Whom do you trust?

**Personality and previous behavior of the
counterpart as determinants of trust and
cooperation in a trust game**

B.A. Thesis
Dafina Georgieva Petrova
Spring 2010
Jacobs University Bremen

First Name : Dafina

Last Name : Petrova

Matriculation No.: 10000573

e-mail-address: d.petrova@jacobs-university.de

Major: Integrated Social and Cognitive Psychology

Supervisor: Prof. Adele Diederich

Instructor: Prof. Margrit Schreier

Word count: 6,790

Submission Date: 28.04.2010

Signature

28.04.2010

Date

Student Declaration

I hereby declare that :

1. This thesis is my own work. I have acknowledged material from the work of other people and I have clearly marked and given references to all quotations.
2. I permit the lodging of a copy of this thesis in the IRC, which shall be made available for the academic use of staff and students.

Signature

28.04.2010

Date

Contents

1. Introduction

1.1. Personality - behavior distinction

2. Theoretical background and previous findings

2.1. Types of models of cooperative behavior

2.2. Information sources

2.3. The trust game

2.4. Types of strategies in cooperative exchange

2.5. The role of personality in bargaining games: The Mach scale

2.6. Previous behavior

3. Hypotheses

4. Method

4.1. Participants and Procedure

4.2. Experimental design

5. Results

5.1. Trust

5.2. Reciprocation

5.3. Machiavellianism of the participants

5.4. Participants' comments

6. Conclusion and discussion

7. References

8. Appendix

Abstract

This experiment investigated cooperative behavior in a two-person one-shot trust game by manipulating two types of information about the players - personality and previous behavior. When information about the counterpart was consistent, participants used the simple strategy to trust and reciprocate to trustworthy individuals and do the opposite with untrustworthy ones. When information was inconsistent, mainly behavioral information determined the decision to trust. Personality information mattered more for the choice to reciprocate. Different response patterns emerged depending on whether the participants were trusted with a big or a small endowment. Results support intention-based models of cooperative behavior, confirming that people use the opportunity to punish and reward. The Machiavellian personality scale designed to detect non-cooperators was administered but previous findings were not replicated.

1. Introduction

Cooperative behavior is not only beneficial for individuals, it could be also necessary for survival. These are some of the reasons it has been widely studied in the social sciences. Most importantly, cooperation is based on trust and reciprocation. Trust in the context of this paper involves engaging in a social exchange, which is characterized by uncertainty and vulnerability. To understand what the concept of trust is, we need to differentiate between trust and its related concept of trustworthiness. It is the lack of trustworthiness in interaction partners that creates the problem of trust and it can be solved by making people more trustworthy. If something conceptually entails or causes trustworthiness, then indirectly it tends to cause trust (Hardin, 2002). Reciprocation is the response to trust and is linked to the concepts of generosity, fairness and punishment. We try to make the most of the available information about our interaction partners to classify them as trustworthy or not and to decide how to act.

The research area which tries to mathematically capture behavior in such strategic situations, in which an individual's success in making choices depends on the choices of others, is game theory (e.g., Camerer, 2003). It uses various experimental games to study human behavior¹. Game theoretical research on what type of people we trust has focused mainly on reputation formation after repetitive interaction (Brandts & Figueras, 2003), leaving one-time interactions less explored. Technology growth together with the increasing “strength of weak ties” (Granovetter, 1983), however, make one-time interactions more likely and information gathering easier, so more research is needed to shed light on the important processes and factors in such situations. Most of the game theoretical experiments assume that we know nothing about the counterpart but often in real life this is not the case. We inquire about our future partner and the information we gather is not always correct and can be sometimes contradicting. Such information asymmetries are mostly taken for granted in the research literature (Guennif & Revest, 2005, as cited in Boero, Bravo, Castellani, Lagana & Squazzoni, 2009) despite the fact that sociological research has shown the importance of the role third parties play in trust processes, especially in large and complex networks, characterized by uncertainty (Corten & Cook, 2008, as cited in Boero et al., 2009). An important area where trust, reciprocation and information gathering intersect is democracy: a citizen gives his or her vote and hopes that the representative keeps the promise. There is a

¹For an example, see 2.3. The trust game.

great potential for both cheating and mutual gain and a lot of uncertainty is involved. Other real life examples are shopping on websites like Ebay (www.ebay.com), borrowing money to a new colleague, hiring new personnel, subletting an apartment, etc. What all of these situations have in common is that one side has to make the first move and trust. It is only logical that having back-up information on the people we interact with will help. For example, an employer would take a good look at the candidate's CV and call on the references to check how the candidate was doing on her previous job. While there are situations where future interaction is expected, there are also such where it is highly unlikely (e.g. shopping on the Internet, getting a taxi in a foreign country). This theoretically increases the chances for cheating and ungrateful behavior and this is what makes one-time interactions so interesting.

1.1 Personality - behavior distinction

This project focuses on two types of information about potential interaction partners – personality-related and behavioral. As these two may be strongly interrelated, it is important to highlight the aspects that make them two separate characteristics in the context of this project. Personality information pertains to the beliefs we have or say we have about the world and how one should behave with others, i.e. a predisposition, an attitude. It is a more global, indirect piece of information based on self-report. Past behavior is simply a direct record which reveals choices made in the past in the same context only and is thus more local. One might expect that the perceived truthfulness and believability of both types of information will be weighted differently in different situations. In a perfect world, people would be consistent, i.e. their benevolent actions would make their personalities good or their good personalities would make them act benevolently. In the real world, however, this is not always the case. We do not always act according to our beliefs, or the beliefs we present to hold.

Research has not yet fully revealed whom we trust and whose trust we reciprocate in one-time interactions with none or very little probability of meeting again. Is it that we trust people, whose characters are generous and fair? Or is it the previous behavior and not the global personal characteristics that matters? How do we cope with contradicting information in such cases? This experiment is a first attempt to answer these questions by putting

common real-life situations in the frame of a simple trust game and manipulating the qualities of the players to see how they influence the behavior of the real participants.

