Do people exhibit more antisocial behavior if the income allocating process has been unfair?

(Draft January 2020)

Abstract

We examine whether an unfair process of income allocation leads to a higher degree of antisocial behavior. In order to test this hypothesis, we run an experiment where we vary the way players' endowments are determined: by a fair, random, or unfair process. The initial distribution has a certain degree of inequality, which is held constant across treatments. After receiving their income, subjects can anonymously reduce the income of another player at a cost. The overall frequency and percentage of destruction is similar and not significantly different across treatments. Surprisingly, even if money is allocated in an unfair manner, subjects do not destroy more.

We furthermore elicit subjects' perceptions about the fairness of the income-generating process. They are in line with the intended treatment effect, but we find almost no correlation between subjects' fairness evaluation and the propensity to burn money. The findings indicate that the degree of antisocial behavior is rather constant in this context and independent of the fairness of the income-allocating process. Subjects' justifications of their decision and insights of related studies suggest that the decision to destroy other's income depends a lot on whether other subjects can be held responsible for the initial (unfair) distribution.

1. Introduction

Many economic experiments have demonstrated that subjects exhibit *antisocial behavior*. The term antisocial behavior in this context is used in the sense that subjects destroy each other's income without any material benefits to themselves, even if it is costly. The amount burned in this way varies a lot, and depends on the exact framework in which subjects are acting. The percentage of destruction in such experiments ranges from below 5% to over 60% of total endowment. Previous studies focus on the effect of variables like *the costs of burning* (Zizzo and Oswald, 2001), *the degree of anonymity* (Abbink and Sadrieh, 2009) and *relative income positions* (Abbink et al, 2011 as well as Grossman & Komai, 2013), just to name a few. For some factors such as anonymity and the price of burning, the results seem expectable and quite clear: Abbink and Sadrieh (2009) find that a higher degree of anonymity leads to more money burning. In the experiment of Zizzo and Oswald (2001) subjects burn significantly more money, when costs are very low (2.5% and 5% of the amount destroyed) compared to a higher price (25%). However, for other factors such as the relative income position, there are contradictory results so far concerning the amount of destruction: Abbink et al (2011) show that subjects with a similar rank in the income distribution destroy more, while Dawes et al. (2007) find an opposite effect.

These studies suggest that the degree of destruction depends a lot on the context of the situation. In most studies discussed above, there exists a certain degree of inequality in the initial income distribution. This potentially creates a conflict: Under some circumstances these differences might be accepted, while on others subjects might have the desire to change the initial unequal allocation. Taking into account prominent theories of justice, one key factor is ***the fairness of the money allocating process***. This aspect has not been examined in the context of antisocial behavior so far.

In this study we investigate, if the degree of antisocial behavior can depend on "how fair" subjects view the money allocating process. The natural hypothesis we want to test is, if participants burn more money, when the endowment allocating process has been unfair, compared to when it has been fair.

If we use the term *fair* in the context of our study, e.g. for labelling the treatments and forming the hypothesis, it mostly corresponds to the concept called *process fairness/procedural fairness*. Concretely this means that subjects' incomes are positively correlated to their time and effort spent during the experiment. We also use the term to describe the subjective fairness evaluations of our participants concerning the endowment-allocating process. These fairness evaluations are also based on the concept of process fairness and the respective question is framed accordingly.

Process fairness has to be distinguished from the one of *outcome/distributional fairness.* Both play an important role in assessing a distribution of resources. Distributional fairness emphasizes the resulting distribution, while process fairness focuses on the way how this distribution is generated. So one can say, process fairness takes more an ex ante view, while outcome fairness judges the result more ex post. Let us consider for example a lottery, which awards the winner a high prize and the rest receives nothing. If everyone has the same probability of winning, this mechanism satisfies the criterion of process fairness, but not that one of outcome fairness.

Generally, in economic experiments, endowment is mostly provided in the form of windfall gains -the most common way- or as earned money. In the latter case participants usually perform a (real-effort-) task according to which their initial income is assigned. Evidence suggests the process determining initial income affects subsequent subjects' behavior: In an Ultimatum Game lower offers are accepted if the proposer has earned his role by winning a logic game (Hoffmann et al 2008). Similarly, Oxoby and Spraggon (2008) observe larger transfers in a Dictator Game, if the recipient has created the amount of money to be divided by his performance in a problem solving task.

Most fairness norms and theories on procedural justice emphasize the role of effort as an important factor justifying differences in the distribution of (financial) resources (see Konow 2003, p.1207). In contrast there are differences resulting from luck, or factors which are congenital such as abilities or talents. Usually people favor a more even distribution if initial differences result from external reasons (e.g. a handicap) compared to internal reasons (e.g. lower effort provided or bad decisions made in the past) (see Farvelli, 2007 and Konow, 2001). Römer (1998) goes even one step further and argues, that effort can also be partially seen as some characteristic of type, for which people cannot be held responsible completely.

For further information on theories of distributional justice see Capppelen et al. (2007, 2013) and for an overview of results in empirical social choice see Gaertner and Schokkaert (2012).

A closely related paper is Fehr (2018). The author examines in a lab experiment the relationship between increasing inequality and the tendency to burn other's income. He finds that increasing inequality leads to more money burning, but only if the underlying process creating this higher inequality is unfair. In contrast, if higher inequality can be unambiguously attributed to higher effort, subjects do not destroy more income.

Taking all this together, a possible conjecture can be that the perceived fairness of endowment determination is an important factor influencing the decision to change/destroy others' income.

The results of this study are important for the real world. One can think of many situations, in which resources are assigned by different procedures, which vary in their perceived fairness. For example companies provide different remuneration schemes depending on individual performance. Individual performance is though often not easy to measure and its value contribution can be hard to disentangle. The main difference to our experiment probably is that antisocial behavior usually is not directly possible in the real world. But in some cases even (costly) sabotage might occur. However, an unfair process can in any case trigger negative emotions and harm individual productivity. So, from society's point of view it is an essential question, if the resulting allocation is approved or not.

