

Do Antisocial preferences exist?

A review of experiments on antisocial preferences

Tom Straatman

25-07-2012

Abstract

For many decades research has been done on social preferences. Only recently researcher started to look for the occurrence of antisocial preferences. They designed new games to test for the ‘dark side’ of human nature. We tried to find out if antisocial preferences really exist or if they are just a side effect of the design of the games. We discussed fairness, inequality and reciprocity. We also checked for the occurrence of experimenter demand effects. In the last part of our research we wanted to see if and how social- and antisocial preferences would fit together. We found the positive burning rates found in the experiments were mainly because of fairness and inequality but there was also evidence for highly antisocial behavior. We found that social- and antisocial preferences would fit together and designed a scale to do this. Antisocial preferences do exist but they only cover a small part of human behavior.

Introduction:

For many decades, research has been done to investigate preferences. Preferences can show how the attitude of people is toward a set of objects in a decision making process. So preferences induce certain behavior. Researchers are investigating these preferences and then try to fit the behavior found into models. In the beginning researchers assumed that people are rational and developed the so called rational theory (Green 2002). To understand this theory we must look at a few assumptions that have to be made. According to Green the rational choice theory is based on a few assumptions. First, people face a set of choices. They have to choose between different possibilities with different outcomes (just like in real life). Second their preferences are complete, this means that the options given cover the entire range of possibilities (you prefer A over B, B over A or you are indifferent between A and B). The third assumption is that the preferences are transitive (if A over B and B over C than it must be that A is chosen over C). The final assumption is that people choose the most preferred alternative. Rationality has many definitions, but here we will see rationality as self-interested behavior. A person will only look at his or her own payoffs when considering a set of choices.

However outcomes found in experiments suggests that there is evidence against rational behavior. Positive donations in ultimatum games and dictator games (Charness and Rabin 2002) suggest that people are not purely self-interested. An ultimatum game is a game in which there are two players. The first player can make an offer to split an amount of money (the proposer) and the other can choose to accept or reject (the receiver). When the receiver rejects an offer both players receive nothing.

To make this clearer we will give an example. Player 1 receives \$10 and has to divide this between himself and another player (player 2). Player 1 is free to choose the amount he will send to player 2. However player 2 is allowed to reject the offer done by player 1. In this case both receive \$0. If player one decides to send \$3 to player 2 and player 2 accepts this offer, player 1 has a payoff of \$7 and player 2 has a payoff of \$3.

Rational game theory states that the receiver will accept every positive offer (even the lowest), because this increases the utility of the receiver. The proposer knows this and will therefore bid the lowest amount possible and the receiver should accept. However in most research amounts larger than the lowest possible are offered and even get rejected. To see this in the light of the earlier example. The lowest amount player 1 can offer is \$0,01. Research has found significantly higher offers than this which get rejected. The dictator game is a similar game, only here the receiver has no possibility to reject the offer. According to rational theory the dictator will keep all the money to himself giving nothing to the receiver. Again here research has found large positive amounts (larger than the lowest amount possible). These are just two of the enormous amounts of games developed to look for preferences.

Because of this evidence researchers started investigating social preferences (Charness and Rabin 2002). We define social preferences as preferences which deviate from the rational game theory. These social preferences show people do not only care about their own payoffs. An example of these preferences is altruism. The reason for research on social preferences was that researchers wanted to design models that incorporate social preferences. These models might be used to predict behavior in real economic situations. Models might explain how people react on changes in prices or how people behave when subjected to different tax schemes. Charness and Rabin (2002) found that participants are interested in raising social welfare even sacrificing endowment to increase the welfare of participants with a low payoff. Participants also care for reciprocity. They lose their willingness to sacrifice their own payoff to help others if other people are not willing to do the same. They also punish unfair behavior.

As we can see, a lot of research has been done to investigate preferences. In most of this research experimenters only looked at social preferences like altruism, fairness etc. Social preferences are based on the assumption that people do not only care about themselves. The well-being of others is also important for a person's own well-being. Social preferences can be used to explain the positive amounts bid in (among others) the ultimatum game and the dictator game. A lot of interesting models have emerged over the years that show people do not only care about themselves.

Recently authors have found a new branch of interest concerning preferences. These experiments are designed to look at the 'dark side' of human nature. They explore the possibility of antisocial preferences (envy, spite, etc.). Antisocial preferences are preferences where people receive a payoff from the hurting of others. Among the reasons for the interest in these antisocial preferences were the sights of random destruction of property, violence against other people and stealing. Antisocial preferences are based on the possibility that people increase their own well-being by decreasing the well-being of others. Experiments like the joy-of-destruction game, the vendetta game, money burning and the destructor game were designed to look for the occurrence of antisocial preferences. Because antisocial preferences can show that there is a dark side of human nature, we think a lot of people don't want to believe this. One always hopes for the goodness of human nature. We will look at the research done on this subject in order to answer the research question of this paper: Do antisocial preferences really exist?

This will be done by reviewing the literature concerning this subject and analyzing the experiments. In order to find an answer to our research question we will have to look at a few sub questions: First we will explain fairness, inequality and reciprocity. These concepts are important to know, as they might have an influence on the outcomes of experiments. Second we ask how one can investigate antisocial preferences. Here we will review the experiments done by several authors and then deal with the

phenomena of experimenter demand effects and try to gain insight in the role of fairness, inequality and reciprocity. Third we want to know if and how these antisocial preferences fit in with the social preferences. The research done suggest social- and antisocial preferences are each other's opposites, but is this really true? Maybe social- and antisocial preferences can explain human behavior when combined.

We use fairness, inequality and reciprocity in our research because these can be forces which might influence social preferences. Sometimes people punish others in order to create a socially desirable allocation of resources or maybe people punish because they are bad. Also a lot of research has been done on these concepts. As we will see this research found that behavior is strongly influenced by these three concepts.

After these sub questions have been answered we can use these findings to answer the research question and give some suggestions for further research.