In the next section, I give an overview of relevant models and experimental findings supporting the view that people have other-regarding preferences and care about the motives of others when making cooperative choices, demonstrating that people use available information about the characteristics of others. Then I introduce the trust game used in this experiment and related results showing that it is a proper tool for the purpose of the current project. In the following paragraphs I present evidence of uncovered strategies people use in cooperative exchange, relevant for the hypotheses. Lastly, I focus on the two types of information to be manipulated - personality and behavior - by referring to previous research done in a similar setting and introducing the tools for measurement.

2. Theoretical background and previous findings

2.1. Types of models of cooperative behavior

Cooperative behavior in psychology and economics is mainly studied using experimental bargaining games. Various models have been created trying to explain individual economic behavior. Two approaches to cooperation have dominated the research field in the past years - outcome based and intention based (McCabe, Rigdon, & Smith, 2003). According to the outcome based models, the intentions players attribute to each other are of no importance and it is the properties of the outcome that matter (e.g. that one earns more money from the interaction). Intention based models, on the other hand, emphasize the role of reading the intentions of the other person.

There has been a lot of evidence inconsistent with outcome based models, (e.g., McCabe et al., 2003; Nelson, 2002). A large body of evidence in experimental economics demonstrates that one of the main assumptions of economists, that people act mainly according to self-interest, may be wrong (Fehr & Schmidt, 2005). Furthermore, there are numerous experiments documenting how people show altruistic behavior without having any personal economic gain (for a review see Nelson, 2002, or Fehr & Schmidt, 2005). Such findings disproved the initial belief that people are not concerned for the well-being of others and that they are not motivated by other-regarding preferences. Scientists started to look into under which conditions these other-regarding preferences have a role and how they can be properly described and modeled. It turned out that models taking self-interest as the main

behavior motivator do much worse at predicting behavior than models taking other-regarding preferences into account (Fehr & Schmidt, 2005). Of course, it is not the case that all people are either selfish or other-regarding. Rather it is important to know under which conditions the well-being of others matters to us. More precisely, experimental evidence shows that intentions play a role in rewarding and punishing behavior. De Quervain et al. (2004) found that people derive satisfaction from punishing trust abusers, even when this decreases their economic payoff. Using positron emission tomography (PET), the authors detected activation in the dorsal striatum, a part of the brain associated with processing of awards after a goal-directed action, thus demonstrating the neural basis of altruistic punishment of defectors. In their review of experimental evidence and new theories, Fehr and Schmidt (2005) conclude that participants want to punish unfair intentions and unfair people. They also note that the findings regarding rewarding benevolent behavior are mixed and the motivation to reward is of less (if any) importance.

Clearly, information that could tell us something about people's intentions will be relevant in a context where cooperation can take place. For example, it has been established that prior social or moral information about trading partners exerts influence on the decision to cooperate (Delgado, Frank, & Phelps, 2005). Participants read vivid descriptions of life events indicating praiseworthy, suspect or neutral moral character of their game counterparts and had to decide whether to trust them. Participants tended to make more risky choices with partners of good moral character, which was supported by a reduced activation in the caudate nucleus, a part of the brain linked to learning feedback. Apparently, when the participants encountered a good or a bad person, as opposed to a neutral one, they discounted feedback trial-and-error learning information, relying on the characteristics provided by others.

2.2. Information sources

The effect of information about the other player has been mainly studied in repeated games, which allow for reputation formation (e.g., Brandts & Figueras, 2003). It has been found to significantly increase cooperation in certain situations (e.g. Novak & Sigmund, 1998, as cited in Barrera & Buskens, 2009). In one-shot games, however, reputation formation is not possible, so another option is using third party reference. For example, Barrera and Buskens (2009) demonstrate that knowledge of someone's reputation does not need to be based on direct experience with the counterpart in order to be useful. One of the

examples of how people make use of secondary sources mentioned by them is gossip. They also report that the effects of third party information can have a positive effect on cooperation, as long as the third party is perceived as responsible for the information they transmit.

2.2 The trust game

One of the best tools to study cooperative behavior is the so-called trust game. It was initially developed by Berg, Dickhaut and McCabe (1995). There are two players. Player #1 can send between zero and A to Player #2. The experimenter then triples the amount sent and Player #2 receives $3xA$. Player #2 is then in a dictator position and is free to return anything between zero and $3xA$ to Player #1 (see Fig. 1).

Figure 1. A simple sketch of the trust game.

Usually, both parties have equal starting budget (e.g., Berg et al., 1995, Gunnthorsdottir, McCabe, & Smith, 2002). The typical finding is that Players #1 entrust approximately half of what they have and Players #2 send back at least what they have been given or slightly more (Fehr & Schmidt, 2005). In the experiments by Pillutla, Malhotra and Murnighan (2003), participants returned more as more money was sent, which was interpreted as a function of felt obligation; they reciprocated less when they were sent less. In the experiment by Berg et al. (1995) thirty out of thirty-two players trusted their counterparts and a substantial proportion of players reciprocated so that both participants made a profit. Such findings seem to be robust and not due to framing. Ortmann, Fitzgerald and Boeing

(2000) used various treatments to prompt strategic reasoning but none of them led to reduction of the investment.

It should be noted however, that usually experiments have each person play one of the roles only - either a trustor or a reciprocator. Burks, Carpenter and Verhoogen (2003) had participants play both roles which resulted in a significant decrease in trust and reciprocation in comparison to other conditions and experiments. These findings go in line with the so-called “reduced responsibility” hypothesis proposed by the authors, according to which this decrease is due to the participants knowing that their counterparts will have more opportunities to make money in the game². This speaks once more in favor of the other-regarding models.

