J. A. List (2004): "The Hidden Costs and Returns of Incentives-Trust and Trustworthiness Among CEOs," *Journal of the European Economic Association*, 2(5), 743–771.

Fehr, E., and B. Rockenbach (2003): "Detrimental effects of sanctions on human altruism," *Nature*, 422(6928), 137–140.

Fehrler, S. (2009): "Prosocial Behavior as a Signal of Trustworthiness," Available at SSRN: http://ssrn.com/abstract=1502565.

Fischbacher, U. (2007): "z-Tree: Zurich toolbox for ready-made economic experiments," *Experimental Economics*, 10(2), 171–178.

Gächter, S., and E. Fehr (1999): "Collective action as a social exchange," *Journal of Economic Behavior & Organization*, 39(4), 341 – 369.

Hume, D. (1896): *A Treatise of Human Nature*. 3 vols. (Orig. published 1739).

Levine, D. K. (1998): "Modeling Altruism and Spitefulness in Experiments," *Review of Economic Dynamics*, 1, 593–622.

Mullen, B., R. Brown, and C. Smith (1992): "Ingroup bias as a function of salience, relevance, and status: An integration," *European Journal of Social Psychology*, 22(2), 103–122.

Rege, M., and K. Telle (2004): "The impact of social approval and framing on cooperation in public good situations," *Journal of Public Economics*, 88(7-8), 1625 – 1644.

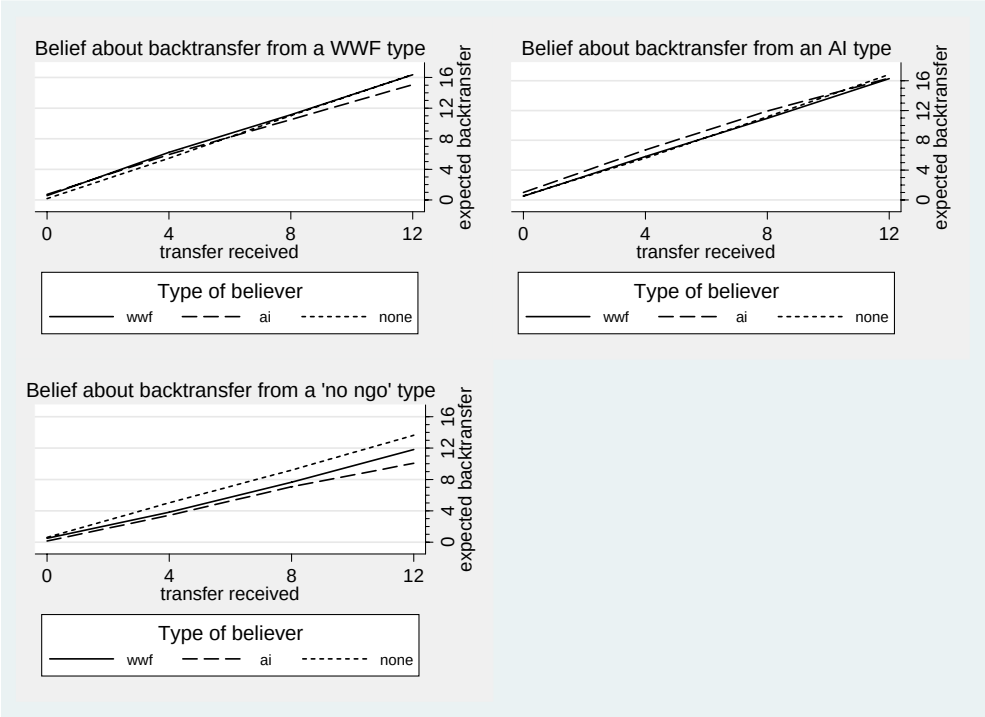
Soetevent, A. R. (2005): "Anonymity in giving in a natural context–a field experiment in 30 churches," *Journal of Public Economics*, 89(11-12), 2301 – 2323.

Tajfel, H., M. Billig, R. Bundy, and C. Flament (1971): "Social Categorization and intergroup behaviour," *European Journal of Social Psychology*, 1(2), 149–178.

Appendix

Figure 5 shows the beliefs of first movers about backtransfers from second moverws grouped by the types of first movers.

Figure 5: Beliefs of NGO types about trustworthiness of different second movers (who make the backtransfer).



Regressing the belief about the backtransfer on the first mover's transfer gives us estimates of the slopes of the different lines in Figure 5. The lines for AI and WWF type second movers are significantly steeper than the lines for No-NGO types in all three pictures (at 5% in the first two and at 10% for No-NGO type first movers). In these regressions there are four observations from every first mover, one for each possible transfer level. This is taken into account in the estimation of the standard errors by treating these four observations as one cluster each and using Taylor linearized standard errors. The difference between the slopes is tested using an adjusted Wald test.