2. Experimental Design

The experiment is pen and paper based and consists of three treatments:

1. **Treatment "Fair"**,
2. **Treatment "Random"**
3. **Treatment "Unfair"**.

All treatments have the same basic structure, consisting of two parts: First, the endowment/income[1] determination phase and in the second part the destruction decision. Finally, subjects fill out a questionnaire and then receive their final payoff.

As endowment, half of the participants receive a **high (10€)** and the other half a **low (5€) amount of money**. The only difference between the treatments is the way in which participants are assigned these values:

In the random treatment, the endowment is determined by a lottery. Subjects pick up a sealed envelope from a box containing a note that they either are allocated 5€ or 10€. They are told that 50% of the participants receive the high amount and the other 50% the low one. This treatment is designed as a "baseline", to be comparable with most of the other money burning experiments with similar parameter values, in which endowment is provided in the form of "windfall gains".

In the other two treatments, the subject pool is divided into two groups, the "early group" and the "late group". Members of the early group have to perform a real effort task, involving correcting IQ-tests from another experiment. The subjects of the late group are told to show-up 30min later for the experiment and do not have to do any work. In the fair treatment all members of the early group receive 10€ as endowment and participants from the late group are assigned 5€. The payment scheme in the unfair treatment is exactly reversed: That means, subjects showing up early only receive 5€ and the others, who do not have to do any work, obtain the higher value of 10€.

The whole procedure is common knowledge. That means, in each treatment all subjects are informed what kind of "task" both groups of the respective treatment have to complete and how they are rewarded for it.

After the phase of endowment determination each subject is given the opportunity to spend a fraction of his or her income to anonymously destroy some or all money of a randomly selected other participant. The second part of the experiment is identical across all

[1] Note that we use the terms endowment and income in the context of this study synonymously

treatments. The costs of burning money are 10% of the chosen amount, so for every Euro a subject wants to destroy, he or she has to pay 10 Cents. They can choose any value ranging from zero up to the total income of the other person (either 5€ or 10€). Furthermore, in the instructions it was pointed out, destruction is only optional and one does not necessarily have to subtract any money. Framing here was as neutral as possible to avoid experimenter-demand effects in any direction. All decisions are made anonymously. For that purpose subjects generate a code, which corresponds to their identity and decisions during the experiment. Destruction decisions are made using strategy method: Every subject indicates which amount she or he would like to reduce, if the other person has an income of 5€ or 10€. Remember that subjects know which task the other participant (with an endowment of 5€ and 10€) had to perform beforehand. Afterwards, participants are randomly paired in groups-of-two. In each of these groups, only one of the two decisions is actually carried out ("unilateral destruction"). So in the end only half of the destruction decisions are implemented. The idea is to prevent motives like preemptive retaliation or negative reciprocity. If both decisions were to be implemented, it would be possible, that some players would not want to burn any money at all, but have the belief the others would do so and therefore would want to preemptively retaliate. This is maybe one of the reasons for surprisingly high burning rates in experiments, like for example, Zizzo & Oswald (2001) or the occurrence of vendettas in repeated money burning games (Bolle et al, 2013).

After everyone has made his or her choice, the experimenters randomly draw which decisions of the groups are actually implemented and then calculate the resulting payoffs. Meanwhile, subjects fill out a questionnaire, for which they receive additional money (3€). In the questionnaire subjects are asked, as how fair they rate the money allocating process and what are their motives for burning (or not burning) money.

At last, subjects receive feedback about which decisions are carried out and their final payoff.

The experiment was conducted in 2013 in the experimental lab of the University of Heidelberg. Overall 119 subjects, mostly students, took part (42% had their major in Economics). Recruitment was carried out with "ORSEE" (Greiner, 2004). Each treatment included 3 sessions, with an average duration of 45-60min. Further details are summarized in the table below. Average earnings were around 10€ per subject, with payoffs ranging from 3-13€.

Treatment	Procedure	Number of sessions	Number of subjects
Fair	Real-effort task	3	41
Random	Lottery	3	34
Unfair	Real-effort task	3	44

Table 1

3. Hypothesis

The first hypothesis is about whether subjects evaluate the fairness of the process of endowment determination differently across treatments.

Hypothesis 1: The fair treatment is rated as fairer than the random treatment, which again is rated as fairer than the unfair treatment.

Hypothesis two compares destruction decisions between the treatments[2]. Considering the discussion above about social norms and theories of distributional justice, it is natural to suppose that people destroy less money, if they perceive the endowment allocating process as rather fair[3]. This hypothesis is also supported by concepts from social psychology such as "Equity Theory" by Adams (1963). Accordingly, people tend to accept income differences as long as the proportion of effort to payoff is similar across all participants. If people perceive this relation as unbalanced, they experience negative emotions, leading to actions restoring a more even situation. In our setting, we would expect subjects viewing the income allocation as unfair to reduce the income of other players to some extent until subjective equity is restored. In the unfair treatment of our experiment this relation is clearly more unbalanced than in the fair one. Therefore one would expect higher destruction rates there to reduce differences in the effort to payoff-ratio. In a similar fashion, models of Inequality Aversion (as e.g. Fehr & Schmidt (1999) predict this kind of behavior, if effort is seen as some form of monetary costs.

Hypothesis 2: Most destruction is chosen in the unfair treatment unfair and least in the fair treatment.

To explain the composition of overall destruction, one has to look more into detail of the behavior of specific income classes. There are four different cases to distinguish, concerning the endowment of the decision maker and the target of destruction: {(low, low), (low, high), (high, low), (high, high)}. The first value refers to the endowment of the decision maker and the second to the endowment of the target.