Fairness, inequality and reciprocity

Fairness

People care for fairness. If earnings were not gained in a fair way (stealing, randomly assigned etc.) people do not take this lightly. We can see this in real life. Think of investigations concerning prior knowledge in stock and bonds trading, the punishments for stealing, the Dutch tax on winnings in lottery and gambling and even the tax on the inheritance. This last two we might have to explain. Because of the randomness of the gambling and lottery people might see this as unfair. We think that because of this, taxes are put on the earnings gained by lottery participation or by inheritance. People want to lesser the distortion in endowment distribution. Also the intentions matter. Falk et al. (2008) found evidence for the importance of the intensions of fairness on positive and negative reciprocity. They did an experiment and with the results of that experiment to test whether a model could be formed. When the attribution of fairness is ruled out in their model, the responses are substantial weaker (lesser positive- and negative reciprocity). Some percentage of participants even behaved as ignorant individuals. They did not take the behavior and endowment of the other players into account.

Another aspect of fairness is called procedurally-based fairness. Procedurally-based fairness is used when looking at the process of allocation of certain endowments. The process of distribution has is linked to with fairness, people want a fair process of distribution. Why do people care about procedural fairness? De Cremer and Blader (2006) did several studies to look why procedural fairness is so important for people. They designed questionnaires to test for the need to belong. Explanation and further elaboration of this experiment goes beyond the boundaries of this research. Important for us is that De Cremer and Blader (2006) found evidence that the care for procedurally-based fairness

has to do with the desire to belong to a certain group. Because of the need to belong to a group, say an income group. People will more closely watch how the process of distribution takes place than when they do not care about the income group they are in. When this doesn't happen in a fair way, they can reciprocate. This to ensure a place in the desired group or to punish the ones responsible for failure to reach or stay in a certain group. In real life protests may occur or some people even steal to retaliate the unfairness of the distribution

Inequality

Inequality is a definition which is closely related to fairness. As unfair gaining might lead to inequality (stealing leads to a person having more than the other). Inequality is the condition of being unequal. An example of this is the difference between the rich and the poor (their wealth is not the same). Inequality manifest itself in many ways, one of those ways is rank egalitarianism. Rank egalitarianism is the inequality aversion amongst people. In their minds people might have ranked their fellow men. They can use different measures like endowment, luck, appearance etc. People might envy others who are richer or better than themselves. Because of this they want to make sure everybody has the same. The other side is people who are richer might prefer to donate some of their money to subjects who do not have a lot of money, to lower the gap between themselves and the poorer people.

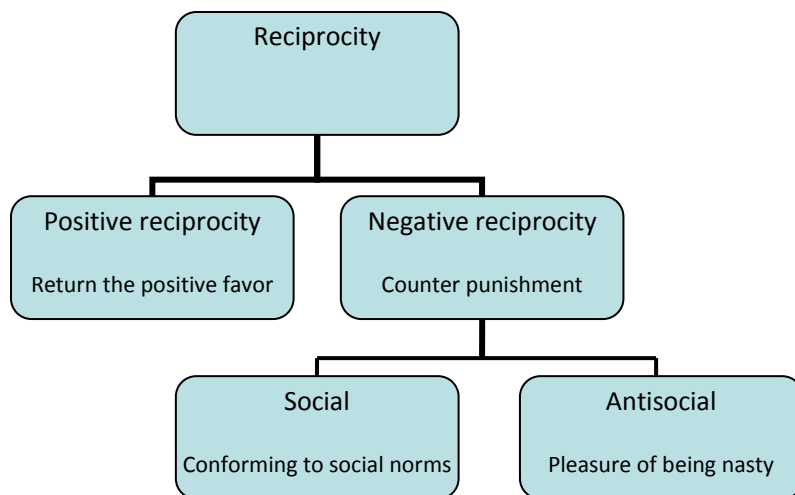
Inequality is also influenced by the so called reference point effects. Reference points are indications for people to check if something is good or bad for them. People always have a certain idea of the value of something and use these values as a benchmark to evaluate new information. We can see this at companies who use current interest rates as a benchmark for future rates. They check whether or not interest rates are better or worse than interest rates in the past. So the past interest is a reference point to determine if the current interest rate is good or not. If you change the references point (by stating an assignment different) someone might change his or her view on things. Another example is this one: you get a raise in your paycheck of 5%. You are all excited because your reference was your previous paycheck. So in this case your reference point was your previous wage and has an influence on your behavior (maybe you will work a little harder). Then you find out that all your colleagues got a raise of 10% and you get angry at your boss because you did not get as much raise as your colleagues. The reference point has changed from your previous paycheck to the pay raise of your colleagues. You see that if we frame the raise in a different way the exact opposite feeling occurred than in the first example. People can use these reference points to check whether or not inequality occurs and adapt their behavior according to the information. So reference effects play a part when looking at inequality.

Reciprocity

The concept of reciprocity is closely linked to fairness and inequality. We can define reciprocity as a response to perceived kindness or unkindness. As this definition suggests, two kinds of reciprocity emerge. On the one hand we have positive reciprocity. This form of reciprocity has to do with returning the favor. If people experience generosity of other people they are inclined to answer this generosity with generosity. The other form is negative reciprocity, which means the exact opposite. If someone does not behave according to social norms, are unkind or is free riding, others will answer this bad behavior with bad behavior or punishment. This last form of reciprocity can be seen in as a social preference as well as an antisocial preference. In the experiments of (Charness and Rabin 2002) we can see the social aspect of reciprocity.

Herrmann et. al (2008) conducted a public goods experiment. In this experiment participants are asked to donate some or all of their endowment to a project. This project generates a certain amount of endowment which is divided between all the participants (even if they did not contribute). Herrmann et. al 2008 found that participants who have not contributed to a project were punished by the other participants. Participants punish other participants who do not behave according to social norms. So we think these punishments are meant to get other participants to get back in line. We can see this form of reciprocity as a social preference because this means people care for the behavior of others. They want the others to conform to social norms.

However there is also another form of reciprocity, an antisocial preference. Participants who do not behave in line with social norms often get punished. Sometimes these participants retaliate against their punishers or just punish others. A reason for this might be they feel like their punishment is not fair or they receive a certain endowment of punishing back. Another explanation could be that participants are bad people and just want to punish others (without any intention to put others back in line with social norms). Maybe some participants derive pleasure from hurting others. It might also be that people are curious to check out how hurting others feel. People want to know if they are capable of conducting bad behavior or want to know how it feels to hurt others. We can also explain antisocial reciprocity when we look at rank egalitarianism. People want to reciprocate the fact there is inequality in the group. The experiment done by Herrmann et. al will be elaborated in the paragraph were we review experiments done on antisocial behavior. Figure 1 on the next page shows how we see reciprocity.



