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

This Version: November 11, 2021

Abstract

We examine whether an unfair process of income allocation leads to a higher degree of antisocial behavior. First, an initially unequal distribution is determined by either a fair, a random or an unfair process. In this context the fair payment rule induces an approx. equal effort-to-pay ratio, while in the unfair case this ratio is strongly imbalanced. Participants can then anonymously reduce the income of another player at a cost. Our main finding is that money burning rates are relatively low and similar across treatments. In contrast to the literature, even in the unfair treatment disadvantaged participants do not destroy more from those who are better off. Only in the reversed direction there is an effect: participants who suffer from the unfair mechanism are very rarely the target of destruction, irrespectively of the income class of the decision maker.

According to reported fairness evaluations, the unfair mechanism is perceived as clearly less fair then the other mechanisms. However, there is no correlation between fairness evaluations and the propensity to reduce money. This suggests that subjects' decisions are not affected by the (un-)fairness of the allocating process. Putting together evidence from a debriefing questionnaire and related studies suggests that the decision to burn money depends a lot on whether there is individual responsibility for the (unequal) distribution.

Keywords: Process Fairness, Antisocial Behavior, Money Burning

JEL-Codes: C91, D31, D63

1 Introduction

An unequal distribution of income and wealth is found in many countries (see e.g. the World Inequality Report 2018 by Alvaredo et al.). While a certain degree of inequality may be inevitable and could even be beneficial for society as a whole, a large degree may create dissatisfaction and social conflict.[1] Prominent theories of justice suggest point at the *the fairness of the allocating process* being a key factor for the tolerance of such inequalities (see e.g. Konow, 2001, 2003; Cappelen et al., 2007). This suggests that individuals' fairness perceptions about the income generating process significantly affect their attitudes towards the resulting distribution. If they view the income generating process as fair, they might accept some degree of inequality. Whereas if the process is perceived as being unfair, individuals might have negative emotions/become frustrated. According to a prominent theory from social psychology ("frustration–aggression hypothesis", see e.g. Berkowitz, 1989) these negative emotions may lead to more aggressive behavior. In our framework, individuals are offered the chance to become "aggressive" by reducing other persons' payoffs. The aim of our experimental study is to examine the latter aspect and how it relates to individual fairness perceptions.

In our context, *fairness* refers to the relation of effort participants have to exert in a task and their respective payment. All payments in our experiment are performance-independent, but the amount of effort (working time) differs substantially across conditions. Under the fair payment-rule this relation is balanced throughout all participants/groups, while in the unfair case it is strongly imbalanced. Applied to real-life contexts, the fair treatment of our study reflects the kind of compensation schemes which are very common in labor contracts: Employees receive a fixed payment that is proportional to the amount of working time they have to provide. Regarding the unfair condition, one may think of remuneration schemes where in some cases those persons who spend less time or effort receive a higher compensation (examples could be *mistakes* due to imperfect monitoring/measuring of performance, *discrimination or corruption*: bonuses are paid not according to effort/performance but according to certain characteristics out of control of the employee, e.g. common membership in some interest group or institution, unfair privileges granted by third-parties only to some persons etc.).

If the individual frustration about the unfair payment scheme is high enough, it may lead to harmful behavior towards others (see for example several studies which find a negative effect resulting from an unfavorable mechanism/outcome on behavior, like for example sabotaging others: Ambrose et al., 2002, increased violence against other persons: Munyo & Rossi, 2013, less acceptance of inequality: Grimalda et al., 2016 and more money burning: Fehr, 2018 and Grosch & Rau, 2020).

We use the term *antisocial/harmful behavior* for participants destroying each

[1] The relationship between inequality and subjective well-being is relatively complex, because in many cases inequality is also correlated with other factors, which can have with a negative impact on well-being (as for example unemployment). Therefore, there is no conclusive evidence about the exact relationship (see Schneider, 2015 for an overview of findings of different studies). However several studies report that especially members of low income groups suffer from an unequal distribution, see Alesina et al., 2004 and Oishi et al., 2011

other's income at a cost with no material benefit.[2] One of the underlying motivations could be to reduce the degree of (perceived) inequality. While in our experiment the degree of inequality is always constant from a monetary perspective, the effort-to-pay ratio differs substantially.[3] Following from concepts from social psychology such as the "Equity theory" by Adams (1963) and insights from theories of justice, individuals tend to accept income differences as long as the proportion of effort to payoff is similar across participants. Instead, if this proportion is perceived to be unbalanced, participants tend to react negatively. Therefore, we would expect disadvantaged individuals, especially those who find the income allocation unfair, to be more likely to reduce the income of others. In a similar fashion, models of inequality aversion (as for example Frohlich et al. (2004)) can also predict such behavior, if effort is interpreted as a form of monetary cost.

In the real world it is often not easy to measure the true extent of antisocial preferences, either because the revealing of such preferences is not possible or related actions are associated with potentially high costs of punishment. Nevertheless, it might be the case that several individuals intrinsically have such preferences, especially if they feel they have been treated unfairly. For this reason we examine this research question by the means of a lab experiment.

In our study, participants are given the possibility to destroy some or all of another person's income. A lot of so-called "money-burning" experiments have demonstrated that participants exhibit a moderate or even large degree of antisocial behavior (see for example Zizzo & Oswald, 2001; Abbink & Sadrieh, 2009; and Abbink et al., 2011).[4] The amount destroyed varies a lot and depends on the exact framework under which participants are acting (frequencies of money burning range from less than 10% up to 70%). Several studies investigate the relationship between inequality and money burning. A common finding is that participants with higher incomes are more likely to be the target of destruction (see Cetre, Lobeck, Senik, & Verdier, 2019; Zizzo, 2003; Dawes et al. (2007); Grossman, Komai, et al., 2013 and Fehr, 2018).

In the experiments discussed so far, income differences are predominantly due to chance (except the study of Fehr, 2018). Other merit-based factors are not relevant for the determination of the distribution. However, as the literature suggests, these factors play a crucial role for the tolerance of unequal distributions: Akbaş et al. (2019) analyze which main factors influence the fairness assessment of a distribution and how these factors effect redistributional choices. The authors find that in cases of either unequal opportunities or a lack of agency (possibility to influence the outcome by own decisions) there is higher desire for redistribution. Further evidence is provided by several studies where participants in some way could "earn" their initial endowments: For example Barr et al. (2015), Frohlich et al. (2004), List (2007) and Oxoby and Spraggon (2008) all find that recipients in Dictator Games are offered better allocations, if they positively contributed to the creation of endowment by their own effort or performance in a prior task. In a

[2]The terms harmful and antisocial behavior are used synonymously in this paper.

[3]By the term "effort-to-pay ratio" we refer to the ratio of effort costs E translated into a monetary value divided by the related monetary payoff: $\frac{E}{P}$

[4]Following the literature, we interchangeably use the terms income *reduction*, income *destruction* and *money burning*.

similar fashion Hoffman et al. (1994) find an entitlement effect in the Ultimatum Game, if the right to be the proposer was determined beforehand according to one's performance in a task.[5] All these results suggest that participant's feeling of entitlement towards their income influences their willingness to accept the initial allocation. In the framework of our study, the degree of entitlement is closely related to the fairness of the income allocating process.