According to the theories mentioned before, a substantial part of the predicted differences in overall destruction should stem from the combination (low, high). In this combination there are the highest differences in the effort-to-payoff ratio, especially in the unfair treatment. This leads to hypothesis 3:

Hypothesis 3: Destruction in the combination (low, high) is highest in the unfair treatment and lowest in the fair treatment.

---

[2] The argument in this paragraph is based on the assumption that subjects actually evaluate the endowment-allocating process in the treatments as differently fair. This assumption will be confirmed later on. Especially the process in the unfair treatment is rated as clearly less fair than the one in the other two treatments.

In the cases, in which both subjects have the same endowment (low, low) and (high, high) the situation is completely symmetric. Both parties have to perform the same kind of task and receive the same reward for it. There might be a certain number of subjects having inherently antisocial preferences. As there is no clear reason why this number should differ depending on the treatment, we expect no treatment effect in these combinations.

Hypothesis 4: Destruction in the combinations (low, low) and (high, high) is the same across all treatments.

Finally for the combination (high, low), most theories do not predict subjects with a higher endowment would destroy earnings of those with a lower endowment. Only in the fair treatment it is possible to imagine a subject receiving 10€ to feel disadvantaged. This would be the case, if she evaluates the additional 5€ of endowment as inferior to the effort she had to exert beforehand.

Hypothesis 5: Destruction in the combination (high, low) is highest in the fair treatment and lowest in the unfair treatment.

One could refine hypothesis 5 by adding, that destruction in the combination (high, low) is lower than in the combination (low, high). But the comparison of these two cases is more complicated, because the amount of endowment of the target differs. Therefore one cannot really compare absolute values in destruction between these two cases. Possible solutions might be, to look at the percentages or frequencies of destruction here.

4. Experimental Results

4.1 Fairness Evaluations

We first check if the treatments work as intended. In hypothesis 1 we anticipate that participants will find the unfair treatment least fair. To test this hypothesis we ask all subjects at the end of the experiment to evaluate as how fair they perceived the endowment assigning mechanism in their treatment. They can rate the process on a scale ranging from 1-5, where 1 means they perceived the mechanism as "very fair", 3 corresponds to "neutral" and 5 to "very unfair". The results are shown in table 2 below.

Treatment	Number of subjects	Fairness evaluation (average)
Fair	41	2.4
Random	34	2.5
Unfair	44	4.3***

Table 2

As expected, the participants rate the unfair treatment as clearly less fair than the other two treatments. The scale should be interpreted as ordinal, therefore we use a Wilcox-rank-sum test to compare the fairness perceptions pairwise between treatments. The difference of treatment Unfair compared to each of the other ones is highly significant ($p<0.001$). While there is no significant difference between treatment Fair and Random ($p=0.96$). A Kruskal-Wallis rank sum test of comparing all treatments simultaneously leads to the same result. The results (partly[4]) support hypothesis 1:

Result 1: The fair treatment is rated as similarly fair as the random treatment. Both are rated as much fairer than the unfair treatment.

4.2 Destruction Decisions (pooled)

In our experiment we measure destruction in two ways: Either as *destruction frequency* or as *percentage of destruction*. The first case corresponds to the number of decisions, in which one subject wants to reduce the payoff of another subject, divided by the total number of decisions. The latter case corresponds to the amount of money intended to burn divided by the endowment of the other subject. The hypotheses in general refer to both measures.

Overall destruction is moderate and less pronounced compared to other money burning experiments[5]. As explained before we measure destruction activity in two ways: Destruction frequency and percentage of destruction. We have two destruction decisions per subject. This amounts to a total of 238 decisions. The average burning frequency over all treatments is 23.5% and the average percentage of destruction is 10.8%. Results for each treatment are summarized in table 3 and illustrated in figure 1 (standard deviations are given in parentheses):

Treatment	Number of decisions	Destruction frequency	Percentage of destruction
Pooled	238	23.1% (0.39)	10.7% (0.21)
Fair	82	20.7% (0.40)	11.8% (0.26)
Random	68	25.0% (0.41)	11.8% (0.21)
Unfair	88	23.9% (0.37)	9.0% (0.15)

[4] As the difference in evaluations between the fair and the random treatment is not significant, the hypothesis is only partly confirmed. Maybe the difference between these two treatments would have been bigger, if participants would also have been told the allocation mechanisms of the other treatments.
[5] For example in Zizzo and Oswald, 2001 about 70% of subjects burn money at a cost of 10% of the chosen amount and in Abbink & Sadrieh ca. 25% do so, but at higher costs of 20%.