2.4. Types of strategies in cooperative exchange

Burnham, McCabe, and Smith (2000) propose a very simple but effective heuristic strategy people potentially use. They present evidence that people have a preconscious friend-or-foe (FOF) mental mechanism for evaluating the intentions of another person. This mechanism is viewed as a genetic adaptation, which works on the basis of emotional structures built to achieve mutually advantageous reciprocity. It operates in the following way: If the counterpart is a friend, perform behavior A; If she is a foe, perform behavior B. The authors report that the simple manipulation of changing the term used to refer to the counterpart from “opponent” to “partner” results in a significant difference in trusting behavior. “Partners” activate the FOF mechanism and are evaluated as “friends”, leading to more trust on the side of the participant.

Psychological literature is abundant with examples of the presence of a reciprocation norm in society. It is a simple rule which says that we should try to repay what another person has provided us (Cialdini, 2001). Bicchieri, Duffy and Tolle (2004) model how such a behavioral pattern, or a social norm, of trusting and reciprocating develops among boundedly rational agents in a complex environment in the absence of upper control. Using the trust game (Berg et al., 1995), they demonstrate that a social norm of trust and reciprocation is unlikely to be associated with a particular strategy. More precisely, it is likely to be a result of

² The authors base this hypothesis on the subjects’ comments on two open-ended questions, that they feel less obligation to their counterparts in the two-role protocols because they supposedly know that their counterpart will play with other people, too. This hypothesis is, however, not tested directly.

several conditional strategies. Conditional here means that under no condition do the players act according to a simple rule to trust or reciprocate no matter what. Rather, conditional means using a rule like “trust only if you are trusted” as opposed to “trust no matter what”. Using conditional strategies like the FOF mechanism is a behavior very likely to be observed in participants in the trust game when they have been provided some information according to which they can evaluate their interaction partners.

2.5. The role of personality in bargaining games: The Mach scale

Christie and Geis were the first to develop a nowadays widely used scale designed to measure a person's tendency toward Machiavellian behavior, i.e. “in particular cynicism, a penchant for manipulateness, and the belief that the end sanctifies the means” (Christie & Geis, 1970, as cited in Gunnthorsdottir et al., 2002, p. 54), named after the 15/16th century Italian politician and author, Niccolo Machiavelli. The scale consists of 20 statements, e.g., „it is safest to assume that all people have a vicious streak and it will come out when they are given a chance”, to which people are asked to respond on a Likert-scale. Machiavellian behavior is categorized by some as a type of intelligence, which allows us to manipulate others in order to attain particular goods and has been observed not only in humans but also in primates (Knecht, 2004). The literature refers to people receiving a high score on the scale, exhibiting strong preference for Machiavellian behavior, as “high Machs” and to people on the opposite end as “low Machs”. High Machs are characterized by a cool and detached attitude towards social norms (Mudrack & Mason, 1995, as cited in Gunnthorsdottir et al., 2002). The Mach scale has been mainly used for identification of non-reciprocators. It has been found to correlate highly with economic opportunism (Sakalaki, Richardson, & Thépaut, 2007). Gunnthorsdottir et al. (2002) find that although Machiavellianism does not predict trust, it predicts reciprocity quite well: high scorers tend to defect when this is of their advantage, compared to those who score lower and tend to reciprocate. Burks et al. (2003) present some contradicting results: in their experiment the Mach scale was a strong predictor of trusting behavior with high Machs trusting the least but failed to predict reciprocation. Overall, the predictive properties of the scale show mixed results.

What has been investigated so far is how the player's score on the Mach scale is related to his or her actions, or how other relevant characteristics like appearances (Snijders & Keren, 2001) or the faces of the people we interact with (van 't Wout & Sanfey, 2008)

influence the decision whether to trust. To my knowledge, there has not been any research done yet on how information about the personality of the counterpart (e.g., their standing on the Mach dimension) is related to cooperative behavior.

2.5 Previous behavior

In an experiment by Bracht and Feltovich (2009), the investors in a trust game received information either directly from the other player or from a source with high credibility - the experimenter. The direct information was in the form of a statement, or “cheap talk” stating the intentions of the counterpart. The indirect one was an observational record of his past behavior in the same game. They found that allowing observation results in substantially higher cooperation and efficiency, while cheap talk had little effect. In this case, participants did not use the player’s promise to decide but rather an objective indirect measure from the same context.

Another line of evidence about the effect of behavior information is the common finding that we tend to interpret it as a strong sign of character more than of situational circumstances, even when there are strong situational incentives to explain the behavior. We tend to commit the so-called fundamental attribution error (Ross 1977, as cited in Snijders & Keren, 2001). It has to do with our expectations that actors will behave consistently with their beliefs, which is sometimes not the case in real life.³ The fundamental attribution error, also called correspondence-bias, has been mainly studied in the field of attitude formation. In one of the most popular and often replicated type of experiments, participants are presented with an essay written by someone else, either in support of or against a certain policy or person (Jones & Harris, 1967). Then the participants are informed that the author either freely chose which side to take, or was forced to write pro or against, irregardless of their true attitude. The striking finding is that although fully aware that the author of the essay had no choice, participants exhibited a tendency to attribute correspondence between behavior and personal attitude. Thus, information about the other player’s previous behavior will definitely be relevant when making a choice and it would be interesting to observe how personality and past behavior, as defined in this experiment, interact.

³ Note that the fact that character and behavior are perceived as strongly interrelated supports the expectation that information about them will matter to participants when making a choice. Yet, the experiment clearly differentiates between the characteristics of the two types of information and allows for the effects of their different properties to be investigated.

3. Hypotheses

In the following experiment the participants filled out the Mach IV scale and answered demographic questions. Then they played the trust game with many different partners, whose characteristics were manipulated in terms of their personality (Machiavellian tendencies) and past behavior (payback rate of reciprocation in previous rounds of the game). Based on the reported findings, one could expect that the provided information will matter for the participants' choice how much to trust and reciprocate. Below it is hypothesized exactly in what way trust and reciprocation are expected to relate to the different characteristics of the players. For the sake of clarity, from now on the real persons who came to the lab would be referred to as "participants" and the computer-generated people as "players".