Some recent experimental work that studies antisocial behavior incorporates fairness aspects into the income allocating process. Fehr (2018) examines if increasing inequality leads to a higher degree of harmful behavior towards other individuals. One of the main results is that increasing inequality does indeed lead to more money burning, but only if the underlying income allocating process is unfair. In contrast, if higher inequality can unambiguously be attributed to higher effort, participants do not destroy more. An important difference between our study and the one of Fehr is that income differences are not linked to (unethical) behavior of other persons. In his study, destruction observed under the unfair payment regime could be interpreted as a sort of punishment towards individuals who might have increased their income by unfair means. Furthermore, in our study we can disentangle the effect based on fairness perceptions of the process from the effect based on different degrees of inequality.

The closest related study to ours is Grosch and Rau (2020). In their experiment, the authors use different reward schemes for a real-effort task: performance-based, random and an exclusion from payment called "Discrimination". The latter can be viewed as an unfair lottery, taking away a subject's performance-based reward after completing a task. This procedure is likely to be highly frustrating, especially for those who expect to be awarded one of the bonuses. Afterwards, participants play a variant of the "joy of destruction game" (first employed by Abbink & Sadrieh, 2009). On average, the authors observe moderate overall destruction levels. However, under the unfair payment regime the group of non-earners exhibits significantly more antisocial behavior, compared to both the other treatments and to the respective treatment in our study. In contrast to their design, payments in our experiment are not linked to performance. Probably, this increases the entitlement of participants to their payments and hence their frustration levels. Eventually this might trigger harmful behavior to a larger extent. Furthermore, in their study monetary and moral costs for money burning are lower compared to ours. Plausibly, this further increases the effect just described.

Our study contributes to the existing literature in multiple ways: First, we disentangle effects resulting from the inequality in outcomes from those driven by the (un-)fairness of the underlying mechanism. Often these concepts confound each other. Furthermore, we implement a payment scheme, which does not depend on individual decisions or performance (a very common form of labor contract in real-world applications). It is an important question if inequalities arising in such environments are widely accepted or if they trigger envy and harmful behavior against co-workers. Income differences under this regime may have significantly

[5]Demiral and Mollerstrom (2020) do not find this effect. However, in their study all participants had to perform a real-effort task, whereas in our study some participants do not have to undergo the task. This makes it likely that in our study participants who have to do the task feel more entitled to their payment.

different impacts compared to a situation where workers have a strong feeling of entitlement to their payments. Another important feature of our design is that we measure participants' fairness evaluations of the income allocation process. This yields information on how they perceive the respective mechanism, allowing us to directly link participants' evaluations to their behavior. This provides a more detailed picture about the relationship between the perceived fairness and burning behavior. Finally, participants' self-reported motives help to obtain further information about the key factors relevant for the decision to destroy money.

2 Experimental Design

The experiment consists of two main parts: The income determination phase and the destruction decision. As initial income (endowment), half of the participants receive a high amount (10 Euro) and the other half a low amount of money (5 Euro). There are three treatments, which vary the degree of fairness by using different income allocating rules. This induces different effort-to-pay ratios:

1. **Treatment "Fair"** (equal effort-to-pay ratio)
2. **Treatment "Random"** (unequal effort-to-pay ratio)
3. **Treatment "Unfair"** (highly unequal effort-to-pay ratio)

In treatment *Random*, initial income levels are determined by a lottery. Participants pick up a sealed envelope from a box. This envelope contains a note informing them that they either receive 5 Euro or 10 Euro. Participants are additionally told that half of the participants are allocated the 10 Euro, while the other half receives 5 Euro. This treatment is intended as a "baseline", to be comparable with other money burning experiments where endowment is provided in the form of windfall gains.

In the other two treatments, *Fair* and *Unfair*, the subject pool is divided into two groups. Members of the first group have to arrive 30min before the other group. During this period, they perform a real effort task, involving marking IQ-tests from another unrelated experiment. The participants in the late group do not have to do any work. In the fair treatment, all members of the working group receive an initial income of 10 Euro and participants from the late group receive 5 Euro. The difference in payment values is chosen such that it approximately reflects the difference in effort (time spend during the experiment), using a compensation scheme common at that time. As a result, in the fair treatment the effort-to-pay ratio is supposed to be approx. equal for both groups. The payment scheme in the unfair treatment is exactly reversed: Participants performing the effort-task receive 5 Euro and the participants from the late group are given the higher income of 10 Euro. The general procedure is common knowledge to all participants within a specific treatment. However, they are not informed about the income allocation processes in the other treatments.

After the first phase of income allocation, each participant is given the opportunity to destroy some or all money of a randomly selected other participant. This part of the experiment is identical across all treatments. The costs of burning

money are 10% of the selected amount. Every participant makes two burning decisions, using the strategy method.[6] They indicate how much to reduce in case the other participant has an income of 5 Euro and in case he or she has an income of 10 Euro. Participants can choose any value ranging from zero up to the total income of the other person (either 5 Euro or 10 Euro). Framing was as neutral as possible to avoid any type of demand effects.[7] Once the burning decisions are recorded, participants are randomly paired in groups of two. In each of these groups, only one burning decision is actually carried out ("unilateral destruction"). The intention behind this is to minimize motives like preemptive retaliation or negative reciprocity.[8] At the end of the session, participants fill out a debriefing questionnaire for which they receive an additional compensation.[9] Participants are asked about their fairness rating of the money allocating process and their motives for their burning decision. Finally, a random process determines which destruction decisions will be carried out and participants are informed about their final earnings.

Implementation Details

The experiment was pen & paper based and was conducted in 2013 in the experimental lab of the University of Heidelberg. The average duration per session was about 45-60min (including the real-effort task). Overall, 119 participants, mostly students, took part. The sample size was determined before any data analysis. Anonymity with respect to both the other participants and the experimenters was ensured by the use of a personal identity code. Average earnings were around 10 Euro per person, with payoffs ranging from 3-13 Euro. Details about participant and session numbers per treatment are summarized in table 1. Each subject made two burning decisions, therefore the number of observations is two times the number of participants in the respective treatment. In the results section, we report all relevant measures and manipulations.

Table 1: Summary of participant numbers

Treatment	**Income Allocation Process**	**Sessions**	**Participants**	**Observations**
Fair	Effort-based	3	41	82
Random	Lottery	3	34	68
Unfair	Effort-based (reversed)	3	44	88
Total		9	119	238

[6]This may have some influence on the amount of destruction as both decisions are now set in relation to each other. However, as we are mainly interested in differences of burning rates across treatments, there seems to be no obvious reason why this should effect treatments differently.

[7]Furthermore, in the instructions it was pointed out that one does not necessarily have to subtract any money.

[8]If both decisions were implemented, it could be possible that some players engage in money burning only because they have the belief the others will do so and therefore would want to retaliate.

[9]Participants had to answer all questions for being eligible for the extra payment.