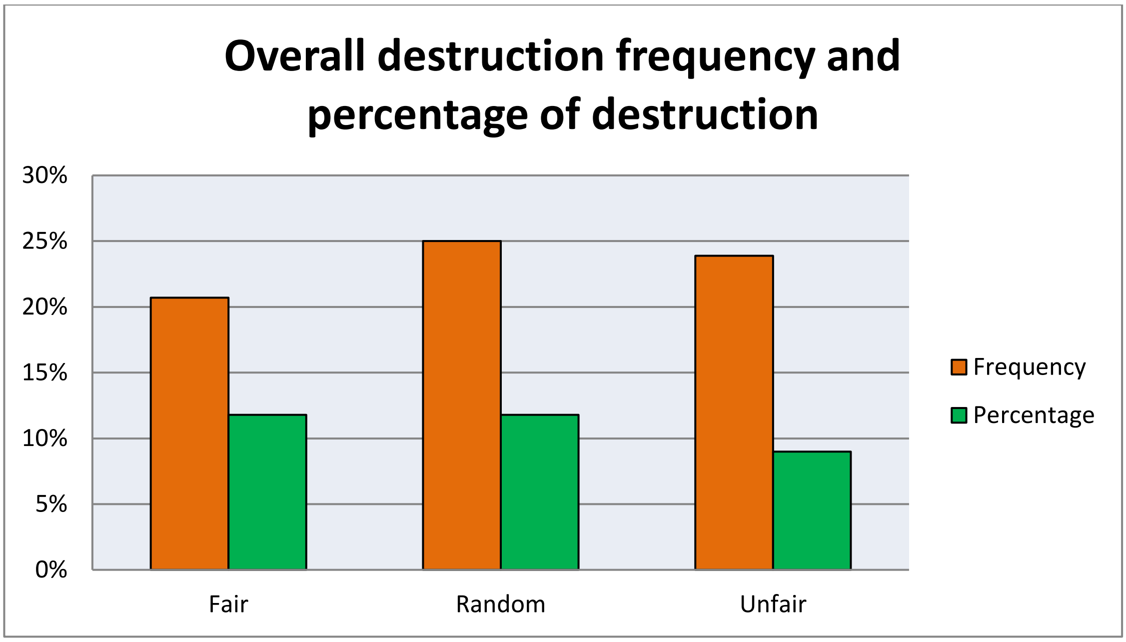

Figure 1

For the analysis we focus on the percentage of destruction, as this variable contains additional information compared to the mere frequency. Note again, that subjects make their decisions using the strategy method, specifying for each of the two cases (the other player has either an endowment of 5€ or 10€) the amount they would like to reduce from her. For the computation of our variables (burning frequency and percentage of destruction) we take both decisions into account, no matter, if they are actually carried out or not. Therefore the percentage of destruction per subject is calculated by adding up both intended destruction values and dividing them by the sum of the target endowments (which corresponds to 5€+10€=15€ in every case).

For the statistical analysis we use a Wilcoxon-rank sum/Mann-Whitney test (for pairwise comparisons of treatments) and a Kruskal-Wallis test to compare destruction decisions simultaneously across all treatments. All significant differences are indicated in the diagrams with arrows and stars. We first look at the pooled values and later examine the behavior of specific income classes both of the decision maker and the target. We additionally run a linear regression with the amount of destruction as outcome variable and the treatments as dummy variables (+controls). We do this both for pooled values and for controlling for specific income classes. Results of the regressions can be found in section 4.5. For the discussion and conclusion we focus on the results of the Wilcoxon-rank sum test, as some of the underlying assumptions of the regression analysis are not fully met (such as e.g. the assumption that the error terms are normally distributed).

Looking at the pooled values, there are no significant differences across treatments concerning the amount of destruction (Kruskal-Wallis test, $p=0.74$). A pairwise comparison of treatments shows similar results. In the unfair treatment, average values are even slightly lower than in the random one.

Result 2: There are no significant differences concerning overall destruction rates across treatments.

There are two potential explanations: Either the treatments do not have a strong effect on individual decisions to destroy somebody else's income, or we have multiple effects going in opposite directions and balancing each other on average.[6]

To check the second point we take a closer look at the behavior of specific income classes.

4.3 Behavior of specific income classes

Case 1: Combination (low, high)

In the first case we examine destruction behavior of subjects receiving a low endowment (5€) targeting subjects with a high endowment (10€). Average percentage destruction[7] per treatment is displayed in the diagram below:

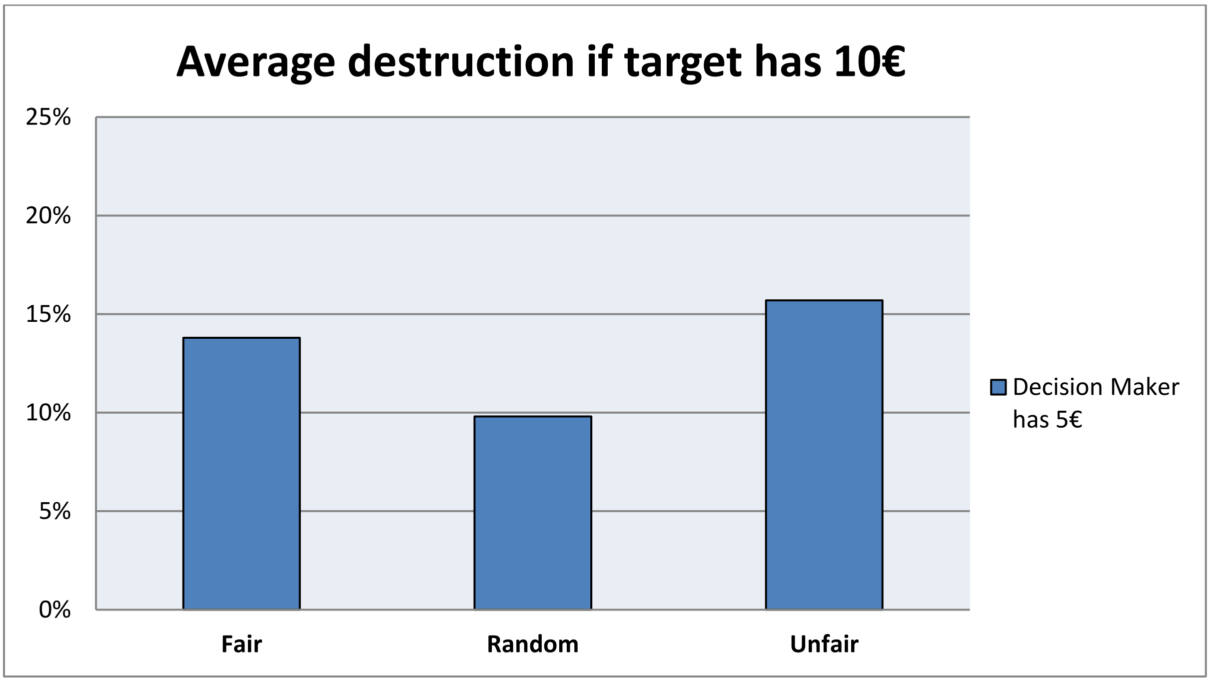

Figure 2

[6] It could for example be the case that in the unfair treatment subjects burn more money from players with a high endowment of 10€ compared to the other treatments, but less from players with the lower value of 5€. This would overall also lead to similar destruction rates across treatments.