Consistent information

When the information about the player is consistent, e.g. high Mach score and low payback, participants will activate the FOF mental mechanism and will trust and reciprocate more to those players, whose previous actions and personality make them appear more trustworthy compared to those whose actions and personality present them as untrustworthy. In other words, the conditional rule would be: If a player is consistent, act according to what he or she deserves.

Inconsistent information

Trust

Given inconsistent information, (e.g. High Mach score and high payback) participants will rely more on the information about the player's previous behavior. This information is more "up-to-date", showing the latest behavior in the context of the game, as opposed to knowledge about the personality, which is more global. Conditional rule: If the player behaved badly, do not trust. If the player behaved in a good way, trust.

Reciprocation

Participants will use both types of information into account. Because they are in a dictator position and there is no danger of loss, they might use the opportunity to "punish" undesirable behavior or reward benevolent one, expressed through payback or personality. Conditional rule: If the player is a high Mach or did not behave in a good way in the past, do

not reciprocate fairly. If the player is a low Mach or behaved generously in the past, reciprocate fairly.

Mach hypothesis

High Machs will tend to trust and reciprocate significantly less than low Machs.

4. Method

4.1. Participants and Procedure

The participants were 71 students, 26 males, mean age of 20 ($SD=1,8$), from Jacobs University Bremen, who took part in the experiment for course credits or chocolate. The experiment was conducted in English, a language all students were fluent in. On their arrival in the lab they read and signed a consent form. Afterwards they filled in a demographics questionnaire⁴. The participants then filled out a paper version of the Mach IV scale with 5-point Likert-scale answer options. Then they met once more with the experimenter, who explained in detail what the scale measured, what its scores meant and how they should be interpreted⁵. The participants were then informed in detail about the rules of the trust game, which started with some practice trials. In order to make sure that the meaning of scores for both types of information and the rules of the game were clear, short sketches with explanations were provided for each participant and could be used at all times. It was emphasized that every game will be played with a different person and that this person was described by the characteristics shown. The participants were told that they would not know how much the players returned to them. They were informed about the opportunity to win a second credit if they paid attention and played carefully. This prize was introduced in order to stimulate sticking to the rules and increase the incentive to pay attention, because the game was long and required some thinking. The receiving of additional prize was actually

⁴ Demographics included age, gender and nationality. Additionally, there were questions about religion and religious affiliation and one item from the World Values Survey (<http://www.worldvaluessurvey.org/>), designed to measure general trust: "Generally speaking, would you say that most people can be trusted or that you need to be very careful in dealing with people?". Participants had to choose between two alternatives: a) Most people can be trusted. b) Need to be very careful". None of those items showed significant differences, so no further results are reported.

⁵ The participants were not informed about their own score on the Mach scale.

determined by a random factor (the points given out on the last trial in the trust condition). The first half of the participants was given the opportunity, if they wish, to write down their comments on a sheet of paper in the form of a free text. The participants were given their prize and left their e-mail addresses for debriefing.

4.2. Experimental design

The experiment consisted of two parts – the trust move (where the participant is Player #1) and the reciprocation move (where the participant is Player #2). Each participant played both roles. A concise description of the experimental design can be found in Table 1.

Table 1.

Experiment Design.

		Independent Variables	Dependent Variables
Between-subjects		Role order	
		Characteristics of the counterpart	
Within-Subjects	Trust	Personality Mach score (high, low)	Points entrusted (0-10)
		Behavior (high, low)	Expected payback (0-30)
	Reciprocation	Personality Mach score (high, low)	
		Behavior (high, low)	Points reciprocated (0-30)
		Points sent (high, low)	

There were 4 conditions for the trust move and 8 for the reciprocation move, manipulated within subjects (for a detailed description see Table 2). The experiment was programmed in Superlab 4.0 and the players appeared in a random order for each participant. There was one between-subject factor - the order of the trust and reciprocation moves. Some participants were first trustors, while others took the role of reciprocators first. Each participant played 200 games on total, each with a different person. A hundred games were in the trust move (Player #1), where for each condition there were 25 games, and another 100 in the reciprocation move (Player #2), with approximately 12 games in each condition.

The characteristics of the 200 people the participants played against were computer-generated (see Appendix 1). The game was played with points instead of real money and in contrast to the usual scenario Player #2 did not have a starting budget⁶. The information about the counterpart was presented as short sentences showing scores, the meaning of which was explained to the participant (for details and examples see Appendices 1 and 2).

Table 2.

Experimental Conditions.

	Condition's name	Characteristics of the counterpart		
		Mach	Payback	Points received
Trust	CT (consistent trustworthy)	low	high	-
	CU (consistent untrustworthy)	high	low	-
	IH (inconsistent high)	high	high	-
	IL (inconsistent low)	low	low	-
Reciprocation	CTH (consistent trustworthy- high)	low	high	many
	CTL (consistent trustworthy- low)	low	high	few
	CUH (consistent untrustworthy- high)	high	low	many
	CUL (consistent untrustworthy- low)	high	low	few
	IHH (inconsistent high- high)	high	high	many
	IHL (inconsistent high- low)	high	high	few
	ILH (inconsistent low- high)	low	low	many
	ILL (inconsistent low- low)	low	low	few

5. Results

Five participants were excluded from the analyses because they misunderstood the game. They entrusted or thought they received more points than they actually could on most

⁶ When designing the experiment at the time, not the original paper where the game was first introduced was used. Some of the papers using the game do not mention anything about the starting budget of Player #2, so this detail was overlooked. However, this mistake could have had a positive influence by reducing the complexity of the game (the participants were burdened with enough calculations). However, this small change could have produced any other unknown or unwanted influences.

occasions, which indicated a major misunderstanding of at least one part of the game. Such cases are not commonly reported in other studies but here additional factors for consideration were included, which could have been confusing for the less concentrated participants. Ortmann et al. (2000) report a similar possible misunderstanding in an experiment using the trust game. They, too, included additional information that needed to be taken into account. Single trials for some of the remaining participants which were also “impossible” in the rules of the game were coded as missing data, possibly due to involuntary double-click or low concentration. There were two outliers, who were excluded for extreme trust and cooperation. The final sample for analyses was $N=64$. Means for the dependent variables trust and reciprocation for the different conditions were calculated and the data was analyzed with SPSS 17.