3 Hypotheses

As explained before, the treatments impose different effort-to-pay ratios (equal, unequal and highly unequal). According to theories of social justice, inequality aversion as for example Adams (1963), Fehr & Schmidt (1999) and Frohlich et al. (2004) and the frustration–aggression hypothesis (e.g. Miller, 1941; Berkowitz, 1989), participants might take measures to reduce the imbalance in the effort-to-pay ratio. The only option participants have for doing so is to destroy some income of another person. If this makes sense from an individual's perspective depends on how much utility they receive from a reduction of inequality vs. the costs of money burning.

Applying this logic, these models predict destruction to stem predominantly from disadvantaged participants targeting the class of subjects with a higher income (pairing 5, 10). According to Adams (1963) and Frohlich et al. (2004) the amount of destruction is expected to be highest in the unfair treatment and lowest in the fair treatment because the effort-to-pay ratio is most imbalanced in treatment *Unfair* (and least imbalanced in treatment *Fair*) This causes the highest degree of frustration for participants with a low income in treatment Unfair.[10] The same prediction can be derived from the model of Fehr & Schmidt (1999), if one interprets effort as some form of monetary costs. Otherwise this model would predict no treatment differences in destruction for any specific income pairing (as the difference in incomes is the same throughout all treatments for a given pairing). Furthermore, results from studies examining entitlement effects in real-efforts tasks suggest that participants with 10 Euros have the strongest entitlement to their income in the fair treatment and the weakest in the unfair treatment. As discussed in the introduction, this may lead to higher frustration and eventually money burning levels if the payment scheme is perceived as unfair. Following these reasonings the first hypothesis is:

Hypothesis 1: When the decision maker has a low income and the target has a high income (pairing 5, 10), destruction rates are expected to be highest in the unfair treatment and lowest in the fair treatment.

Concerning the reversed case, a decision maker with a high income targeting a person with a low income, most models would not predict any destruction at all, as this would increase the inequality in the effort-to-pay ratio even further. The same logic applies taking into account entitlements. However, as findings of other studies suggest, there might be a certain baseline level of destruction for reasons we will discuss below. In line with the argumentation from before, this baseline level of destruction might be somewhat reduced in the unfair treatment and increased in the fair treatment.[11]

Hypothesis 2: When the decision maker has a high income and the target has a low income (pairing 10, 5), destruction rates are lowest in the unfair treatment and highest in the fair treatment.

[10] In Frohlich et al., 2004 this is referred to the concepts of "just deserts".

[11] In the fair treatment a decision maker with an income of 10 might even feel slightly worse off compared to a participant receiving 5 Euro, but who did not have to do any work.

For the cases when the decision maker and the target have the same income, none of the models/theories would predict any treatment differences, as there is no difference in the relative position between participants (both are in the same privileged or unprivileged position). Nevertheless, there might be a certain baseline level of destruction for reasons such as "joy of destruction" (Abbink & Sadrieh, 2009), experimenter-demand effects or a preference for "being active". Some experimental studies even observe higher burning rates in pairings with equal/similar income values (Abbink et al., 2011), while others find an opposite effect (Dawes et al., 2007). Furthermore, it could be that the increased frustration about the payment mechanism triggers negative emotions in general. For some participants this could increase the preference of burning money independent of the income pairing (see various studies which find such negative effects on behavior, like for example sabotaging others: Ambrose et al., 2002, increased violence against other persons: Munyo & Rossi, 2013, less acceptance of inequality: Grimalda et al., 2016 and more money burning: Fehr, 2018 and Grosch & Rau, 2020). Given this evidence, the prediction for equal income pairings is not clear.

To be in line with the models' argumentation, we maintain the hypothesis that there are no treatment differences for these pairings.

Hypothesis 3: When the decision maker and the target have the same income (pairings 5, 5 and 10, 10) destruction rates are expected to be the same across all treatments.

Given the models prediction that destruction behavior results from imbalances in the effort-to-pay ratio and increased frustration about the payment scheme, this would lead to certain treatment differences on the aggregate level. One of the main expected drivers for these differences is the behavior of decision makers with a low income destroying some money from those with a high income (see hypothesis 1). From this the last hypothesis results:

Hypothesis 4: Aggregate destruction is lowest in the fair treatment and highest in the unfair treatment.

4 Results

4.1 Fairness Evaluations

At first, we examine if the treatments have an effect on participants' fairness perceptions of the compensation scheme. All participants are asked to evaluate how fair they find the income determining mechanism they experience. They can rate the process on a scale ranging from 1-5, where 5 means they perceived the mechanism as "very fair", 3 corresponds to "neutral" and 1 to "very unfair". Ratings took place before knowing the results from the money burning stage in order not to be influenced by those. Average fairness evaluations are shown in table 2.

Table 2: Average Fairness evaluations

Treatment	**Obs.**	**Fairness rating**	**Rating by I=5**	**Rating by I=10**
Fair	41	3.6	3.6	3.6
Random	34	3.6	4.0*	3.2*
Unfair	44	1.7**	1.5	1.8

Fairness ratings on a scale from 1 (very unfair) to 5 (very fair)
** differences w.r.t both other treatments are significant at the 1% level
* differences between income classes are significant at the 5% level

As expected, participants see the unfair treatment as clearly less fair than the other two treatments. According to a Wilcoxon rank-sum test the difference of treatment *Unfair* compared to each of the other treatments is highly significant (both times $p < 0.001$). There is no difference between treatments *Fair* and *Random*. It seems that subjects equally like a merit-based compensation scheme as in Treatment *Fair*) and a scheme, which is not merit-based but grants an equal expected payoff (lottery). Furthermore, it could be that if they would have to rank all three mechanism simultaneously, there would be some differences between the fair and the random treatment.

If one splits fairness ratings by income class, in treatment *Fair* average ratings are exactly the same. This provides some evidence that -as intended- participants perceive the additional reward of 5 Euro as approx. equally valuable as the extra effort they have to provide. A bit surprising may seem that in treatment *Random* the ones who are assigned a higher income by the lottery rank this treatment as less fair than those who receive the lower income. In any case the random mechanism is evaluated as relatively fair by both income groups. In contrast, the unfair mechanism receives a very low fairness rating by any income class. There is a tendency of the disadvantaged group perceiving it as even less fair. As we expect most destruction stemming from exactly those participants, the manipulation worked in inducing the respective differences in fairness perceptions across treatments.