[7] Note again that these are all intended values as the actual destruction, which is finally implemented is randomly determined.

Destruction rates are highest in the unfair treatment, but they do not differ too much across treatments. Comparing treatments pairwise using a Wilcoxon rank-sum test, we find no significant differences in destruction rates.

Result 3: There are no significant differences across treatments concerning destruction rates for the combination (low, high).

Hypothesis 3, in which we expected substantial differences in destruction rates across treatments, is not supported by the evidence. Possible explanations will be discussed later.

Case 2 and 3: Combinations (low, low) and (high, high)

For these two cases average destruction rates are displayed in figures 3 and 4. In the first case, when both subjects have an endowment of 5€, it seems there are some differences across the fair und the unfair treatment. But due to low destruction rates on general these differences are not significant (Wilcoxon rank sum test, p=0.36). The other pairwise comparisons of treatments as well as the case, when both subjects are endowed with 10€ also show no significant differences.

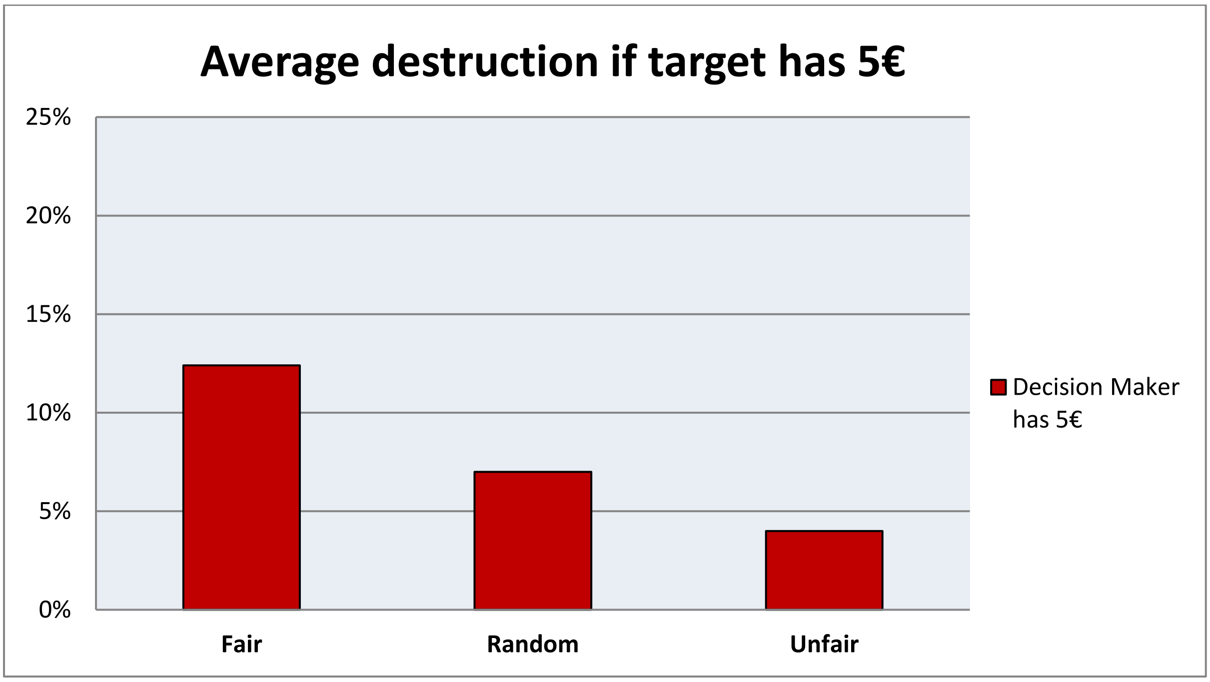

Figure 3

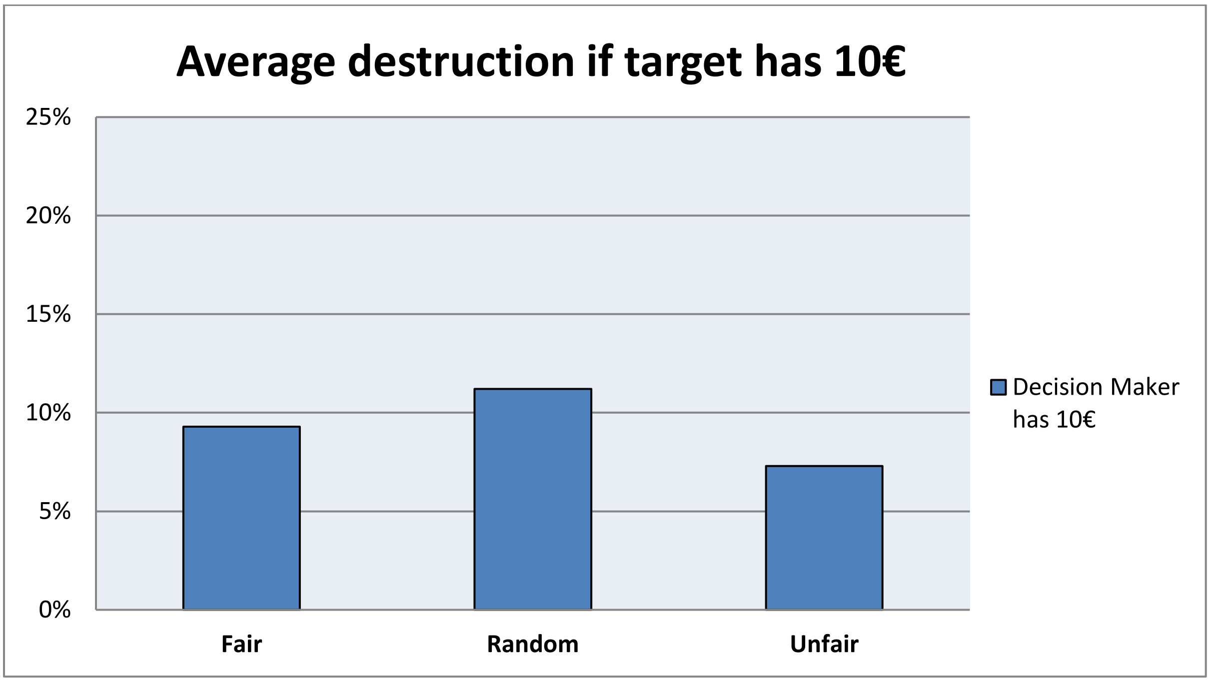

Figure 4

So, hypothesis 4 can mostly be confirmed.

Result 4: There are no significant differences across treatments concerning destruction rates for the combinations (low, low) and (high, high).

Case 4: Combination (high, low)

In the last case, most models and theories predict low destruction rates overall, as the decision maker is in a privileged position anyway. Results are displayed in figure 5. In contrast to those predictions, differences across treatments are noticeable in this case, especially between the random and the unfair treatment. They are close to being significant (Wilcoxon rank sum test, p=0.14). The evidence does not support hypothesis 5[8].