Although response times were taken, no analysis was performed. The participants took a very variable amount of time to get used to the game and depending on whether they started off ready with a simple strategy or they thought extensively about each trial, overall response times varied greatly between subjects.

The results section is structured in the following way: First, the trust move is analyzed, reporting replication of previous findings and testing the hypotheses. Second, the results related to the hypotheses from the reciprocation part of the game are reported. Last, findings from the Mach score of the participants and their comments on the game are discussed.

5.1. Trust

Overall, trust was low ($M=3.93$, $N=64$, $SD=1.2$). Mean trust in experiments using this game typically is higher, around 50% of the player's possession. Looking at the raw data for all valid trials, before calculating the means, showed that entrusted amount and expected payback correlated highly ($r=0.696$, $p<0.0005$, $N=5857$). This shows consistency from the participants' side - the more they sent, the more they expected to gain from the interaction. On Figure 3 the mean gain for the computer-generated Players #2 is depicted as a function of the desired gain of Players #1. This is how the final pay-offs would look like if it happened what Players #1 (the participants) wanted. We can see that Players #1 would keep the most points for themselves and only big investments lead to a high payoff for both parties. Equalizing outcomes were preceded by the highest levels of trust; a similar finding was

reported previously by Pilutla et al. (2003) Those who trusted the most were also likely to leave many points for their counterparts.

Figure 3. The outcome of the trust move as desired by the participants. $N = 64$. Mean desired gain of Players #1 = (10 points + expected payback). Mean gain for Players #2 = ((mean entrusted points $\times$ 3) – expected payback).

A two- factor (player's Mach score and player's previous behavior) repeated measures ANOVA was conducted to investigate the relationship between the four conditions and the participants' trusting behavior, which was the dependent variable. There was a main effect for Mach score ($F(1,63)=194$; $p<0,0005$, $\eta^2 = 0,76$), showing that participants trusted players with low Mach scores more as opposed to high Mach players. There was also a main effect for payback ($F(1,63)=277$; $p<0,0005$, $\eta^2 = 0,81$), indicating that participants trusted players with higher previous payback scores more, as opposed to players with lower payback scores (see Fig. 4). Most importantly, there was also a significant interaction ($F(1,63)=215$; $p<0,0005$, $\eta^2 = 0,77$). When information was consistent, participants trusted players who

appeared trustworthy on both factors contrary to those who were consistently untrustworthy. Thus, the consistent information hypothesis was confirmed for trusting behavior. We can see that both personality and previous behavior had an impact on the decision. Yet, behavior explained more variance and gave slightly more weight to the decision when information was inconsistent. This is best seen on Figure 5 depicting the interaction. It is visible that Mach score of the player had a strong influence on trust but only when the past behavior score was high. When players did not reciprocate fairly in the past, their Mach score played only a small role in the participants' decisions, which confirms the inconsistent information hypothesis.

Figure 4. Mean trust in the different conditions.

Figure 5. Mean trust as a function of Mach score and previous payoff.

5.2. Reciprocation

A three-factor (player's Mach score, player's previous behavior and number of points entrusted) repeated measures ANOVA with a between-subjects factor (role order) was conducted. An order effect was found $F(1,62)=13$; $p=0,001$, $\eta_p^2=0,17$), meaning that playing as Player #1 first was related to a significantly higher number of reciprocated points (Fig.6).

Figure 6. Mean reciprocation in the different conditions as a function of order.

Furthermore, there was a main effect of the number of points received ($F(1,62)=295$; $p<0,0005$, $\eta_p^2 =0,83$), player's Mach score ($F(1,62)=74$; $p<0,0005$, $\eta_p^2 =0,54$) and past behavior ($F(1,62)=43$; $p<0,0005$, $\eta_p^2 =0,41$). Logically, the number of points received was the main determinant of reciprocation, explaining most variance in the dependent variable. Receiving more points was associated with reciprocating a larger amount. Also, the main effects showed that a low Mach score and a high score of previous payback lead to more reciprocation as opposed to a high Mach and a low previous payback score. Consistently trustworthy players were reciprocated the most and consistently untrustworthy players were reciprocated the least, confirming the consistent information hypothesis for the second part of the game. Interestingly, both personality and previous behavior determined the participants' decisions, showing a different pattern depending on the points received. When the participants had received many points (Fig. 7), the previous behavior was the main determinant of the decision how much to give back. When the previous behavior score was low, the Mach score had only a small influence on the points returned. When it was high, the

Mach score played a significant role and players with low Mach scores were awarded more points than players with high Mach scores. The situation was different when the participants received just a few points (Fig. 8). When the player's Mach score was high, past behavior determined the reciprocated amount, and those who had been good in the past received many points back. When the Mach score was low, however, past behavior did not matter, and players were rewarded a relatively large amount of points. In this situation, the Mach score put more weight on the decision. The inconsistent information hypothesis for reciprocation was confirmed by both personality and past behavior playing a role depending on the situation. The participants used the dictator position they were in to punish undesirable behavior and personality and reward the desirable ones.

Figure 7. Mean reciprocation as a function of the Mach score and previous payback when many points were received from the player.

Figure 8.. Mean reciprocation as a function of the Mach score and previous payback when few points were received from the player.