4.2 Aggregate Destruction Behavior

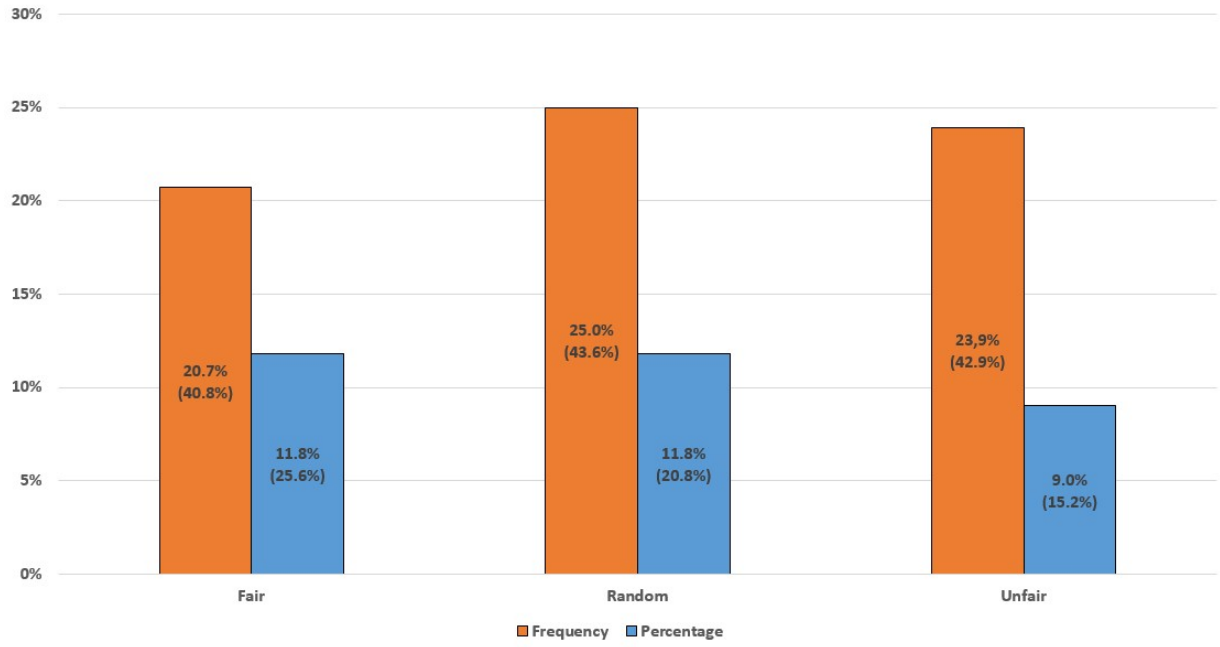
In our experiment we measure destruction behavior in two ways, either as *burning frequency*, or as *percentage of destruction*. The burning frequency measures the number of times a participant chooses to reduce the payoff of another participant by any positive amount $D > 0$. The percentage of destruction corresponds to the amount of money intended to burn divided by the income of the other participant (either 5 Euro or 10 Euro): $\frac{D}{I}$. When measuring aggregate percentage values, we sum up participants' individual decisions D_1 and D_2 and divide by the sum of the targets' income values (I_1+I_2=15 Euro): $\frac{D_1+D_2}{15}$. This yields an average percentage value per person, which is not influenced by the exact allocation of the two decisions

to the income classes of the targets.[12] For the analysis one could alternatively use absolute values D_1 and D_2 per person. With respect to the results of the statistical tests, this yields the identical results as using percentage values.

Results of the average burning frequency and percentage of destruction per treatment are presented in figure 1. The average burning frequency ranges from 20% to 25% and the average percentage of destruction is around 10% in all treatments. Overall destruction levels are moderate and noticeably lower compared to other money burning experiments. This is especially true with respect to the unfair treatment.[13] There are no significant differences across treatments for destruction rates (Kruskal-Wallis test, $p = 0.74$). A pairwise comparison of treatments by a t-test indicates the same result. Therefore hypothesis 4 is not supported by the data.

The finding of no difference across the three treatments may be surprising. We will discuss potential explanations for this observation later in more detail. One possibility could be that there are multiple effects at play, which counterbalance each other to some extent. This is something we will have a closer look at through the destructive behavior in specific income pairings in the next subsection.

Figure 1: Aggregate destruction frequency and percentage of destruction



Bars report values of the mean and standard deviation (in parenthesis)

[12]Otherwise, it would matter if a decision maker subtracts 1 Euro from an income of 5 or 10 Euros.

[13]For example, in the experiment of Zizzo and Oswald (2001) the burning frequency is 70% for the same cost parameter of 10%. In Fehr (2018) in the treatment with a low degree of fairness ("Bonus & Cheating") the overall burning rate is 42%.

4.3 Destruction Behavior by Income Pairings

There are four different possible pairings w.r.t. income of the decision maker and the target (the respective number of observations are listed in brackets):

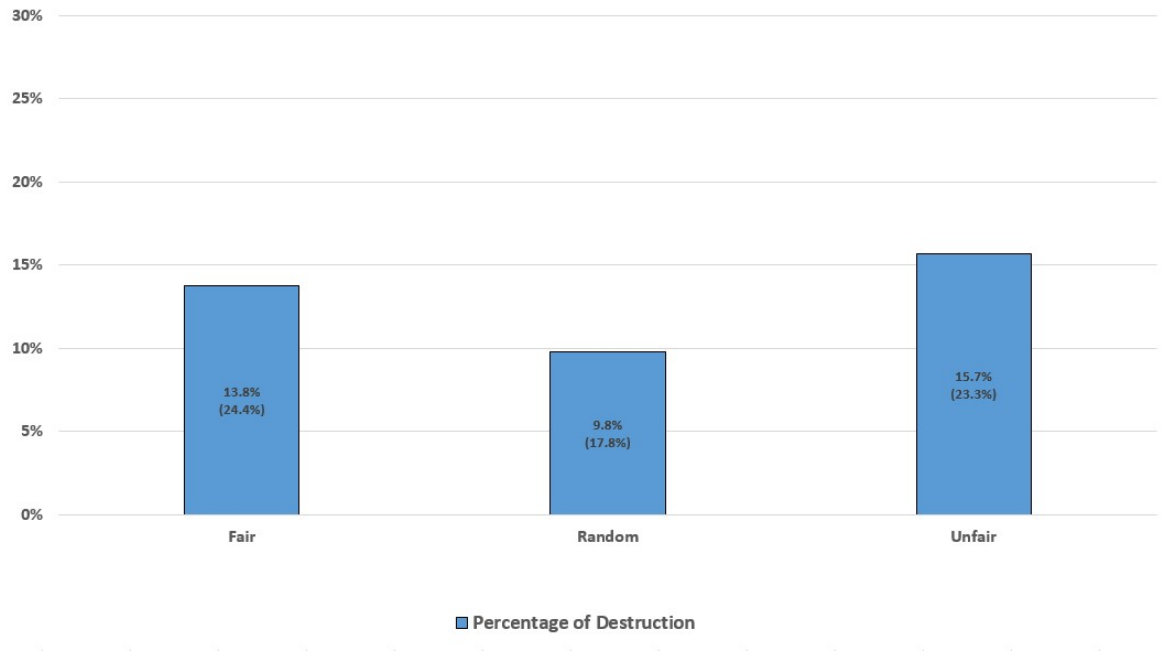
1. **Pairing (5, 5)** = Decision Maker: 5 Euro, Target: 5 Euro (n = 60)
2. **Pairing (5, 10)** = Decision Maker: 5 Euro, Target: 10 Euro (n = 60)
3. **Pairing (10, 5)** = Decision Maker: 10 Euro, Target: 5 Euro (n = 59)
4. **Pairing (10, 10)** = Decision Maker: 10 Euro, Target: 10 Euro (n = 59)

As discussed before, differences in destruction behavior are expected to be predominantly found in pairings with unequal incomes. For this reason, we focus specifically on those cases.

Disadvantaged Decision Maker: Pairing (5, 10)

According to hypothesis 1, we expected to observe most destruction for this paring, with rates being highest in the unfair treatment. Figure 2 displays the amounts destroyed per treatment. In the analysis, we focus on the percentage of destruction, as this contains more detailed information than the burning frequency.