Result 5: There are no significant differences across treatments concerning destruction rates for the combination (high, low).

[8] As we will see in the regression in section 4.5, the interaction term of the unfair treatment and the income of the decision maker is close to being significant. So it seems there is the tendency that in the unfair treatment subjects with 10€ destroy less money from those having 5€ compared to the other treatments.

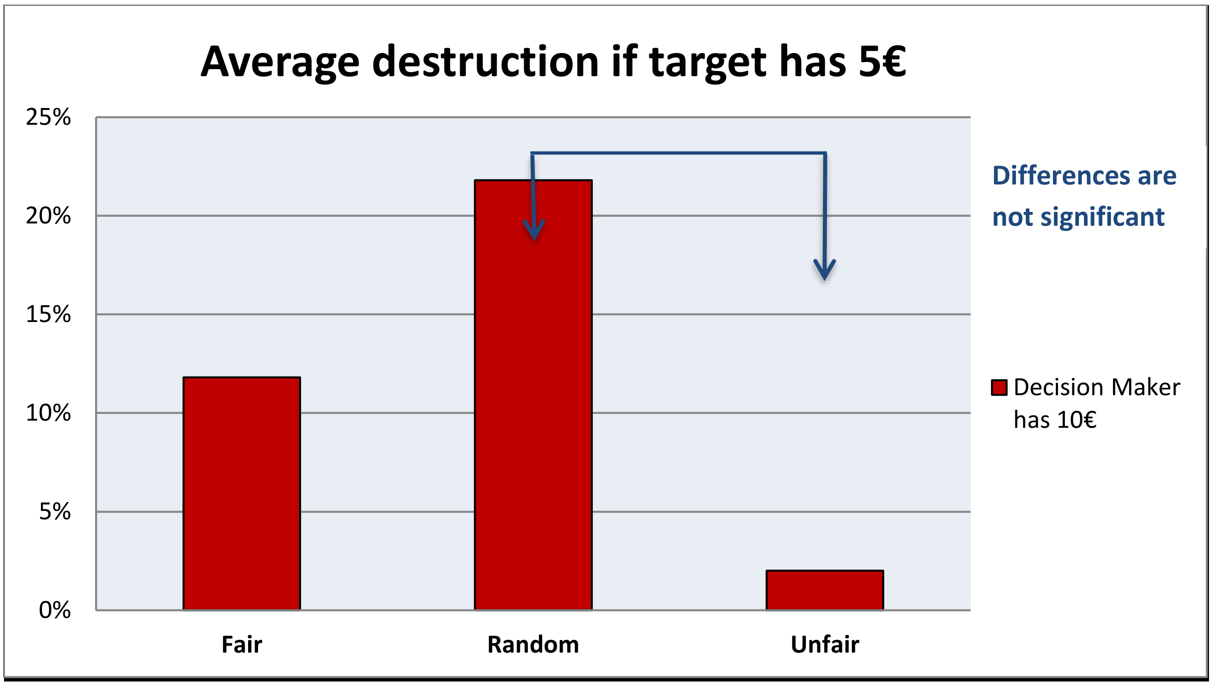

Figure 5

As mentioned before, the reason for most results being not significant are probably low destruction rates overall. To illustrate that point, we take a closer look at the data on an individual level for the last combination (high, low). In the random treatment in 5 out of 17 decisions, subjects decide to reduce somebody's payoff. Additionally, every time they chose a rather high amount and in 2 cases even the entire endowment of the group member is destroyed. In contrast, in the unfair treatment money is burned only in 3 out of 22 decisions and only to a very little extent in each case. That leads to high percentage differences on average, driven only by a few single decisions. Two of those subjects burning money in the random treatment state in the questionnaire as reason for their decision: *"They expected the other participant would have done the same, if he or she had been given the opportunity to do so."* Obviously the motive of preemptive retaliation still seems to be a relevant factor, even though the design allows only unilateral destruction decisions.

The only significant effect found is *within* the unfair treatment for pooled decisions, while examining which endowment class is more prone to be target of destruction. Pooling in this case means, we consider both the decisions made by subjects having an endowment of 5€ or 10€. The endowment of the target is held constant in this case. As one can see in figure 6, the percentage of endowment destroyed of those in the unfair treatment, who received an endowment of 10€ is much higher than the amount reduced from subjects having 5€. These differences are highly significant (Wilcoxon rank sum Test, p=0.016; two tailed) and would even be larger, if one compared absolute values instead of relative ones. As the diagram suggests, this effect is mainly driven by the fact that subjects with an endowment of 5€ (target has 5€) is less money subtracted in the unfair treatment compared to the other treatments. In the other treatments, differences are insignificant

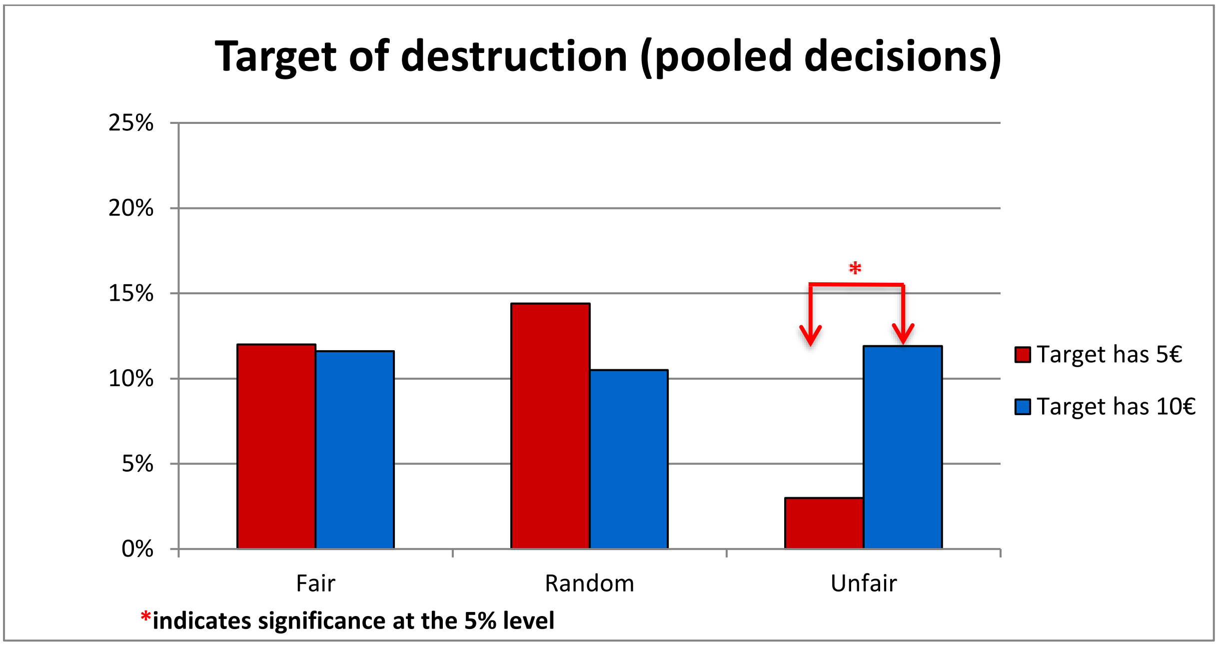

Figure 6

Result 6: <u>Within</u> the unfair treatment the percentage destruction targeting subjects having a high endowment is significantly higher than the percentage targeting those with a low endowment.

Besides counterweighting effects of different income classes, another reason for the similar burning rates across treatments could be that participants' burning decisions are not strongly affected by their fairness evaluations. In the next section we analyze the assessments of those subjects burning any positive amount of money and those who decide not to do so.

4.4 Correlation of fairness evaluation and destruction decision

As shown before, subjects rate the unfair treatment as clearly less fair than the other two treatments. Nevertheless, their judgments do not seem to (strongly) affect their destruction decisions. Looking at the correlation of one subject's burning decision (yes=1 or no=0) and her fairness evaluation of the treatment, there is almost no correlation at all (Spearman's rho=0.11; Test of independence p>>0.1). The figures are similar, if you control for specific income combinations. Therefore it seems, those subjects who make use of the opportunity to destroy another player's endowment, do not mainly act in this way, because they perceive the endowment determining mechanism as unfair.

Result 7: Negative fairness evaluations do not trigger subjects' decisions to reduce another player's payoff in this context.

4.5 Regression Analysis

In this section we additionally report the results of the regression analysis. This allows us to perform the analysis with demographics and other controls. Results are similar as before. Analogously to section 4.2, we first look at pooled values of destruction. We use the percentage of destruction as the dependent variable. An alternative specification would be a regression with the destruction decision as binary outcome. But as discussed, this analysis the magnitude of destruction and provides us with more detailed results.

Results of the pooled regression are displayed in table (X). None of the treatments have a significant effect on destruction rates. Male subjects and economists show more antisocial behavior, while people, who donate to charity destroy significantly less money.

Table 1: Regression: Pooled destruction (in percent)

	Treatments only	Treatments with controls
T_fair	-0.028	-2.465
	(4.864)	(4.827)
T_unfair	-2.815	-6.329
	(4.788)	(4.973)
Male		6.170
		(4.097)
Economics		7.747^{+}
		(3.935)
Charity		-8.134*
		(4.050)
Constant	11.784**	12.818*
	(3.596)	(5.296)
Observations	119	119
R^2	0.004	0.083

Standard errors in parentheses

Destruction is measured by adding up both intended values and dividing them by sum of endowments of the targets (=15)

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

In table (Y) we report the results by specific income combinations, both for the decision maker and the target. In each regression the income of the target is held constant. Both models are estimated with and without controls. The income of the decision maker is used as an explanatory dummy variable. We also look at interaction effects between the treatment variable and the income of the decision maker. The treatments have no significant effect on the destruction decision. When the target and the decision maker have an income of 5€, a higher percentage of the target's income is deducted than in the case when the decision maker has an income of 10€. This effect is weakly significant, but only in the random treatment. In the fair and unfair treatment this relationship is exactly reversed. There, less money is subtracted from subjects with 5€ by those with an income of 10€. As before, the interaction effect is not statistically significant. This might be the case, because of rather low destruction rates in general.