5.3. Machiavellianism of the participants

The mean Mach score was 57,2 (SD=8,4, min. 41, max.74, N=66)⁷ on a scale from 20 (no Mach) to 100 (absolute Mach)⁸. The Mach score and the mean trust of the participants did not correlate significantly ($r = -0,15$, $p = 0,22$). Nor did the Mach score and the mean reciprocation rate ($r = -0,08$, $p = 0,50$). Categorizing the participants as high and low Machs, as is typically done in such experiments, did not give any significant differences either thus not providing any support for the Mach hypothesis. There was a visible tendency for the extreme cases on the Mach dimension to show corresponding trusting behavior (i.e. low trust for high Machs and high trust for low Machs) but those cases were not enough to give a statistically reliable estimate. A larger sample size, which would provide more cases to the extreme, has the potential to shed light on the relationship between Machiavellianism and trusting behavior.

⁷ N=66 because the two outliers are included in the Mach analysis.

⁸ The original Mach IV scale is administered using a 7-point Likert scale (scores varying from 20 to 140). In this experiment, a 5-point Likert scale was used.

5.4. Participant's Comments⁹

Participants' comments on their strategies matched with the results. As Player #1, most of the participants mention that they trusted most when the player had both a low Mach score and high previous payback. A common pattern was to choose the previous payback score as a determinant of the decision when the information was inconsistent. However, several participants mentioned that it was quite hard to decide. Overall, the decision was based on the hope that they will "get back more at the end". For the role of Player #2, participants' comments focused on splitting fairly the profit with Player #1 (mostly in half), rewarding previous cooperation and punishing selfishness.

6. Conclusions and Discussion

Trust overall

Trust overall was low¹⁰, which was expected because players knew that they would play both roles, leading to reduced responsibility for the counterpart. This is, however, not the only possible explanation in this case. Participants were aware that they are playing against computer-generated players. A study by McCabe, Houser, Ryan, Smith and Trouard (2001) demonstrated different brain activation patterns when participants were playing the same trust game against a person or a computer. The prefrontal regions were more active when cooperators were playing with a human subject, supporting the hypothesis put forward by the authors, that these regions are responsible for binding contingent information that allows subjects to evaluate the mental states of others. This finding suggests that there is a difference between playing against a computer, who acts according to a pre-generated algorithm, and a real person.

Trust and reciprocity: Consistent information

The high correlation between the entrusted amount and the expected payback has the implications that in the trusting move, participants' actions were not truly benevolent - they

⁹ Less than half of the participants wrote comments after the experiment. The results reported are just common patterns noticed by the experimenter. A more systematic analysis should be conducted (e.g. content analysis) but due to limits of word space this is not done here.

¹⁰ Reciprocity was bounded by the points sent by the computer-generated players, which were either very few or many, so no comparison to other experiments could be made for the reciprocity rate.

obviously expected some kind of gain from their action. When the information about the player one encountered was consistent on both factors, participants made a clear choice to trust those who were trustworthy and to avoid cooperation with those who were untrustworthy. The same pattern emerged for the reciprocation move, where good players who acted fairly in the past were overall the most rewarded. This conditional strategy goes in line with the FOF heuristic, allowing easy categorization of players as friends or foes.

Trust and reciprocation: Inconsistent information

When Players #1 had to decide whether to trust in the face of inconsistent information, they relied more on the behavioral data. It is not yet clear why they did so but possible reasons are that participants thought that Machiavellian tendencies were not so indicative of the player's strategy in the context of the trust game. It could be very well that a Machiavellian decides to be generous in order to make more profit. In this case, intentions of the players were not very clear and what mattered most to the participants was the likelihood to profit more from the exchange, which they estimated from the previous behavior score, indicative of the behavior of the players in previous trial of the same game. Yet, more research is needed to validate this claim.

Reciprocation in more detail

Results from the reciprocation part of the experiment showed an unexpected order effect. Playing the reciprocator first was associated with less reciprocation, or in other words, being the trustor first was related to higher reciprocation rates. It could be that a perspective-taking mechanism is the reason behind this finding. Having already been Player #1 means having experienced the tough decision to invest relying on someone else's fairness. This could have lead participants to be very generous to players who trusted them with many points. Another reason could be that when participants were Player #2, they were always trusted, even with one point only. Participants who played the trustors first had the chance to invest 0 points only, which never happened to them in the reciprocation part. This could have made them feel more thankful for the expressed trust.

Perhaps the most interesting finding linked to other-regarding preferences is the different patterns that emerged when participants reciprocated to players who sent them many vs. only a few points. When participants received many points, their decision how much to

reciprocate was mostly determined by the Mach score. If the player was a high Mach, he was punished with low reciprocation regardless of the previous behavior. When the Mach score was low, however, participants rewarded high previous payback players with more points. In the situation where just a few points were given to the participant, a low Mach score was rewarded regardless of the previous payback. When Mach score was high, previous payback was taken into account. These patterns fit into the predictions of intention-based models- participants used the opportunity to punish and reward. What is most interesting is that what they punished was not the behavior in the context of the game, expressed by the previous payback score, but rather the ungrateful personality and selfish and manipulative tendencies, expressed by the Mach score.

Machiavellianism

Surprisingly, the Machiavellian attitude of the participants, measured by the Mach IV scale, did not replicate previous findings. Neither significant differences in trust and reciprocation between high and low Machs emerged, nor the commonly found gender difference. Most likely, this is due to the small sample size, which did not allow for many extreme cases to be included. Another possible research question tackles the very curious interaction between the Mach score of the participants themselves and the Mach score of the game counterparts. This has potential to give very interesting results, especially having in mind recent models based on perceived similarity, which see it as a basis for cooperation (Fischer, 2009). The interaction between Machiavellians who are aware of each other is a fruitful ground for further research.