Figure 2: Destruction behavior of disadvantaged group (5 vs. 10)



Bars report values of the mean and standard deviation (in parenthesis)

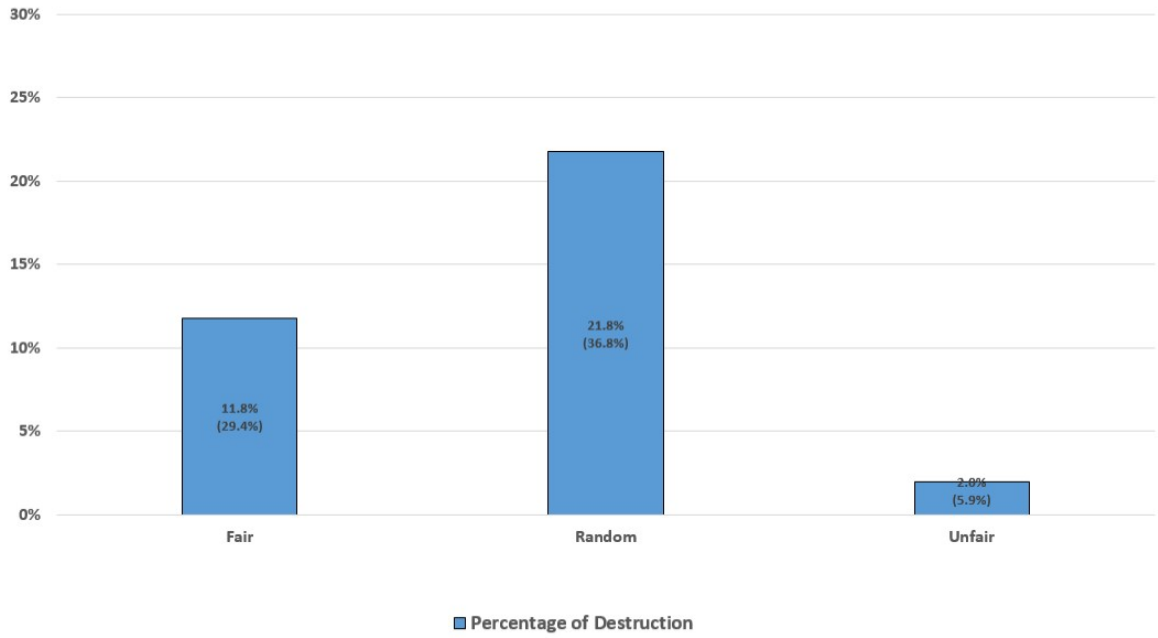
Compared to aggregate values, destruction levels tend to be slightly higher for this pairing. However, there are still no significant differences between the fair and the unfair treatment (two-sided t-test, $p = 0.80$). The same is true for a comparison

between treatments Random and Unfair (two-sided t-test, $p = 0.39$). The results do not support hypothesis 1. We will discuss possible explanations for this result in the next sections.

Advantaged Decision Maker: Pairing (10, 5)

Figure 3 presents the percentage of destruction by treatment when the decision maker has a high income and the target a low one.

Figure 3: Destruction behavior of advantaged group (10 vs. 5)



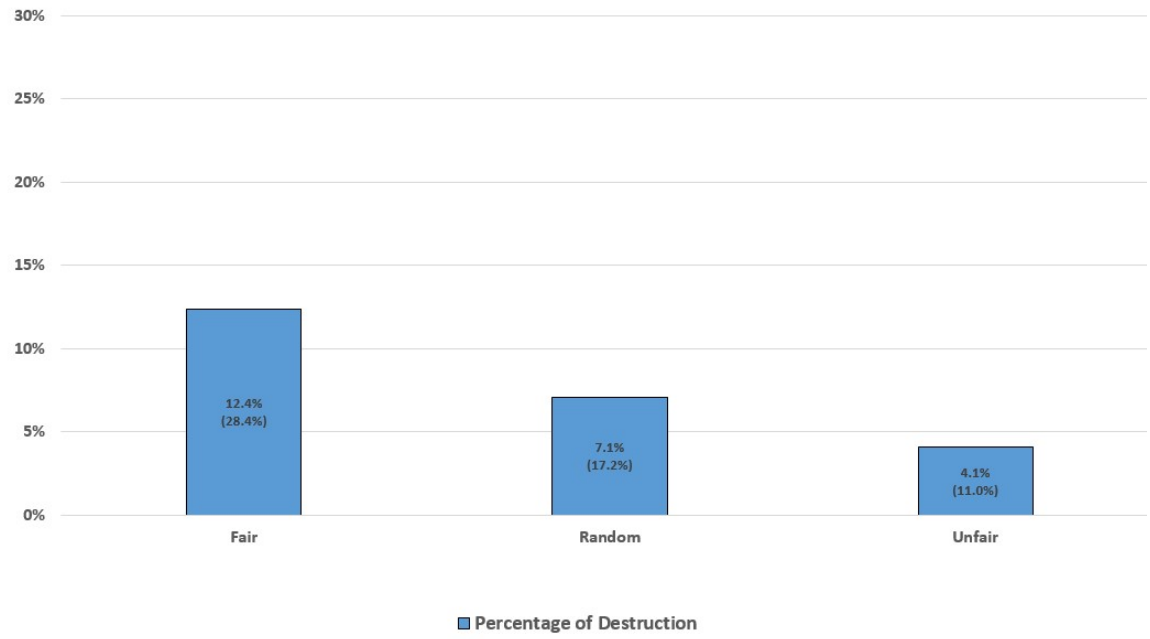
Bars report values of the mean and standard deviation (in parenthesis)

Clearly, less money is reduced in the unfair treatment. Differences in the percentage of destruction between the random and the unfair treatment are significant (two-sided t-test, $p < 0.02$) and between the fair and unfair treatment the tendency is the same (two-sided t-test, $p = 0.13$). Maybe somewhat surprising is that money burning is the highest in the random treatment. Looking at the burning decisions in detail reveals that the high percentage value in treatment *Random* is driven by only a few observations. In these cases, the entire income of the target is destroyed, which is unusual. Additionally, as reported before the income allocation processes in the fair and the random treatment are perceived as equally fair (see section 4.1). This would imply that the main differences are expected to arise in the comparison of the unfair treatment with any of the other two treatments. Overall hypothesis 2 partly can be confirmed.

Equal income pairings: (5, 5) and (10, 10)

For the parings with equal income levels (5, 5) and (10, 10) we did expect some baseline level of destruction but no strong treatment differences. Results are more or less in line with expectations: There are no significant treatment differences for any pairing. Therefore, hypothesis 3 can be confirmed.