(Fig 1)

Reciprocity consists of positive reciprocity and negative reciprocity. With positive reciprocity we mean that people reciprocate positive behavior with good behavior and with negative reciprocity we mean people reciprocate negative behavior with negative behavior or conducting negative behavior as a counter punishment. So negative reciprocity consists of a social part (getting people back in line with social norms) and an antisocial part (counter punishment or negative reciprocity for a person's own pleasure or curiosity satisfaction).

How can one investigate antisocial preferences?

Just like the games for social preferences researcher have developed games to look for the occurrence of antisocial behavior. Different forms of the joy of destruction game, the destructor game and the money burning game have been introduced (Zizzo and Oswald 2001, Zizzo 2004, Abbink, and Sadrieh, 2008, Abbink, Masclet, and van Veelen, 2009, Abbink, and Herrmann, 2011, Kessler, Ruiz-Martos and Skuse, 2012) also a game called the vendetta game was introduced (Abbink, and Herrmann 2009). All these games are designed to look for the occurrence antisocial preferences. In this section we will review, explain and give comments on these experiments.

Money burning game

The search for antisocial preferences started with a research done by Zizzo and Oswald (2001). They analyzed their experiments done in 1998 at Warwick University and in 1999 at Oxford University (In 2002 they ran a similar experiment from which they also analyzed the results). In this experiment they used the so called one shot Money Burning game. One shot means participants played the game for only one round. In this game participants could (after a practice round) earn money units (doblons) by betting money. The first two (of four) persons could bet more money per round (this is relevant for answering the fairness question later on). After people earned money units, they entered the burning stage. In this burning stage people could first practice the burning and again the first two subjects

receive an extra bonus. The practice round was done to show participants the consequences of their actions. After the practice they decided if and how much of the endowment of the other players they would burn (destroy). This simply means a participant who got burned loses money. This destruction was not always free. In some treatment groups participants had to give up some of the money they earned to burn the money of others. A distribution was made of different prices of burning in different groups. So burning would come to a cost for the person who burns money units. After the burning stage participants could convert their money units to real money.

We think that this experiment has a problem called experimenter demand effects. This is a problem in experimental economics and is based on the possibility that subjects will act to what they think the experimenter wants from them (what he or she is testing for). In this case they get the opportunity to burn money, so they might think burning is appropriate. This could lead to distorted results. In their research Zizzo and Oswald (2001) try to correct for these effects by avoiding the use of the word 'burning' in the experiment description and stressing out that the subjects could not earn any money units by eliminating (burning) the earnings of other subjects. However in our opinion more could be done to avoid these effects. One could embed the burning stage in several other stages. Researcher might even add games that explore social preferences to avoid experimenter demand effects.

However the findings in these experiments of Zizzo and Oswald (2001) are very interesting and similar to experiments in which there were better measures to counteract experimenter demand effects. Zizzo and Oswald (2001) found that over 50% of the subjects decided to burn money of others (and were even willing to pay for this). They also looked if there was a price elasticity for burning. This means that they wanted to check if the price of burning rose the burning rates went down. Up to a point the price elasticity was almost zero (when the highest price for burning was reached burning rates dropped). The fact that people were willing to give up their own money to burn the money of others shows that people are not purely self-interested or purely altruistic (otherwise they would allow the richer subjects to keep their money).

The question arises which of the factors (envy, fairness or something else) contributes the most to this. Because of the fairness aspect of the experiment (extra betting possibilities randomly assigned participants and more endowment randomly assigned to participants) we think fairness plays a great part in the findings of these experiments. Subjects can view the randomly awarded extra betting money and a bonus as an unfair advantage and maybe want to punish them for that.

Zizzo (2002) ran a similar experiment, but in this case only one random decision was chosen from all the burning decisions made by participants (after all subjects had made their burning decision). Almost 50% of the subjects engaged in burning money. We think however that because of this random implementation the decision to burn is a lot easier. Participants could feel less bad about their burning

decision because of the chance of implementation is smaller than when one's burning decision is implemented with certainty. Participants might even feel it is not their fault the other got burned. This is because they might see it as bad luck for the other their burning decision was chosen. The decision to implement was not in the power of the burner.

Zizzo (2004) introduced a similar experiment. Here the researcher used a similar Money Burning Game; only in half of the groups the possibility of stealing was introduced. Participants were also allowed to redistribute their own earnings to other participants. In the stealing condition people were rewarded for their action to take money from other subjects. Again we might see the same problem of experimenter demand effects as stated in the previous section. According to the rational game theory discussed in the introduction we should see a stealing rate of 100% in the stealing condition. However they found that stealing was substantial but always less than 100%. This means that people are not purely self-interested. Burning was much lower in the stealing condition (though not zero). Some participants still use burning even though stealing has a reward attached to it (participants were allowed to keep the stolen money themselves). An explanation for this might be participants want to punish other participants who are out of line with social norms without any gain for themselves (pure altruism). This form of punishment is a social preference (as we have explained in the reciprocity part). When stealing was not allowed the burning rates increased drastically (from 8% to 20,20% on average). Although no percentage for stealing was presented in the paper, the author suggests that it is much higher than the burning percentage when stealing was not allowed. According to Zizzo (2004) this might implicate that a part of the stealing was motivated by antisocial preferences. However we may not underestimate the role of reward. Because of the fact participants gain money when stealing it might be that participants think stealing is good. One is rewarded for bad behavior, which might lead to more of this bad behavior. We can also see this with children. When a child is crying it might get a treat to get them to stop crying. The child might think it got rewarded for crying, creating the incentive to cry when it wants a treat. Zizzo (2004) thinks the role of reward in the stealing condition is not the only one because of the stealing rate found (which was lower than 100%). However we think it could be the case some of the participants stole because of the reward they received for it.

Another experiment done on antisocial preferences was the one shot money burning game by Abbink, Masclet and Van Veelen (2009). They wanted to check for the occurrence of reference point effects.

Their experiment included two treatments. In the first treatment participants played a money burning game. The participants received an initial endowment, after that they got the opportunity to decrease the endowment of other participants incurring some costs (the endowment of the burner was lowered with 10 money units). This treatment was called the negatively framed treatment. In the second treatment participants could choose to raise the income of the other participant with 50 money units

and earn 10 themselves. This treatment was the positively framed treatment and the choice not to increase the money of other participants was seen as the choice to burn. In this treatment burning also would come at a cost. Participant who did not decide to increase also did not receive the increase in endowment (which can be seen as a cost of 10 money units). In the experiments Abbink, Masclet and Van Veelen (2009) used a range of initial endowment, so they were also able to see if inequality played a role. A range from (50,50) initial endowment till (600,50) was used in the negatively framed disadvantageous inequality treatment and in the negatively framed advantageous inequality the researchers used a range from (70,50) till (600, 500). In the disadvantageous inequality with positive framing they used a range from (40,0) till (40.550) and in the positively framed advantageous inequality the researchers made use of a range from (60,0) till (590,0). The experimenters found burning rates which did not significantly differed between treatments. They found that the different framing did not cause the different outcomes.

Joy of destruction game

Another game designed, is the so called Joy of destruction game. This game was designed and used by Abbink and Sadrieh (2008). Two players have earned money by doing a task. After they have done the task the participants choose simultaneously to destroy the earnings of the other or to do nothing. Abbink and Sadrieh (2008) first use a multiple shot game (ten rounds were played). So they game was repeated for ten times to see if reciprocity would lead to more burning. Interesting extra aspects of this game were the treatments used. There was an open treatment and a hidden treatment. In the open treatment destruction could only be done by the participants and the other could see how much was destroyed. In the hidden treatment however there was a random destruction by nature. A dice was rolled in order to determine if and how much was destroyed. The participants who got burned only saw the total amount of destruction and not how much was done by 'nature'. This gave participants a possibility to hide behind nature (unless destruction was very high). The game was repeated to check if retaliation would start a vendetta (more and more burning) or that the fear of retaliation was enough to deter participant from destroying.

Abbink and Sadrieh (2008) found that destruction rates were about 8,5% in the open treatment and quickly fall down to almost zero after a few rounds. In the hidden treatment however destruction rates were significantly higher (39,4%) and stable at a positive rate. We think the higher positive rates found in the hidden treatment can have many reasons. Maybe they are nasty persons and do not want to get caught, it could be the case that participants want to punish antisocial behavior (which is a social preference as we have seen in the reciprocity part) and are not afraid of retaliation in the hidden treatment. It could even be the case that participants are curious about how it feels to be nasty and try it out in a hidden environment.

The experimenters try to avoid experimenter demand effects by embedding the game in multiple different tasks. They let the participants do several different tasks so they could hide the true reason of the experiment. Because of the diversity of tasks participants could not really gain insight in what is investigated and could not form expectations about what is tested. Because of this participants might be less likely to play the game to conform to the experimenter's research.

Another variant of this game is the Mini joy-of-destruction game (Abbink and Herrmann 2011). This is the one-shot variant of the joy of destruction game but an extra feature is that just like in the money burning game participants incur a cost of destruction. Abbink and Herrmann (2011) also use a hidden and an open treatment. However participants had to incur a cost of 1 money unit to destroy 5 money units of the other participant. In the experiment Abbink and Herrmann think the hidden treatment can be seen as a reduction of the moral costs of nastiness. It was not tested why people destroyed more in the hidden treatment. Another aspect in which this experiment is lacking is avoidance of the experimenter demand effect. Participants just get some money and then are allowed to destroy some of the other participant. No effort was made to counter experimenter demand effects. This means the outcomes of the experiment could be biased because participants might want to conform to the experimenters' research (that is finding antisocial preferences).

Destructor game

The destructor game (Kessler, Ruiz-martos and Skuse 2012) is a reversed dictator game. As we have seen in the introduction, the dictator game is a game in which one participant is assigned to split an amount of money and the other participant is passive. In the Destructor game one participant is assigned to be the 'destructor' and the other is the passive victim. The 'destructor' is allowed to destroy an amount of money earned by the passive receiver. There is also a random destruction which takes place in this game. At random some passive subjects get some of their earnings destroyed. So it is possible for the destructor to hide behind nature when destroying the others' endowment.

(Kessler, Ruiz-martos and Skuse 2012) experimented with their Destructor game. They did an experiment in which they used a one shot Destructor game. Participants earned 1000 tokens by filling in a personality questionnaire. After that, participants were randomly assigned to the role of destructor or passive victim. Each destructor was allowed to destroy 0%, 20% or 40% of the endowment of the victim. There also was a change of nature destroying. Nature would destroy the endowment of 20% of the victims with an equal chance to destroy 0%, 20% or 40% of the endowment. So destructors could to some extent hide behind nature. In this experiment retaliation plays no role, because it is a one shot game and participants therefore cannot react to the action of the other participants. In this experiment all other motives for destruction ((pre-emptive) retaliation, money reasons etc.) were removed. This was done by making in a one-shot game and there could be no preemptive retaliation because only one

participant was allowed to destroy money. Also the destructor did not receive any money when destroying the endowment of the other participant so money could not be a reason for destruction. The experimenter demands effects were countered by embedding this game in five others tasks.

Still the experimenters found that 15,5% of the destructors choose to destroy. Because other motives for destruction were removed, this meant that some people might like to destroy the earnings of others. The most frightening part of this experiment is that they also tested for the personality of the participant but did not found a significant difference between the personality of the participants and their burning behavior. No personality trait can predict the choice to destroy money from another participant (so all the participants might be capable of destroying). The researcher used a NEO-FFI personality questionnaire which tests for different personality traits (neuroticism, extraversion, openness, agreeableness and conscientiousness). The Leveson Self-Report Psychopathy questionnaire was also used to assess the personality. This tests measures impulsivity and checks for poor behavior control.

Vendetta game

This game introduced by Abbink and Herrmann (2009). The purpose of this game was to gain insight in the motives underlying hostile group relationships. This game looks a lot like the Money burning game and de Joy of destruction game. In this game two groups of four participants do a task to earn money. After this task each player is allowed to pay a certain amount to reduce the earnings of all of the members of the other group (so no one can be singled out). To check for the impact of reciprocity this game is repeated for ten rounds.