Table 1: Regression: Destruction behavior of specific income classes (in percent)

	Target has 5	Target has 10	Target has 5	Target has 10
T_fair=1	5.322	4.045	3.416	1.278
	(7.659)	(7.149)	(7.520)	(7.243)
Income_10=1	14.706^{+}	1.412	14.153^{+}	0.989
	(8.052)	(7.516)	(7.724)	(7.439)
T_fair=1 X Income_10=1	-15.287	-5.971	-16.701	-5.122
	(10.892)	(10.166)	(10.547)	(10.158)
T_unfair=1	-2.968	5.917	-6.732	2.107
	(7.581)	(7.076)	(7.452)	(7.178)
T_unfair=1 X Income_10=1	-16.797	-8.912	-15.451	-8.622
	(10.721)	(10.007)	(10.258)	(9.880)
Male			3.792	7.078
			(4.597)	(4.428)
Economics			9.554*	6.375
			(4.387)	(4.225)
Charity			-12.360**	-6.426
			(4.499)	(4.333)
Constant	7.059	9.765^{+}	11.626	9.975
	(5.694)	(5.314)	(7.224)	(6.958)
Observations	119	119	119	119
R^2	0.072	0.016	0.177	0.072

Standard errors in parentheses
In this regression the income of the target is held constant
and the income of the decision maker is used as dummy variable
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

5. Discussion and Conclusion

In this paper we analyze the extent of antisocial behavior in a money burning experiment, in which endowment is determined in different ways, varying in their degree of fairness. This is the first study, which focuses on the fairness of the money allocating process, while in most other comparable experiments, money is provided in the form of windfall gains.

Altogether, we find moderate degrees of antisocial behavior in all treatments. However, there are no significant differences in destruction rates across the treatments, even if one controls for specific income classes. One reason might be that individual burning decisions in this context do not to depend a lot on the perceived fairness of the treatment. This is shown by the results of the questionnaire. There is no correlation between the individual fairness assessment of the procedure and the decision to burn money.

Below we discuss several possible explanations for this result:

First, it could be that the fraction of people showing antisocial behavior is rather constant and independent of the context: Some evidence for this hypothesis comes from Sadrieh & Schröder (2012) and Le Zhang & Ortmann (2012). They both suggest, there is some relatively constant fraction of people exhibiting both pro- and antisocial behavior, depending on the exact choice set they are offered (e.g. Joy of Destruction Game or Dictator Game). Sadrieh and Schröder assume these so called “influencers” derive their utility mainly from having the power to change the payoff of others, no matter in which direction. This argument goes in a similar direction as experimenter demand or boredom effects. By these effects participants also gain utility from the process in which payoffs are generated and not only from the final distributional consequences (e.g. They want to be “active” or have “fun”, while taking part in the experiment). This is especially true for low stakes scenarios, as is most often the case in standard economic experiments. But there is also abundant evidence against this argument: For example in Zizzo & Oswald (2001) or in Fehr (2018) the amount of destruction varies a lot as treatment parameters change.

An alternative explanation is uncovered by looking into responses to our questionnaire: Several participants state as their reason for not reducing somebody else’s payoff, that they do not like the endowment allocation mechanism, but see it *not as the fault of the other subject*. Two aspects are important here: First, strictly speaking, the procedure of endowment determination is ex ante also random across all treatments: Subjects are randomly assigned to one of these treatments, as well as to the early or late group. Presumably, some participants perceive the whole mechanism as a sort of (unfair) “lottery”, in which some participants are lucky (the ones who do not have to do any work and receive the high endowment) and others are not. Adopting this view, one could argue that the criterion of process fairness is satisfied here, as all participants had ex ante the same chances being in each of the possible positions.

And secondly, subjects in our experiment have no opportunity to balance initial differences. Possibly, because of this they are not blamed or held responsible for these differences and therefore are not target of destruction more often. This would explain why destruction rates are at a low level in all treatments.

For both explanations we find clear evidence in other studies:

In Bolton et al (2005) “unfair” (in the sense of strongly unequal) outcomes in an Ultimatum Game are widely accepted, if they are determined by a fair (random) procedure, in which the players both had the same chance to receive the favorable outcome. Furthermore they find that unfair offers are only frequently rejected, if the proposer had the chance to choose a more equal allocation, but not if he had no other choice. In a similar fashion, Blount (1995) finds that much lower offers than common are accepted in the Ultimatum Game, if these offers are randomly generated by a computer instead by the participant himself. Fehr and Fischbacher (2004) also provide additional evidence for the second point: They examine second and third order punishment in distributional choices. In their experiment subjects play a Dictator Game. In one treatment the recipient himself can after seeing the decision costly punish the dictator. In another treatment a neutral third party has the right to do so. Punishment for low transfers is widespread and used both by the recipients and also neutral third parties. So, there is clear evidence, that if people violate common sharing norms, they get sanctioned. The main difference to our experiment is that people have no option to change initial endowments. Thus we can presume that destruction rates would have been much higher - especially in the unfair treatment - if subjects were offered an option to redistribute incomes.

So we can conclude that antisocial behavior in our experiment is probably so low, even in the unfair treatment, because subjects cannot be held responsible for the resulting (unfair) distribution.

The latter explanation could be tested by performing a similar experiment and adding an additional stage, in which subjects with the high endowment can transfer some share to the ones with the low endowment before the destruction decision is made. Or instead the design could be changed such that the high endowment is not directly allocated, but subjects are assigned the right to choose one of these values (that means in the unfair treatment participants from the late group would have the right to choose which endowment they would like to receive). Then we would expect a much stronger treatment effect on destruction rates targeting those subjects who did not redistribute.