Figure 4: Destruction behavior of pairing (5 vs. 5)



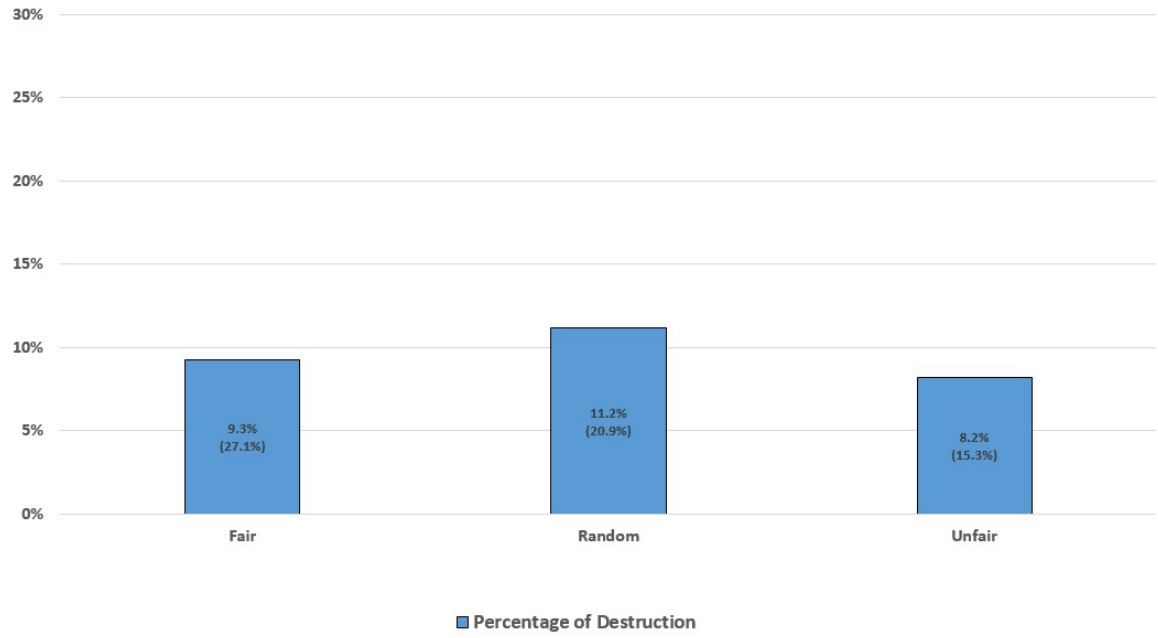
Bars report values of the mean and standard deviation (in parenthesis)

The biggest effect is found for the pairing (5, 5) when comparing the fair and the unfair treatment (two-sided t-test, $p = 0.21$). Especially for the pairing (5, 5) destruction rates are noticeably lower in the unfair treatment.[14] A potential reason could be that those participants indeed are dissatisfied with their payment but do not want to harm someone in this poor position even further.

It seems that the possibility of increased frustration caused by treatment unfair does not lead to higher burning rates on the aggregate level, as observed in the related studies of Fehr (2018) and Grosch and Rau (2020). A possible explanation we will discuss in more detail below could be that payments are not directly linked to performance.

[14]There might be a concern that this is caused by a demand-effect due to the simultaneous elicitation of two destruction decisions. However, applying this logic in the reversed direction, one could also argue that such a demand-effect would increase harmful behavior against those with a higher income when destruction against targets with 5 Euro is set as a reference point.

Figure 5: Destruction behavior of pairing (10 vs. 10)



Bars report values of the mean and standard deviation (in parenthesis)

Sample sizes and power of the statistical tests

As sample sizes of individual categories of pairings are not very high, one may question if the corresponding power of the statistical tests is high enough to detect possible treatment differences. When determining sample sizes for this study, we orientated by results of previous money burning studies at that point in time. There, one could observe relatively high destruction frequencies (e.g. approx. 70% for the very same cost parameter in the study of Zizzo & Oswald, 2001 and ca. 40% in the hidden treatment of Abbink & Sadrieh, 2009). Therefore, the assumption of treatment differences in burning rates of 20-30% does not seem unreasonably high. Given the observed effect sizes in more recent related studies[15], a treatment difference in destruction rates of approx. 20%, an a-priori power analysis would indicate a need of $n = 26$ observations per treatment group.[16] This need is only slightly above the number of observations we have in each category of income pairings (ca. 20 observations) and clearly below the numbers in the analysis of aggregate burning rates. Therefore, we attribute differences of our results in comparison to findings of other studies to differences in the decision framework. We will discuss the latter point in more detail in the next section and see our findings shedding further light on the complex interplay of fairness aspects and harmful behavior.

[15]In Fehr (2018) overall treatment differences in burning rates between the treatments "No Bonus" and "Bonus & Cheating" are 22%. In Grosch and Rau (2020) treatment differences in the percentages of destruction of persons with a low income ("non-earners") are 15% when comparing treatments "Discrimination" and "Random".

[16]using a value for the standard deviation of $sigma = 0.25$, which is between the observed value from this study and that in Grosch and Rau (2020) and values for $\alpha = 0.05$ and $\beta = 0.8$, which are common in experimental economics.

4.4 Fairness evaluations and burning decisions

At last, we have a closer look on how individual fairness perceptions influence the decision to burn any money. As we have seen, participants evaluate the unfair treatment as clearly less fair then the other two treatments. However, their judgments do not seem to have a strong impact on burning decisions. There is, in fact, no correlation between burning decisions ($yes = 1$ or $no = 0$) and fairness evaluations ($Spearman's\ rho = -0.06$; Test of independence $p = 0.36$, $n = 238$). Even when focusing only on the parings with a disadvantaged decision maker and a high-income target (5, 10), the correlation remains low ($Spearman's\ rho = -0.21$; Test of independence $p = 0.11$, $n = 60$). In this specific pairing, one would expect the strongest effect on destruction behavior. This analysis seems to suggest that the perceived fairness of the income allocating process is not a main driver of a participant's decision to burn money.

5 Discussion

In this study, we investigate if the fairness of the money allocation process has an impact on antisocial behavior in terms of money burning. Overall, we find relatively low levels of destruction, which are similar across treatments. Even under an unfair process, those who are disadvantaged do not harm those who are better-off significantly more. This is true, despite these participants reporting that they are clearly dissatisfied with the allocating process. On the reversed direction, in the unfair treatment clearly less money is reduced from disadvantaged subjects, independent of the income class of the decision maker. Probably, most participants take into account that those individuals are already in a bad position and do not want to harm them even further. Overall these patterns do not lead to any treatment differences on the aggregate level. Aggregate destruction levels are even slightly lower in the unfair treatment.

A plausible hypothesis was that the opportunity of money burning would predominantly be used by participants with a low income targeting those who have a high one. However, when comparing behavior with results from related studies, destruction levels for this specific income pairing are noticeably lower in our experiment. This is especially true with respect to the unfair treatment:

In Fehr (2018), in the treatment with the lowest degree of fairness (Bonus & Cheating) participants who are not in the top earning position[17] exhibit burning rates between 45% -50%. In this treatment there is a possibility to increase one's own performance by unfair means. Hence, the relationship between effort and income is not fully transparent and possibly distorted. Probably those receiving an income which is lower than expected have a strong belief that the other participants increased their performance by unfair means. Money burning in this case could be interpreted as both a sort of punishment for the (unfair) behavior of others and a way to channel negative emotions/frustration.[18]

[17] This refers to participants who do not have the top income rank in groups-of-four.