Two treatments were used by Abbink and Herrmann (2009). The first one was exactly as described above. Participants could earn money by answering questions to gain a certain amount of endowment. When the answer was correct, 60 money units were awarded, when the answer was wrong participant could try again, but it lowers the endowment with 5 money units with each try. To avoid losses the lowest endowment a participant could earn with answering the question was 50 money units. After the earning stage participants could lower the endowment of all four members of the other group. This destruction costs 10 money units for the destructor and would lower the endowment of all group members with 10 money units (so 40 in total).

The second treatment included a price. This price was randomly awarded to one person who destroyed money (this was done in each group). This price was lower than the costs of destruction. The reason for this was to ensure participants would have no economic incentive to destroy.

Interesting to see is that in the first treatment Abbink and Herrmann found destruction rates were around 20% in the first round, but quickly decreases to about 10% in the next few rounds. In the

second treatment destruction rates were about 60% in the first round and decrease to around 40%. In the second treatment destruction rates were much more volatile compared to the first treatment. The authors conclude the high rates in the first rounds were the result of pure nastiness and spite. However this could not trigger vendetta's because the fear of retaliation is deterring enough to stop high destruction.

We think Abbink and Herrmann (2009) are a little bit quick in their conclusions. Because of preemptive retaliation participants might be tempted to higher destruction rates in the first rounds. Although the concept of preemptive retaliation is mentioned in a lot of research, we found that none of the researchers checked what the motive is of preemptive retaliation. It might be a social or an antisocial preference. Experimenters simply assume preemptive retaliation means people are bad. One can explain this behavior as the expectation to get burned by others, which suggests people believe others behave in an antisocial way. However correcting for possible behavior which is not in line with social norms might make preemptive retaliation a social preference. Participants want to correct in advance the possible behavior that is not in line with social norms and therefore punish others in advance (which is very inefficient, but not impossible). The explanation given by Abbink and Herrmann (2009) was that people are evil or expect other to be evil and therefore the burning found is evidence of antisocial behavior.

Comparison of new games

All these games are designed to search for the occurrence of antisocial preferences. We should check how consistent the outcomes of these researches are. To get a better view on the outcomes of the experiments we have put all the outcomes in a table (see table 1)

Author	Game	Burning rate
Zizzo and Oswald 2001	Money burning game	62,5%
Zizzo 2002	Money burning game	50%
Zizzo 2004	Money burning and stealing game	With stealing: 8% Without stealing: 20%
Abbink and Herrmann 2009	Vendetta game	Normal treatment: 13% Prize treatment: 40%
Abbink, Masclet and van Veelen 2009	Money burning game	Negativley framed: 25% Positively framed: 24%
Abbink and Sadrieh 2008	Joy of destruction game	Open treatment: 8,5% Hidden treatment: 40%
Abbink and Herrmann 2011	Joy of destruction game	Open treatment: 10,8% Hidden treatment: 25,8%
Kessler, Ruiz-Martos and Skuse 2012	Destructor game	15%

As one can see burning rates differ a lot between different experiments and games. Different outcomes might mean that the tests are not accurate. Keep in mind however not all games are the same. The design of a game can determine how people react. As we have seen in some experiments the experimenter demand effect may have influenced the outcomes.

Because the experiments done by (Abbink and Sadrieh 2008) and (Abbink and Herrmann 2011) look a lot like each other (same treatments used and similar methods used) we may compare these two. The

games were also both researched by the same researcher. We might see how accurate the experiments are. These games are almost the same and for the open treatment the burning rates are similar. For the hidden treatment however we see a difference in burning rates of 14,2% which is a lot. These experiments differ in one way. (Abbink and Sadrieh 2008) used a repeated Joy of destruction game whereas (Abbink and Herrmann 2011) use a one shot game. Another thing that might have influenced the outcomes of these experiments is the participant pool used. The first using undergraduate students from the University of Tilburg (The Netherlands) and the second using students from three universities in Ukraine. Differences in the cultures of these countries might have caused outcomes to differ just like in the public good experiment done by Herrmann et. al (2008). Besides these explanations it could also be possible that the experiment has a design which is not capable to replicate the same outcomes. Only repeating these experiments can proof if the design of the experiments is capable of replicating the same results.

Public goods game

The public goods game is a game that has been around for a long time. A public good is a good that is non-rival and non-excludable. Non-rival means that the use of the good by one person has no implication for the use by others. If we take a pizza for example (Rosen and Gayer, 2008). If one has a pizza and start to eat a slice, somebody else won't be able to eat that same slice of pizza. So there is a rivalry for eating the pizza. Because of this Pizza is not a public good. Non-excludible means that no one could be refrained from the use of the good. Again we take a look at the pizza. We can easily prevent someone from eating our pizza. It is our pizza, so we decide who gets a slice. The fact we can exclude someone from the consumption of the pizza makes it excludable. Another aspect of a public good is the fact there is no market for these goods, this is because it is either too expensive or one cannot determine a prize for it. An example of a public good is the police force. Everybody is able to make use of this and no one could be excluded from using this good. Because this public good is no one's property and everybody is allowed to use this it cannot be valued (everybody sets their own value on these goods). Also it cannot be paid by a person (the police force is paid by the government which is not a person). The government uses the taxes paid by all the people to pay for the police force. In this way everybody has to chip in and pay for it (even though not everybody uses it).

The public goods game is a game in which a group of subjects receive a certain endowment and are than asked to invest money in a project. The return on these projects per person is mostly lower than the amount invested by that person. This means if one invests 1 money unit the return is less than one. On first sight this might not look profitable, but if everyone invest in this can be very profitable. Look for example to participants who receive an endowment of 30 money units and we have an investment in which the return on every money unit invested is 0,6 money unit. Investing one money unit leaves

the investor with 0,6 money units. So investment is not profitable for one investor. Now take 5 investors all having 30 money units. If they all invested all the money units received, they will have a return of $150 * 0,6 = 90$. This is three times their initial investment. An interesting part of this game is that people have the possibility to free ride. Even if someone does not invest a single money unit he or she will still receive a part of the returns (remember no one can be excluded).

Interesting for our research is the fact that a public goods game can be adapted to look for antisocial preferences. The way by which this is done is simple. Experimenters have added the possibility for participants to punish non-cooperators (Herrmann et al. 2008). We used Herrmann et. al (2008) because this article was mentioned a lot by designers of new games. It is possible more researchers did similar experiments. We also do not know when experimenters first used this adaption of the public good game to check for antisocial preferences. The interesting part for us is the fact that non-cooperators could also punish the others.

In this experiment the researchers mainly looked at the cultural differences in the contribution of a public good game. The researchers also looked at the punishment of high contributors (which is interesting for our research). Antisocial punishment was seen as a participant who had punished someone who contributed the same or more to the project. So a participant say Jack has not contributed a single money unit to the project but punishes another participant say Amy who did contribute to the project. Amy was in line with social norms but still was punished by someone who did not conformed to social norms.

They found that antisocial punishment varied a lot across the different subject pools. Antisocial punishment has a significant effect on the cooperation. In the groups in which punishment was allowed contributions were significantly depending on the amount of antisocial punishment. This means in groups where there was a high rate of antisocial punishment contributions were significantly lower than in groups with a low rate of antisocial punishment. Maybe contributing participants were scared of the reciprocity of free riders and therefore decided not to punish them but instead they contributed less to the project (indirect punishment).

The question remains why participants who did not contribute to the project decide to punish participants who did contribute to the project. We think it might be that these non-contributing participants got angry because they were punished and decided to hurt the ones who punished them. It might also be the case that participants who did not contribute thought of their contributing colleagues as cocky and therefore decided to punish them. It is possible antisocial punishment has to do with social norms. In a society where social norms are very strong and widely accepted people will have more incentive to conform to these social norms. In a society where this is not the case people might be more likely to engage in antisocial punishment because social norms are not as strong. The

enforcement of rules could also have impact on the amount of antisocial punishment (some people want to undermine authority or react negatively if they are forced to act in a certain way).

What is the role of fairness, inequality and reciprocity?

Fairness and inequality in the experiments. In the experiments reviewed above, researchers were not only trying to find the occurrence of antisocial preferences but some of them also tried to find what role fairness and inequality had on the burning rates found.

Zizzo and Oswald (2001) designed their experiment to look for rank egalitarianism and check whether deservingness mattered. They found rank egalitarianism and that deservingness had an influence on the burning behavior of advantaged participants. The difference between the relative payoff and the way participants earned their endowment could not be explained by the burning activity of poorer participants (they burned all the richer participants with no regard for the way the endowment was earned). The deservingness factor could be found in the case of the richer participants who earned their endowment by pure luck. Rich participants who earned their endowment by pure luck engaged more often in burning others in comparison with other rich participants who had earned their endowment by working hard. Maybe the undeserving advantaged participants thought they would get burned more often because their endowment was undeserved (they did not have to work hard for their endowment) and therefore started to engage in preemptive reciprocity. However the data found in the experiment pointed out that the deserving factor has more to do with burning in comparison to the reciprocity factor. So people who were undeservingly advantaged got burned more and did not reciprocate more often.

Just like in the experiment by Zizzo and Oswald (2001), Zizzo (2004) found that participants were rank egalitarian (around 80%) and the view of the participants on the fairness of the distribution was influenced by the relative payoff and the way the endowment was earned.

In their experiment Abbink, Masclet and Van Veelen (2009) looked for inequality aversion and framing effects. Although framing had no significant effect on the amount of nastiness exerted by participants, it had some interesting effects on inequality aversion. Inequality aversion means people do not want to have inequality. They want everybody to have the same endowment (or utility). Even when this means one has to give up some of his own utility to restore the equality. Abbink, Masclet and Van Veelen (2009) found some interesting results of framing on inequality aversion. In the negatively framed situation two things emerged. First, it seems like participants act like they are equality averse instead of the assumed inequality aversion. This equality aversion was shown by the higher burning rates in the situation where participants have the same initial endowment. The second phenomenon found was that burning rates were significantly higher when the participant was ahead

(had a higher initial endowment). This might be explained by competition. Participant might think they have to have the highest endowment. Burning someone who is behind to make sure they will stay the richer one of the participants. The reason for the latter phenomenon might be the fact the endowment is distributed at random. Participants get randomly assigned an initial endowment and might experience this as undeserved (they did not have to do anything for it) and expect to get burned anyway, so they exert preemptive retaliation. In the positively framed treatment Abbink, Masclet and Van Veelen (2009) found something different. No high burning rates were found when the initial endowment was equal. In the case of unequal initial endowment burning rates were opposite to the ones found in the negatively framed treatment. Burning rates were significantly higher when a participant was behind in endowment, which is in line with inequality aversion. However the authors note their results may be influenced by the occurrence of the possibilities in real life. Opportunities to destroy are present everywhere while opportunities to create something are very rare.

What we have seen here is that fairness and inequality can be a strong motive to burn, destroy or steal endowment of others. So fairness and inequality might be arguments for the burning and destroying seen in the experiments. As we have seen in all of the reviewed experiments burning and destruction rates were positive. In the experiment of Herrmann et. al (2008) we can see people reciprocate on participants who do not contribute enough. Also in the experiment done by Zizzo and Oswald (2001) reciprocity can be found. Participants were even willing to sacrifice some of their own endowment to reduce inequality.

In most experiments seen above there also might be preemptive retaliation. Because some participants think they will get burned, they burn in advance to reciprocate on the burning that might be done by the other participants. Preemptive reciprocity might explain the higher burning rates in the first round in Abbink and Sadrieh (2008) and Abbink and Herrmann (2009). In these experiments the burning rates were higher in the first rounds, but then lower after a few rounds. We think this is because participants might see no burning of the others and then change their expectations concerning the burning behavior of others. The moral costs of burning have been obviously lowered making burning less costly. The lower moral costs might explain why the experimenters found positive burning rate persisting throughout the hidden treatment. More research has to be done to find out if this is due to the pleasure of people to hurt others or that other factors might play a role.

We will now conclude on the role of fairness, inequality and reciprocity. As we have seen earlier, these concepts are closely linked. Fairness and inequality are the main reason for reciprocity. Fairness does not only incorporate the fairness of distributions, but also the fairness of behavior and procedures leading to this distribution. We think fairness and inequality are the main drivers behind the antisocial preferences found in the experiments. In all the experiments reviewed, positive burning amounts

occur. However, only the public goods game (Herrmann et al. 2008) is stated how they measured antisocial behavior. In the rest of the experiments described earlier researchers are talking about antisocial preferences, but do not look at why people really are burning or destroying. This can have many other reasons (get others back in line with social norms etc.) and most of the authors state burning and destroying as the 'dark side' of human nature and ignore other possible explanations. People might do it out of inequality concerns (Abbink, Masclet and van Veelen 2009) or even to punish another participant who does not behave in line with social norms (Herrmann et al. 2008)

We do believe a small part of the burning is done by people who gain pleasure out of the pain they incur on others. We think this because in the destructor game (as we have seen) most the other incentives to burn were removed and still destruction took place.

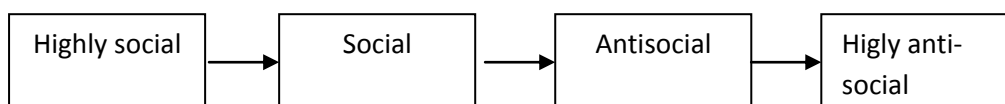
How do social and antisocial preferences fit together?

The evolution of research from social to antisocial preferences seems like a natural one. First researchers found evidence of social behavior. Second they started to explore the range of preferences further. Rejections of high offers in the ultimatum game and dictator game made researcher question rational game theory. Where these rejections based on biased judgments of the receivers who might react emotionally to a low offer (because they were offended). Researchers started to incorporate forms of punishments in these games. A third unbiased participant was added to the game who would act as a referee with a possibility to punish the proposer if he deemed the proposal as unfair. After this researcher found evidence of antisocial punishment in a few games. One example of this is the antisocial punishment in the public good game of Herrmann et. al (2008). Antisocial punishment might have triggered research to investigate the 'dark side' of human nature. Games like the money burning game, the joy of destruction game and many more arose to check for the occurrence of antisocial preferences.

Researchers might see social preferences and antisocial preferences as different things. The reasons for this might be because researchers do not believe the results found in experiments done on antisocial preferences. It might be experimenter demand effects have biased experiments. When this is the case results found in an experiment can be misleading. We thought experimenter demand effects might have influenced the experiment done by Zizzo and Oswald (2001).

We think however research on social and antisocial preferences is very alike. Some of the behavior found has the same explanation. Participants might be altruistic because they do not favor the inequality or the unfairness. People may burn money to do the exact same thing (lowering inequality and unfairness).

Preferences might even be put in a scale. We want to show a scale of preferences. In the research done we can sometimes see unrealistic social behavior (a very high offer in the ultimatum game or dictator game). Then there is the regular social behavior seen as the reduction of inequality and unfairness (either by advantaged or disadvantaged participants). After that we have the more antisocial preferences embodied by the higher burning rate in the hidden treatments seen in the experiments of Abbink and Sadrieh (2008) and Abbink and Herrmann (2011). This might be explained by the lower moral cost of burning when they are allowed to hide their burning behind nature. Also we can see this in the higher destruction rates in the prize treatment of Abbink and Herrmann (2009). The prize lowered the costs of destruction. Another explanation for the antisocial preferences might be curiosity. Participants might be curious about the feeling of doing something bad. Finally there are the preferences we will call pure evil. Just like there are some people who are extremely altruistic, people who are bad exist. In real life this can be seen in many situations. The random destruction of property or the random murder of people etc. Figure 2 shows a scale of preferences. It starts with positive behavior and when going to the right the behavior can be classified as more negative until the most negative behavior which is highly antisocial behavior.



(Figure 2)

This is how we think social and antisocial preferences fit together. There should not be separation between the both. Research done on both social- and antisocial preferences is important to complete any model of behavior. It is always relevant to incorporate all aspects, so to form a complete model of preferences researcher incorporate our scale. Because this is the full range of human behavior. It is important to realize this scale can be applied on an individual level as well as on a population. People might behave in not the same when they face different situations. This might be due to the reference point effect explained under the heading of inequality. Also people are very different compared to each other. So not everybody reacts the same when facing the same situation. We can use the scale to compare behavior within a population when looking all the different reactions.

Conclusion and suggestions for further research

At the beginning of this article we asked ourselves if antisocial preferences existed. First we looked at the concepts of fairness, inequality and reciprocity to gain more insight in these important aspects of behavior. We found fairness was something that mattered when people make choices. People care for fairness, because of the need to belong to a group (De Cremer and Blader, 2005). Also fairness played a role in the distribution of endowment. Fair distributions matter. This can be seen when looking at investigations concerning prior knowledge when trading in stocks and bonds. Sometimes fairness creates an unequal distribution. This leads us to the part of inequality. People most often are inequality averse, meaning they do not like inequality and wanting to eliminate as much of the inequality as possible. Fairness and inequality are very strong motivators for actions which might be seen as antisocial. People want to create a fair and equal distribution of endowment. If this is not the case they might burn money. We think this is not an antisocial preference because people do not hurt others to gain pleasure. We think this burning is because of the social desire for equality. Reciprocity is also a mean to punish people who don't behave according to social norms. We have seen two kinds of reciprocity, namely positive and negative reciprocity. The first is returning good behavior. The second consists of two parts. The first part is social negative reciprocity which was punishing others to make them convey to social norms or to redistribute endowment to create a more equal distribution. The second part we found was punishment of people who have not deserved it. It might be people punished because they did not convey to social norms punish back without any good reason.

Although researchers most of the time looked for the occurrence of social preferences, a new branch of research has emerged. The research on antisocial preferences. Researchers designed new games or conveyed game to look for the occurrence of antisocial preferences. All the experiments we have reviewed, found positive rates of burning and destruction. The accuracy of these percentages remains to be checked because these experiments are not exactly the same. Sometimes they use a price (Abbink and Herrmann 2009), they look for framing effects (Abbink, Masclet and van Veelen 2009), they use hidden and open treatments (Abbink and Sadrieh 2008 and Abbink and Herrmann 2011) or they introduce stealing in the experiment (Zizzo 2004). Because of these differences it is hard to make a good comparison between them.

Also some of the experiments might be biased by the experimenter demand effects (Zizzo and Oswald (2001), Zizzo (2002) and Zizzo (2004)). The experimenter demand effect was that participants altered their behavior to fit to what they think the experimenter wants to find.

Other experiments did not look for the possibility of other explanations for burning (Abbink and Herrmann (2009), Abbink and Herrmann (2011)). They assume people are bad and therefore do not look at any other possibilities for the burning rates found. Maybe participants burned to let others

convey to social norms or they wanted to reduce the inequality between participants. Even though the researchers tried their best to eliminate all the other incentives for burning, they did not completely remove these incentives. Maybe social concerns lie at the base of the reported burning behavior or people are just curious of how it will feel to do something bad. We think only a small percentage of burning is really because people are bad. In the article from Herrmann et. al (2008) we can clearly see when there is no real reason for the reciprocity. In this experiment a few participants punish for no reason but the fact they are angry they got punished or just because they enjoy the pain they incur on others.

So there is evidence for the occurrence of antisocial preferences. Here we mean the occurrence of behavior that is not in line with social norms and punishment is due to no good reason. Mostly we think the research found evidence of the occurrence for behavior that might be seen as antisocial but is actually a social preference. Reducing inequality or punishing others to make sure they stay in line with social norms.

We also tried to find a way to show how these antisocial preferences might coincide with the evidence found on the occurrence of social preferences. In this process we designed our own scale of preferences. Ranging from highly social preferences to highly antisocial preferences. Behavior occurs in stages and for many different reasons. Not every person is the same. Some people are donors and others steal from their fellow man. By introducing this scale we can more easily categorize behavior found.

We think highly antisocial preferences are existing but these experiments should be repeated more often and need to be corrected for their flaws.

Suggestions for further research

The research on antisocial preferences is a relatively new one and a lot of research has to be done. First we would like to advise to repeat the experiments done. Repetition of experiments can be used to see if the same results can be generated. Experiments done differ on essential aspects and are very hard to compare.

Although money is the ultimate good, it may be interesting to see what would happen if we find a way to use real goods. Maybe because of the endowment effect retaliation will be harder, or maybe people do not want to burn because it is much more personal to destroy someone's goods than to destroy their money. We can explain the endowment effect as the fact people value good they own more than would if they did not own it. In a bargaining game participants want more money for goods they just got then others are willing to pay for it. We think that because the good is more valued by the owner punishment done by someone who puts a lower value on the goods might overshoot more easily. We

have not even mentioned the emotional value attached to goods because of the emotional value we think is a lot more personal when you destroy a good than when you destroy money.

Another interesting aspect of the research on antisocial preferences might be gender. Men and women are different in behavior so maybe they are also different concerning antisocial preferences. It would be nice to check if women and men also differ in their burning behavior. None of the reviewed experiments researched this. Herrmann et. al (2008) checked for cultural differences, but not for gender.

References:

- Abbink, K., and Herrmann, B., (2011), The Moral Costs Of Nastiness, *Economic Inquiry*, 49(2), 631-633.
- Abbink, K. and Herrmann, B., (2009) Pointless vendettas, CREED Working Paper Series, 2009.
- Abbink, K., Masclet, D., and van Veelen, M. (2009), Reference Point Effects in Antisocial Preferences, CBESS Discussion Paper 09-03.
- Abbink, K., and Sadrieh, A. (2008), The pleasure of being nasty, Working paper, University of Amsterdam.
- Bereby-Meyer, Y. and Niederle, M. (2005), Fairness in bargaining, *Journal of Economic Behavior & Organization* vol. 56, 173-186.
- Berg, J., Dickhaut, J., McCabe, K., (1995), Trust, reciprocity and social history, *Games and Economic Behavior* 10, 122–142.
- Bolton, G.E., Ockenfels, A., (2000) A theory of equity, reciprocity and competition, *American Economic Review* 90 (1), 166–193.
- Bruni, L., Gilli, M. and Pelligra, V. (2008), Reciprocity: theory and facts, *Int. Rev. Econ.* 55, 1-11.
- Charness, G. and Rabin, M., (2002), Understanding social preferences with simple tests, *Quarterly journal of economics* 117, pp. 817-869.
- Cremer, D. and Blader S.L., (2006), Why do people care about procedural fairness? The importance of belongingness in responding and attending to procedure., *European Journal of Social Psychology* 36, 211-228.
- Falk, A. and Fischbacher U. (2006), A theory of reciprocity, *Games and Economic Behavior* 54, 293-315.
- Falk, A., Fehr, E. and Fischbacher, U. (2008), Testing theories of fairness. Intentions matter., *Games and Economic Behavior* 62, 287-303.
- Fehr, E., Schmidt, K.M., (1999), A theory of fairness, competition, and cooperation, *Quarterly Journal of Economics* 114, 817–868.
- Herrmann, B., Thöni, C. and Gächter, S. (2008), Antisocial Punishment Across Societies., *SIENCE* vol. 319, 1362-1367.
- Hoffman, E., McCabe, K., Shachat, K., Smith, V.L., (1994), Preferences, property rights, and anonymity in bargaining games, *Games and Economic Behavior* 7, 346–380.

Kessler, E., Ruiz-Martos, M., Skuse, D., (2012), Destructor Game, Department of economics
Universitat Jaume I.

Rosen, H.S. and Gayer T., (2008), Public Finance, New York, McGraw-Hill.

Sethi, R. and Somanathan, E. (2003), Understanding reciprocity, Journal of Economic
Behavior and Organization vol. 50, 1-27.

Zizzo, D. J., (2004). Inequality and Procedural Fairness in a Money Burning and Stealing Experiment,
Department of Economics Discussion Paper number 155.

Zizzo, D.J., (version March 2002), Fear the evil eye, Department of Economics Discussion Paper nr.
91.

Zizzo, D.J. and Oswald. A.J., 2001, Are People Willing to Pay to Reduce Others' Incomes?, Annales
d'Économie et de Statistique N°63-64, pp. 39-62.