Another possibility for the lack of results is the Mach IV scale itself. Recently it has been severely criticized on the basis of inconsistent reliability, tentative factor structure and poor (double-barreled) items (Dahling, Whitaker, & Levy, 2009). A new scale, the so-called Machiavellian Personality Scale (MPS) was created as a substitute to measure Machiavellianism. Taking into account the deserved criticism, a replication of the experiment using the MPS may reveal existing differences and interactions, which the Mach IV scale was not able to capture.

General Discussion

This experiment was a first step to understanding how we combine personality (general) and behavior (local) information to decide on our actions in an economic exchange. A next step would be to try to understand the processes and mechanisms behind the described results by improving the design and the sample. More subtle differences could be investigated by creating characters whose properties (Mach and behavior score) come from normal and other distributions, which will allow for analysis of continuous data, as opposed to creating conditions from the extremes. In this experiment participants were aware that they wouldn't meet any of the players for a second time. In real life, however, the probability of meeting someone again varies according to the situations. This probability could be varied to see how the participants' own actions and the interpretation of their counterparts' actions change as a function of how likely they are to meet again. In this experiment, no feedback was given to the participants about their performance. Another possibility for future inquiries is to investigate the effects of feedback on performance, also by changing the parameterization of the environment (e.g. the percentage of Machiavellians or the pay-off structure). This is highly recommended by Hertwig and Ortmann (2001), who emphasize the importance of repeated playing (meaning playing more than once irregardless of whether with the same person or with different ones). In their view, repetition allows participants to study the environment and perform better even when no feedback is provided, which was the case in this experiment. It is important to note that participants had time to adjust to the game and adapt to interpreting the information. Another tool which psychology should borrow from economics, according to Hertwig and Ortmann (2001), is real monetary incentives, because they lead to convergence of the data toward the performance criterion and reduction of the data's variance. This experiment, however, may have benefited from the lack of true monetary incentive. One purpose of the experiment was to explore any potential differences in behavior between high and low Machs and the interaction between these types. The Machiavellian scale was administered before the game was played, which could have primed the participants to take into account their worldviews and beliefs rather than the exact final pay-off. While there is no data to support this claim, introducing a monetary stimulus would have removed this possible effect in the student population. Also, it would have required defining external criteria that make up the best decision. Because we have a multiattribute decision, this would mean determining the importance of the different attributes (Betsch &

Habestroh, 2001), which is what this experiment was investigating. Of course, one could argue that this lack of a clear differentiation between good and bad choices did not give a chance to the high Machs to stand out. Obviously, monetary prizes are a small but very important manipulation in behavioral experiments, combining psychology and economics, and one should be careful not to arrive at the wrong interpretation.

Knowing more about the way we combine and interpret information about our interaction partners in one-shot game can help us understand and model human behavior in relevant situations. One-time interactions are particularly risky for certain individuals and being aware of individual differences can help reduce uncertainty and vulnerability.

7. References

- Barrera, D., & Buskens, V. (2009). Third-party effects in an embedded investment game. In V. Buskens, C. Cheshire, K. Cook, & C. Snijders. (Eds.) *Trust and Reputation*. New York: Russell Sage Foundation.
- Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, 10(1), 122-142. doi:DOI: 10.1006/game.1995.1027
- Bicchieri, C., Duffy, J., & Tolle, G. (2004). Trust among strangers. *Philosophy of Science*, 71(3), 286-319.
- Boero, R., Bravo, G., Castellani, M., Lagana, F., & Squazzoni, F. (2009). Third party reputation in repeated trust games. Retrieved February, 24, 2010 from http://74.125.155.132/scholar?q=cache:T1O-MPZz7aEJ:scholar.google.com/+reference+Boero+Bravo+Third+Party+Reputation+in+Repeated+Trust+Games&hl=en&as_sdt=2000
- Bracht, J., & Feltovich, N. (2009). Whatever you say, your reputation precedes you: Observation and cheap talk in the trust game. *Journal of Public Economics*, 93, 1036-1044.
- Brandts, J., & Figueras, N. (2003). An exploration of reputation formation in experimental games. *Journal of Economic Behavior & Organization*, 50(1), 89-115. doi:DOI: 10.1016/S0167-2681(02)00042-2
- Burks, S. V., Carpenter, J. P., & Verhoogen, E. (2003). Playing both roles in the trust game. *Journal of Economic Behavior & Organization*, 51(2), 195-216.
- Burnham, T., McCabe, K., & Smith, V. L. (2000). Friend-or-foe intentionality priming in an extensive form trust game. *Journal of Economic Behavior & Organization*, 43(1), 57-73. doi:DOI: 10.1016/S0167-2681(00)00108-6
- Camerer, C. F. (2003). *Behavioral game theory*. Princeton University Press.
- Cialdini, R. B. (2001). *Influence: Science and practice* (4th ed.). Boston: Allyn and Bacon.

- Dahling, J. J., Whitaker, B. G., & Levy, P. E. (2009). The development and validation of a new machiavellianism scale. *Journal of Management*, 35(2), 219-257.
doi:10.1177/0149206308318618
- de Quervain, D. J., Fischbacher, U., Treyer, V., Schellhammer, M., Schnyder, U., Buck, A., & Fehr, E. (2004). The neural basis of altruistic punishment. *Science*, 305(5688), 1254-1258. doi:10.1126/science.1100735
- Delgado, M. R., Frank, R. H., & Phelps, E. A. (2005). Perceptions of moral character modulate the neural systems of reward during the trust game. *Nature Neuroscience*, 8(11), 1611-1618. doi:10.1038/nn1575
- Fehr, E. & Schmidt, K. M. (2005). The economics of fairness, reciprocity and altruism – experimental evidence and new theories. In S.C. Kolm, & J.M. Ythier. (Eds). *Handbook of the economics of giving, altruism and reciprocity*. North-Holland.
- Fischer, I. (2009). Friend or foe: Subjective expected relative similarity as a determinant of cooperation. *Journal of Experimental Psychology: General*, 138(3), 341-350.
doi:10.1037/a0016073
- Granovetter, M. (1983). The strength of weak ties: A network theory revisited. *Sociological Theory*, 1, 201-233.
- Gunnthorsdottir, A., McCabe, K., & Smith, V. (2002). Using the machiavellianism instrument to predict trustworthiness in a bargaining game. *Journal of Economic Psychology*, 23(1), 49-66. doi:DOI: 10.1016/S0167-4870(01)00067-8
- Betsch, T., & Haberstroh, S. (2001). Financial incentives do not pave the road to good experimentation. Open Peer Commentary. *Behavioral and Brain Sciences*, 24, 404.
- Hardin, R. (2002). *Trust and Trustworthiness*. Russell Sage Foundation.
- Hertwig, R., & Ortmann, A. (2001). Experimental practices in economics: A methodological challenge for psychologists? *Behavioral and Brain Sciences*, 24, 383-403.

- Jones, E. E., & Harris, V. A. (1967). The attribution of attitudes. *Journal of Experimental Social Psychology*, 3(1), 1-24. doi:DOI: 10.1016/0022-1031(67)90034-0
- Knecht, T. (2004). Was ist machiavellische intelligenz? Betrachtungen über eine wenig beachtete Seite unserer Psyche. *Nervenarzt*, 75(1), 1-5. doi:DOI: 10.1007/s00115-003-1543-0
- McCabe, K. A., Rigdon, M. L., & Smith, V. L. (2003). Positive reciprocity and intentions in trust games. *Journal of Economic Behavior & Organization*, 52(2), 267-275. doi:DOI: 10.1016/S0167-2681(03)00003-9
- McCabe, K., Houser, D., Ryan, L., Smith, V., & Trouard, T. (2001). A functional imaging study of cooperation in two-person reciprocal exchange. *Proceedings of the National Academy of Sciences of the United States of America*, 98(20), 11832-11835.
- Nelson, W. R. (2002). Equity or intention: It is the thought that counts. *Journal of Economic Behavior & Organization*, 48(4), 423-430. doi:DOI: 10.1016/S0167-2681(01)00245-1
- Ortmann, A., Fitzgerald, J., & Boeing, C. (2000). Trust, reciprocity, and social history: A re-examination. *Experimental Economics*, 3, 81-100.
- Pillutla, M. M., Malhotra, D., & Murnighan, K. J. (2003). Attributions of trust and the calculus of reciprocity. *Journal of Experimental Social Psychology*, 39(5), 448-455. doi:DOI: 10.1016/S0022-1031(03)00015-5
- Sakalaki, M., Richardson, C., & Thépaut, Y. (2007). Machiavellianism and economic opportunism. *Journal of Applied Social Psychology*, 37(6), 1181-1190. doi:10.1111/j.1559-1816.2007.00208.x
- Snijders, C., & Keren, G. (2001). Do you trust? Whom do you trust? When do you trust?. *Advances in Group Processes*, 18, 129-160.
- van 't Wout, M., & Sanfey, A. G. (2008). Friend or foe: The effect of implicit trustworthiness judgments in social decision-making. *Cognition*, 108(3), 796-803. doi:DOI: 10.1016/j.cognition.2008.07.002

8. Appendix

Appendix 1

In order to create the players' characteristics (consistent and inconsistent), four beta distributions, defined on the interval (0, 1) parameterized by two positive shape parameters denoted by α and β were generated in MATLAB 2009 (see Fig.2). This type of distribution was chosen because it allows for different shapes. For the generation of the conditions above, a skewed rather than symmetric distribution was needed to represent a population of people falling towards the extremes (e.g., very unfair or a high Machs vs. very fair and a low Machs). In two of the distributions the frequency of the values approaching 0 was significantly higher than those approaching 1 (R1, R3), and the case was the opposite with the other two (R2, R4). The values were multiplied by 100 and then rounded to result in four lists of numbers in the range 0-100. As the generated numbers were more than actually necessary for the experiment, always the most extreme values from the lists were chosen in the order they were sorted from smallest to largest or vice versa. The values of the four lists were paired accordingly to generate the necessary conditions. In order to determine how many points the players sent in the reciprocation condition, a similar procedure was undertaken, which resulted in either very few points or many. Below are two examples of players:

Consistent player (reciprocation condition):

This player's Mach score is 8. This player's payback in the previous round is 79. How many points would you entrust to that person?

Inconsistent player (trust condition): This player's Mach score is 77. This player's payback in the previous round is 84. How many points would you entrust to that person?

Figure 2. Beta distributions used for the generation of consistent and inconsistent characters. Parameter $\alpha=2$ and $\beta=6$.

Appendix 2

Trust move

When the participant was the trustor, the information was presented in the following way: E.g. “This player's Mach score is 24. This player's payback in the previous round is 19. How many points would you entrust to that person?”. After giving an answer, the following question was asked: “How much payback do you expect?”, designed to measure what the participants expected to gain from every interaction.

Reciprocation move

The presentation was similar to the trust move with the difference that a third parameter – the number of points sent - was added: e.g. “This player's Mach score is 26. This player's payback in the previous round is 9. This player has sent you 10 points. How many of them do you give back?”. It should be noted that participants always received at least 1 point from the computer-generated players, i.e. values of 0 were removed from the list for points sent by Player #1. This was done because such trials were of no interest for the current hypothesis, as they would produce no possibility for a response by the participants. They would also potentially create biases or order effects that could be caused by a complete lack of trust by many players the participant encounters. Also having participants play both roles was expected to reduce cooperation additionally, as demonstrated by Burks et al. (2003). Moreover, the neutral term “player” was used instead of the possible substitutes “partner” or “opponent” in order to avoid any priming effects, which could influence the decision process, as found by Burnham et al. (2000) and interfere with the characteristics of the players.