[18] The author elicited emotions and satisfaction levels of participants before their burning deci-

In Grosch and Rau (2020), the percentage of destruction from disadvantaged participants equals 30% in their "Discrimination" treatment (the most close compensation regime to our unfair treatment). This is roughly twice as much as in the respective value in our data. To some extent this could be attributed to the fact that money burning entails zero cost in their study. However, this probably does not explain all the differences in burning behavior. In their framework, earnings are closely related to the performance in a real-effort task. It is well-established that such performance-based tasks create noticeable feelings of entitlement (see e.g. Hoffman et al., 1994 and Karagözoğlu & Riedl, 2015). In contrast of being rewarded according to one's performance, participants are "suddenly" excluded from the possibility of receiving a payment. This random kind of discrimination may cause high levels of frustration, which might increase the willingness to harm others.

While there are differences compared to the studies discussed above, some experimental evidence is more in line with ours: In the study of Jauernig, Uhl, and Luetge (2016) participants can "punish" each other (destroy income) after being exposed to a competitive task. Similar to our results, the authors find that winners from the task are not harmed by losers more often than by other winners and lowest destruction rates are observed among loser pairs. However in their study, the role of winner and loser had no directly payoff-relevant consequences thus creating no economic inequality.[19] Therefore, it remains unclear to what extent results are comparable to ours.

A more general explanation why in a framework like the one of our study overall destruction rates are relatively low could be that participants cannot not actively influence the initial allocation. In this sense, our results complement insights from previous studies dealing with the acceptance or tolerance of unequal allocations. The literature often has found that individual responsibility for the outcome plays a crucial role for the acceptance of an unequal distribution. Typically, this responsibility involves the possibility to influence the allocation through own effort and decisions or by redistributing resources (see e.g. Akbaş et al., 2019). By the assignment to one of the treatments, subjects are basically forced to provide more effort/time under some conditions. This is an important difference to setups where subjects' payments are based on their own decisions or performance.

In line with the argumentation from before, an unfair compensation scheme might be particularly frustrating if it distorts rewards based on own actions/performance. In our study, participants in the privileged group are neither responsible for the inequity in the effort-to pay ratio, nor can they balance it through any kind of positive transfers. These specific design features could explain why others do not "blame" those participants for the unequal distribution and do not have too many negative emotions towards them. This could be an explanation why destruction rates are on a low to moderate level, even in unequal income pairings where the target is in a better position. This conjecture is supported by the data in two ways: First, fairness perceptions of the income allocation process are not correlated with the money burning decisions (see results in section 4.4). This indicates that the

sions. In treatment Bonus & Cheating, the emotion "anger" is highly significant for the likelihood to engage in money burning.

[19]It only slightly increased one's chances of not being hit by the punishment choices.

unfair process probably does not trigger further antisocial behavior in excess to some baseline level of destruction. Secondly, in the debriefing questionnaire several participants report that *"the lack of responsibility of the other person"* was why they decided not to destroy any money from that person.

Supposedly, if participants had the option but did not choose to balance differences stemming from an unfair process, they would have been the target of destruction more often. Evidence supporting this hypothesis comes from studies where allocations can be rejected, leading to mutual payoff losses (see e.g. Blount, 1995 and Bolton et al., 2005).[20] In both studies, even highly unequal allocations are frequently accepted, if they are not within the influence of the "proposing" side. However, if they are a deliberate choice of the other player, rejection rates are significantly higher. Furthermore, when participants fall short of expectations from common sharing norms, they frequently get punished, even by neutral third parties (Fehr & Fischbacher, 2004).

From an applied perspective, it is easy to think of many situations where resources are distributed unequally. In several such cases, the corresponding mechanism or underlying process is perceived as unfair (e.g. wealth accumulated through heritages or financial transfers, unfair differences in wages or job promotions due to imperfect monitoring/mistakes etc.). Such perceptions, which contain a lot of (e.g. country-specific) variation can have a high impact on preferences for redistribution (Alesina & Angeletos, 2005). When it is not directly possible to assist those disadvantaged by such "unfair" processes, it could be that individuals would like to reduce the income gap to the privileged classes. In some cases this could lead to people engaging in harmful/antisocial activities like e.g. vandalism, sabotage[21] or harmful behavior towards co-workers. This can lead to negative consequences for the company or society as a whole. The results of the present study indicate that such aspirations might not necessarily be very pronounced, at least under certain conditions. That means harmful actions are potentially unlikely, even if they are not very costly (e.g. when there is a low risk of detection). This crucially depends on persons' entitlements to their income and the possibilities to correct potential injustices of an unfair allocating process.

[20] In these studies, 'rejecting a proposal' could be interpreted as some form of punishment and not as antisocial behavior. Both have in common that the decision leads to a lower payoff for both parties, but they differ in their underlying motivations. Adding individual responsibility for an allocation blurs the line between these two concepts.

[21] see (Gangadharan et al., 2020) for an overview of studies dealing with sabotage at the workplace.

References

Abbink, K., Masclet, D., & van Veelen, M. (2011). Reference point effects in antisocial preferences. *Working Paper*.

Abbink, K., & Sadrieh, A. (2009). The pleasure of being nasty. *Economics Letters*, *105*(3), 306–308.

Adams, J. S. (1963). Towards an understanding of inequity. *The Journal of Abnormal and Social Psychology*, *67*(5), 422–436.

Akbaş, M., Ariely, D., & Yuksel, S. (2019). When is inequality fair? an experiment on the effect of procedural justice and agency. *Journal of Economic Behavior & Organization*, *161*, 114–127.

Alesina, A., & Angeletos, G.-M. (2005). Fairness and redistribution. *American Economic Review*, *95*(4), 960–980.

Alesina, A., Tella, R. D., & MacCulloch, R. (2004). Inequality and happiness: are europeans and americans different? *Journal of Public Economics*, *88*, 2009–2042.

Alvaredo, F., Chancel, L., Piketty, T., Saez, E., & Zucman, G. (2018). *World inequality report 2018*. Belknap Press.

Ambrose, M., Seabright, M., & Schminke, M. (2002). Sabotage in the workplace: The role of organizational injustice. *Organizational behavior and human decision processes*, *89*(1), 947–965.

Barr, A., Burns, J., Miller, L., & Shaw, I. (2015). Economic status and acknowledgement of earned entitlement. *Journal of Economic Behavior & Organization*, *118*, 55–68.

Berkowitz, L. (1989). Frustration-aggression hypothesis: examination and reformulation. *Psychological bulletin*, *106*(1), 59.

Blount, S. (1995). When social outcomes aren't fair: The effect of causal attributions on preferences. *Organizational Behavior and Human Decision Processes*, *63*(2), 131–144.

Bolton, G. E., Brandts, J., & Ockenfels, A. (2005). Fair procedures: Evidence from games involving lotteries. *The Economic Journal*, *115*(506), 1054–1076.

Cappelen, A. W., Hole, A. D., Sørensen, E. Ø., & Tungodden, B. (2007). The pluralism of fairness ideals: An experimental approach. *American Economic Review*, *97*(3), 818–827.

Cetre, S., Lobeck, M., Senik, C., & Verdier, T. (2019). Preferences over income distribution: Evidence from a choice experiment. *Journal of Economic Psychology*, *74*, 102202.

Dawes, C. T., Fowler, J. H., Johnson, T., McElreath, R., & Smirnov, O. (2007). Egalitarian motives in humans. *Nature*, *446*(7137), 794–796.

Demiral, E. E., & Mollerstrom, J. (2020). The entitlement effect in the ultimatum game–does it even exist? *Journal of Economic Behavior & Organization*, *175*, 341-352.

Fehr, D. (2018). Is increasing inequality harmful? Experimental evidence. *Games and Economic Behavior*, *107*, 123–134.

Fehr, E., & Fischbacher, U. (2004). Third-party punishment and social norms. *Evolution and Human Behavior*, *25*(2), 63–87.

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, *114*(3), 817–868.

Frohlich, N., Oppenheimer, J., & Kurki, A. (2004). Modeling other-regarding preferences and an experimental test. *Public Choice*, *119*(1), 91-117.

Gangadharan, L., Grossman, P. J., & Vecci, J. (2020). Antisocial behavior in the workplace. In *Handbook of labor, human resources and population economics.* Springer International Publishing.

Grimalda, G., Kar, A., & Proto, E. (2016). Procedural fairness in lotteries assigning initial roles in a dynamic setting. *Experimental Economics*, *19*(4), 819–841.

Grosch, K., & Rau, H. A. (2020). Procedural unfair wage differentials and their effects on unethical behavior. *Economic Inquiry*, *58*(4), 1689–1706.

Grossman, P. J., Komai, M., et al. (2013). Within and across class envy: Antisocial behaviour in hierarchical groups. *Working Paper*.

Hoffman, E., McCabe, K., Shachat, K., & Smith, V. (1994). Preferences, property rights, and anonymity in bargaining games. *Games and Economic Behavior*, *7*(3), 346–380.

Jauernig, J., Uhl, M., & Luetge, C. (2016). Competition-induced punishment of winners and losers: Who is the target? *Journal of Economic Psychology*, *57*, 13–25.

Karagözoğlu, E., & Riedl, A. (2015). Performance information, production uncertainty, and subjective entitlements in bargaining. *Management Science*, *61*(11), 2611–2626.

Konow, J. (2001). Fair and square: The four sides of distributive justice. *Journal of Economic Behavior & Organization*, *46*(2), 137–164.

Konow, J. (2003). Which is the fairest one of all? A positive analysis of justice theories. *Journal of Economic Literature*, *41*(4), 1188–1239.

List, J. A. (2007, jun). On the interpretation of giving in dictator games. *Journal of Political Economy*, *115*(3), 482–493. doi: 10.1086/519249

Miller, N. E. (1941). I. the frustration-aggression hypothesis. *Psychological review*, *48*(4), 337.

Munyo, I., & Rossi, M. A. (2013). Frustration, euphoria, and violent crime. *Journal of Economic Behavior & Organization*, *89*, 136–142.

Oishi, S., Kesebir, S., & Diener, E. (2011, aug). Income inequality and happiness. *Psychological Science*, *22*(9), 1095–1100.

Oxoby, R. J., & Spraggon, J. (2008). Mine and yours: Property rights in dictator games. *Journal of Economic Behavior & Organization*, *65*(3-4), 703–713.

Schneider, S. M. (2015). Income inequality and subjective wellbeing: Trends, challenges, and research directions. *Journal of Happiness Studies*, *17*(4), 1719–1739.

Zizzo, D. J. (2003). Money burning and rank egalitarianism with random dictators. *Economics Letters*, *81*(2), 263–266.

Zizzo, D. J., & Oswald, A. J. (2001). Are people willing to pay to reduce others' incomes? *Annales d'Economie et de Statistique*, *63*, 39–65.

[Part 1:] **Experimental Instructions** Code-No. _____

[All Treatments:] Welcome to this experiment and thank you very much for your participation!

Please do not communicate with other participants and switch off your mobile phones. If you have any questions raise your hand and one of the experimenters will come over to your place and answer it privately.

This is an experimental study about decision-making. You can earn a certain amount of money, which will be paid to you privately and in cash at the end of the experiment. Your exact payoff depends on your own decisions and those of other participants.

[Random:] **First, your initial income will be determined by a random process.**
As income, you will either receive 5€ or 10€.

The randomization will be done using a box with a certain number of envelopes in it. Those envelopes contain a note that you are assigned either the amount of 5€ or 10€. The number of envelopes is chosen such that exactly half of the participants receive each amount.

[Fair and Unfair:] This experiment begins at two different points in time. **Half of the participants start 25-30 min later than the other group.** The assignment to the two groups was determined beforehand and notifications about the starting time were sent in accordance. **Those who belong to the group starting earlier are given a task, which they have to work on during this period.**

This task is to mark IQ test sheets, which have been done by a group of elderly participants from another recently conducted study of this department. For this, participants in this study will receive a solution scheme and their task is to determine the amount of points achieved in each exercise.

The participants belonging to the late group do not have to perform any kind of task. They immediately start with the second part of the experiment, together with the earlier group. This part is identical for all participants.

The participants who had to work on the task beforehand receive an income of 10€ [*Unfair:* 5€] and those who started later receive an income of 5€ [*Unfair:* 10€].

[*Unfair:* Note for clarification: The participants in the late group receive the higher income of 10€ for this part]

[All Treatments:] In the second stage, you might have the possibility to reduce the income of one of the other participants. You will be informed about the exact details after the initial income is assigned. All this information is known by everyone and instructions are identical for all participants.

Finally, there will be a debriefing questionnaire and afterwards you will receive your payment.

[Part 2:]

Your income is 5€ [10€] Code-No._____

As described before, now the income of certain participants may be reduced.

You and a randomly chosen other participant form a group-of-two. You may be paired with a person, who has either the same or a different income as you. During the whole experiment, there is complete anonymity about the group composition and the identities of the involved players.

You now are given the opportunity to reduce the income of the other participant in your group by a certain amount. You have to pay a cost of 10% of the chosen reduction value. This is, for each Euro you want to deduct from the income of the other player you have to pay 10 cents. You can freely choose any amount of reduction, however not more than the value of total income of the other person. Decisions are made for both the cases that the other person has an income of 5€ or an income of 10€. You do not have to make use of the possibility to reduce the income of the other person.

After all participants make their decisions, one person per group will be randomly chosen. The decision of this person will be carried out. That means in your group either your own or the decision of the other player will be payoff-relevant. Of course, you have to pay the costs of reduction only if your own decision is finally implemented. You and the other group member make your decisions simultaneously and independently of each other.

Enter your decisions on the sheet of the next page. Afterwards, the group assignment will take place and the payoff-relevant decisions will be determined. In doing so, every participant draws a card with a letter on it, indicating both the group composition and the decision, which is going to be implemented. Please record the letter you have drawn on the respective box on your decision sheet. Afterwards, put the decision sheet in the envelope you have received beforehand. The envelopes will be collected and evaluated by the experimenters to calculate the final profits or all participants.

Decision sheet

Code-No._____

Your income is 5€ [10€]

Amount, you want to reduce from the other person, in case she/he has an income of 5€:

Amount, you want to reduce from the other person, in case she/he has an income of 10€:

Letter drawn for the group composition: