

Suspicious Success — Cheating, Inequality Acceptance, and Political Preferences

Felix Klimm*

March 9, 2018

Abstract

Supporters of left-wing parties typically place more emphasis on redistributive policies than right-wing voters. I investigate whether this difference in tolerating inequality is amplified by suspicious success — achievements that may arise from cheating. Using a laboratory experiment, I exogenously vary cheating opportunities for stakeholders who work on a real effort task and earn money according to their self-reported performances. An impartial spectator is able to redistribute the earnings between the stakeholders, although it is not possible to detect cheating. I find that the opportunity to cheat leads to different views on whether to accept inequality. Left-wing spectators substantially reduce inequality when cheating is possible, while the treatment has no significant effect on choices of right-wing spectators. Since neither differences in beliefs nor differences in norms about cheating can explain this finding, it seems to be driven by a difference in preferences. These results suggest that redistributive preferences will diverge even more once public awareness increases that inequality may be to a certain extent created by cheating.

JEL-classification: C91, D63, D83, H23, H26

Keywords: Cheating, inequality, fairness, political preferences, redistribution

*LMU Munich, Giselastr. 10 (room 103), 80802 Munich, Germany. Contact: felix.klimm@econ.lmu.de
I would like to thank Johannes Abeler, Ingvild Almås, Armin Falk, Ulrich Glockowsky, Stefan Grimm, Lars Hornuf, Konstantin Lucks, Martin Kocher, Klaus Schmidt, Simeon Schudy, and Bertil Tungodden, as well as participants at the 2016 Choice Lab PhD course with Armin Falk at NHH Norwegian School of Economics, 2017 ESA Europe Meeting and seminars at the University of Munich in 2016 and 2017 for helpful suggestions and comments. I gratefully acknowledge the financial support of the German Research Foundation (DFG) through GRK 1928 and CRC TRR 190.

1 Introduction

Whether unequal outcomes are considered to be fair primarily depends on the sources of inequality. People prefer to eliminate income disparities that have resulted from factors beyond individual control such as pure luck, physical handicap, gender, or family background, yet they tend to accept inequalities based on differences in effort, initiative, or the willingness to take risks (Konow, 2000; Fong, 2001; Cappelen et al., 2013; Möllerström et al., 2015; Almås et al., 2016). The sources of inequality also lie at the core of the political debate about redistribution. Whereas right-wingers believe that one's fortunes are mainly the consequences of effort and choices, left-wingers place more emphasis on the notion that uncontrollable luck determines income (Piketty, 1995; Alesina and Angeletos, 2005; Cappelen et al., 2010; Cappelen et al., 2016).

In this paper, I investigate whether the difference in redistributive preferences between the two political camps persists with regard to another source of inequality — cheating. Because everyday life is permeated with cheating opportunities, ranging from an employee tempted to overstate hours worked to potential submission of false claims by a physician, people might be suspicious of the wealth of the successful. For example, the recent leaks of the “Panama Papers” as well as the “Paradise Papers” have revealed that a large fraction of global financial wealth is held in tax havens.¹ Zucman (2014) estimates that annually foregone tax revenues due to offshore tax evasion amount to \$190 billion, which suggests that a significant share of wealth is created by illegal financial activities.²

A particular feature of inequalities which arise from fraudulent behavior is that although cheating is within individual control, it is, as opposed to effort, unlikely to be regarded as fair (e.g., Kirchler et al., 2003). Therefore, it remains an open

¹See, e.g., <http://www.bbc.com/news/world-41880153>, last accessed on March 5, 2018.

²Inequalities based on cheating are not limited to tax evasion, but there are various other forms of performance cheating that cause someone to be more successful than others. For instance, businessmen fabricate their curriculum vitae to get better paid jobs, athletes take performance enhancing doping substances to win prestigious competitions, and firms manipulate software to maximize profits. (see, e.g., <http://www.dailymail.co.uk/news/article-2669969/CV-fake-hired-Myer-considered-companies.html>, <http://www.telegraph.co.uk/sport/othersports/cycling/lancearmstrong/9810199/Lance-Armstrong-tells-Oprah-Winfrey-he-doped-during-all-seven-Tour-de-France-victories.html>, <https://www.nytimes.com/interactive/2015/business/international/vw-diesel-emissions-scandal-explained.html>, last accessed on March 5, 2018).

question whether left-wingers also demand more redistribution than right-wingers in the presence of cheating opportunities. However, with regard to the prevalence of suspicious success, this question needs to be answered in order to understand the origins of different views on the necessity of redistributive policies.

I address this question by conducting a between-subjects experiment, where some participants work on a real effort task (henceforth called stakeholders). Two stakeholders are matched with another and split a fixed amount of money according to their performances. Stakeholders can overstate their performance in the *Cheat* treatment, which does not affect the total income of the two stakeholders but shifts the distribution of income in favor of the misreporting stakeholder. This captures the impact of cheating behavior in many situations of economic relevance. For instance, tax evasion does not alter the amount of money necessary to provide public goods, but at the same time honest tax payers bear the cost of cheating in the long run. Contrary to the *Cheat* treatment, stakeholders' performances are audited in the *Monitor* treatment, which renders misreporting impossible.

Third-party participants (henceforth called spectators) are able to redistribute the earnings of the two stakeholders (following Cappelen et al., 2013). In both treatments, they are fully aware of the rules for working on the real effort task and the (lacking) possibility to misreport own performance. Importantly, in the *Cheat* treatment, dishonest stakeholders cannot be identified, and spectators thus never know whether a stakeholder reported untruthfully. Therefore, if one of the two stakeholders earns considerably more than the other one, spectators might believe that high income results from cheating but do not know whether their suspicion is accurate. This uncertainty leaves room to justify both eliminating inequalities as well as refraining from doing so.

Using a laboratory experiment allows to provide clean evidence on the effect of cheating opportunities on inequality acceptance for two main reasons. First, it allows for exogenous manipulation of the availability to cheat, which is difficult to achieve in field settings given the nature of naturally occurring cheating opportunities. Second, eliciting redistributive preferences from impartial spectators makes it possible

to exclude confounding factors such as selfishness, self-centered inequality aversion (e.g., Fehr and Schmidt, 1999), or reciprocity (e.g., Rabin, 1993).

My results show that the treatment effect depends on political preferences. Right-wing spectators hesitate to redistribute on the basis of potential cheating as they implement the same levels of inequality in *Monitor* and *Cheat*. In contrast, distributive choices of left-wing spectators reveal an increase of 74% in inequality reduction due to cheating opportunities. The analysis of the treatment effect for different levels of pre-redistribution inequality shows that left-wing spectators react to potential cheating only for high levels of inequality. This provides evidence that they believe that the “rich” stakeholder is cheating in these situations.

There are essentially three different explanations for the polarization in redistributive preferences between the two political camps. (i) Right-wing spectators might believe to a lesser extent that stakeholders are cheating than left-wingers. (ii) Right-wing spectators’ norms about cheating differ from those of left-wing supporters: They find it more acceptable to cheat when possible. (iii) Right-wingers prefer not to redistribute due to potential cheating if they do not know whether a stakeholder indeed cheated, although they know that misreporting is prevalent. In order to distinguish between these three explanations, I examine beliefs and norms about cheating and find no differences between left-wingers and right-wingers. Therefore, the political divide in how to deal with unequal outcomes that might arise from dishonest behavior seems to reflect different preferences. This suggests that different views on the importance of redistributive policies diverge even more in the light of scandals about cheating by the “rich and successful” as we know that redistributive preferences are highly elastic to information (Kuziemko et al., 2015).

My paper contributes to several strands of the literature. First, it relates to studies on the determinants of redistributive preferences. Papers that use survey data find that personal characteristics such as gender, race, and education as well as cultural background and past experience of personal traumas (e.g., divorce, hospitalization, or death of a relative) predict redistributive preferences (Alesina and Ferrara, 2005; Alesina and Giuliano, 2011). Moreover, using data from the General Social Survey,

Fong (2001) shows that people who believe that luck causes poverty and wealth support redistribution to a much larger extent than people who believe that effort causes poverty and wealth. In addition, experimental studies indicate that people also care about whether someone can be held responsible for one's own luck by choosing a risky or a safe option (Cappelen et al., 2013; Möllerström et al., 2015). Closely related to my paper, Bortolotti et al. (2017) experimentally investigate redistributive preferences when cheating with regard to a risky outcome, a coin flip, is possible. The authors document a shift in fairness views due to potential cheating in favor of strict egalitarianism, i.e., implementing an equal distribution of income independent of subjects' choices that affected earnings in the first place. Importantly, while Bortolotti et al. (2017) study redistributive preferences in the light of cheating in the luck domain, I focus on situations where people can cheat regarding their performance.

Second, my results show that political preferences matter for accepting inequalities that might arise from cheating behavior. Interestingly, the evidence on whether political preferences affect choices in allocation decisions is mixed. While a number of studies report significant differences across political preferences (Van Lange et al., 2012; Cappelen et al., 2013; Cappelen et al., 2016; Bortolotti et al., 2017), there are two studies that find only weakly significant or insignificant effects of political preferences (Frohlich et al., 1984; Fehr et al., 2006). Therefore, the impact of political preferences on distributional choices seems to depend on the specific context.

Third, giving participants the opportunity to cheat relates my paper to a growing experimental literature on dishonesty (e.g., Gneezy, 2005; Mazar et al., 2008; Fischbacher and Föllmi-Heusi, 2013; Shalvi and De Dreu, 2014; Conrads and Lotz, 2015; Houser et al., 2016).³ While this literature is primarily concerned with the extent and the causes of cheating, my study is one of the few that deal with the consequences of cheating by showing that dishonesty affects the behavior of third parties (Pigors and Rockenbach, 2016; Bortolotti et al., 2017; Cappelen et al., 2017).

The remainder of the paper is structured as follows. Section 2 describes the experimental design in detail. Section 3 presents the results and Section 4 concludes.

³See Abeler et al. (2016) for a meta-study on data from 72 experimental studies on dishonesty.

2 Experimental Design

The experiment consists of two main parts. Part 1 concerns the real effort provision and potential cheating in the matrix task, which was introduced in the literature by Mazar et al. (2008). Part 2 uses decisions from impartial spectators in order to measure redistributive preferences, similar to Cappelen et al. (2013). Thereafter, beliefs and political preferences are elicited.

2.1 Part 1: The Matrix Task

All subjects receive an exercise sheet with 20 matrices, each containing a set of twelve numbers with two decimal places (see Appendix D.1 for an example). Only two of these twelve numbers add up to exactly 10. The task is to find these two numbers and solve as many matrices as possible within 6 minutes.

Two players A (the stakeholders) are randomly matched in order to determine their preliminary income (i.e., income before redistribution). Proportionally to their performance in the matrix task, €10 are split up among these two participants. This distribution of income is rounded to 50 cents.

Treatment variation. After working on the task, the stakeholders are provided with the correct solutions on their screens and are asked to compare them with their own solutions (see Figure C.1 in the Appendix for an example of the stakeholders' decision screen). They then report for each matrix whether they solved it correctly. Subjects in *Monitor* are informed that all exercise sheets are collected to verify their reported performance and that, if necessary, their reports will be changed to their actual performance.⁴ In contrast, subjects in *Cheat* are informed that they will shred their exercise sheet at the end of the experiment, and it is thus impossible to monitor their solutions.

The design of Part 1 has at least three desirable features for the purpose of this study. First, the real effort task mimics a wide range of field settings where

⁴The fraction of stakeholders for whom performances had to be corrected downwards is 10% (upwards 5%). In 57% of these cases, the difference between reported and actual performance was one task.

people engage in performance cheating in order to serve their self-interests. Consider, for instance, an employee who misreports his number of hours worked to receive either a higher wage or more days off. Second, exaggerating own performance implies a negative externality for the other stakeholder, which reflects the adverse consequences of cheating in many “real-world” situations. Coming back to the example of overstating hours worked, honest colleagues might be affected through a lower likelihood of being promoted due to inferior relative performance. Third, having a fixed sum of payments for the two stakeholders excludes efficiency concerns as a motivation for cheating. This is important for redistribution decisions because if cheating was efficiency enhancing, this might confound the moral assessment of such behavior and it would be difficult to account for this motive in the experiment.

2.2 Part 2: Redistribution Decisions

Each player B (the spectator) is matched with a pair of stakeholders. The strategy method is used for the spectators’ decisions. Hence, for each of the eleven possible distributions of preliminary income (going in steps of 50 cents from maximum inequality in the case of (10,0) to full equality in the case of (5,5)), they can transfer money within a pair of stakeholders and consequently determine the two stakeholders’ final income (i.e., income after redistribution).⁵ Redistribution of the preliminary income is possible in steps of 10 cents. In order to be able to unambiguously refer to the stakeholders’ final income, the stakeholder with the higher or equal preliminary income is called player A1 and the other stakeholder player A2 (see Figure C.2 in the Appendix for the spectators’ decision screen).

Spectators receive a fixed income of €10 for their redistribution decisions. Giving spectators at least the sum of earnings of a pair of stakeholders assures that self-centered inequity aversion based on Fehr and Schmidt (1999) does not affect spectators’ behavior.⁶ Moreover, spectators are informed that either their decision or

⁵Brandts and Charness (2011) provide an analysis of 29 studies in order to compare the strategy method with the direct respond method. Since they do not find a single case in which there is a treatment effect using the strategy method that vanishes with the direct respond method, the strategy method is likely to yield a lower bound for this experiment’s treatment effect.

⁶For instance, the Fehr-Schmidt model predicts the spectator to choose full equality when receiving

the decision of another spectator will be randomly implemented, which is designed to increase the number of decisions taken by spectators.⁷

2.3 Belief Elicitation

After stakeholders report their performance and spectators make their redistribution decisions, two beliefs are elicited. First, stakeholders and spectators are asked to guess the average reported performance of the stakeholders in their own session (*belief-own-treat*). Consequently, subjects in *Monitor* report their belief about how many tasks were actually solved, while subjects in *Cheat* guess how many tasks the stakeholders reported to have solved. Second, I elicit beliefs about how many correctly solved tasks the stakeholders reported in the respective other treatment (*belief-other-treat*). Therefore, subjects are informed that stakeholders worked on exactly the same task in a previously run experiment but that reported performance was monitored differently (instructions can be found in Appendix D.4). Subjects learn about the respective other treatment not before beliefs are elicited.

Beliefs are incentivized with €2 for deviations up to one task and €1 for deviations up to two tasks, whereas larger deviations are not paid. Only one of the two beliefs is randomly chosen for payment in order to prevent hedging. Since *belief-other-treat* is based on the first session of the respective other treatment, it is not elicited in the first session of each treatment.

Moreover, I elicit a third belief at the end of each session of the *Cheat* treatment. Subjects guess the fraction of stakeholders who did not report truthfully (*belief-frac-cheat*). It is impossible to incentivize these beliefs because I refrain from individual cheating detection.

2.4 Political Preferences

At the end of the experiment, I ask participants about their political preferences (i.e., which party they would vote for if there were federal elections next Sunday). In

a fixed income of €5, independent of the distribution of preliminary income.

⁷As a consequence, the number of spectators ($n = 182$) equals the number of stakeholders ($n = 182$).

my analysis, being left-wing is defined as indicating to vote for the Social Democrats (SPD), the Green Party (Die Grünen), the socialist party (Die Linke), or the Pirate Party Germany (Die Piraten).⁸ Subjects belonging to the remaining categories are treated as being right-wing. Following this definition, 37.36% of the spectators are classified as left-wing and 62.64% as right-wing. The distribution of spectators' votes closely resembles the results of the 2017 German national election. Thus, in terms of political preferences, the sample of the experiment is similar to the German population. In order to validate my classification of political parties, subjects are asked to indicate where they rate their general political attitudes on a scale from 1 to 10 with 1 being left and 10 being right. Although this question may be susceptible to the central tendency bias (64% of subjects indicate a score of 4, 5, or 6), left-wing spectators rate themselves lower on this scale than right-wing spectators (mean left-wing = 3.94, mean right-wing = 5.37). The difference in ratings across political preferences is significant ($p < 0.0001$, Mann-Whitney U -test, two-sided). Table A.1 in the Appendix shows which parties spectators would vote for, the 2017 German federal election results as well as the average scores of spectators' general political attitudes.

2.5 Procedural Details

The experiment was conducted with 364 participants at the Munich Experimental Laboratory for Economic and Social Sciences (MELESSA) at the University of Munich in May 2016 and May 2017. In total, 184 subjects were assigned to eight sessions of the *Monitor* treatment and 180 subjects to eight sessions of the *Cheat* treatment.⁹ Subjects were students from various fields of study and recruited using the online system "ORSEE" (Greiner, 2015). Each subject was randomly assigned to one of the two treatments and participated in one session only. The experiment was programmed and conducted with the software "z-Tree" (Fischbacher, 2007).

⁸A coalition of the three established parties SPD, Die Grünen, and Die Linke is also called "left-wing coalition", see, e.g., <http://www.telegraph.co.uk/news/worldnews/europe/germany/10318949/Germanys-coalitions-What-happens-next.html> (last accessed on March 5, 2018). The Pirate Party Germany, which was found in 2006, is typically classified as being left-wing.

⁹For each of the two treatments, half of the sessions were conducted in 2016 and the other half in 2017. The time of conducting the experiment does not affect the results (see Section 3.2).

Upon arrival at the laboratory, subjects found a printed version of the instructions of Part 1 at their seats, which was read aloud by the experimenter (myself) to ensure common knowledge about the rules of the real effort task (see Appendix D.1). In addition, these instructions informed participants that only later on they will be assigned to one of the two roles. Hence, stakeholders as well as spectators worked on the matrix task. In this way, spectators were familiar with the difficulty of the matrix task, which was important for eliciting their beliefs about the stakeholders' performances. Furthermore, subjects were told that the spectator can distribute earnings of Part 1 and that the spectator will receive more detailed information on this at a later point in time.¹⁰ Thereafter, subjects received the exercise sheet and were provided with pens for marking their solutions. After timeout, a second set of instructions about how to compare own with correct solutions appeared on the subjects' screens and were read aloud (see Appendix D.2). Subsequently, subjects were displayed their role, and stakeholders self-reported their performance while spectators received detailed instructions about Part 2 (see Appendix D.3) and made their redistribution decisions. Thus, *only* stakeholders compared their own solutions with the correct ones. After that, the exercise sheets were collected and verified in *Monitor*, and subjects in both treatments indicated on a 4 point Likert scale whether they considered it to be fair that preliminary income was proportional to self-reported performance. Next, *belief-own-treat* and *belief-other-treat* were elicited.

Finally, the participants answered a questionnaire about their political preferences, opinion on income inequality in Germany on a scale from 1 (inequality should be reduced) to 10 (inequality should be enlarged in order to provide incentives), socio-demographic characteristics, and their belief about the fraction of stakeholders who did not report truthfully (*belief-frac-cheat*, only in the *Cheat* treatment). Subjects received their payments privately after the experiment and earned €12.22 on average, including an average show-up fee of €4.5.¹¹ Sessions lasted on average 50 minutes.

¹⁰Little information about Part 2 cannot exclude the possibility of strategic effort provision. However, this would not affect spectators redistributive choices because they were elicited with the strategy method and are thus independent of actual performances.

¹¹The show-up fee was €4 in 2016 and €5 in 2017 due to changes of the rules of MELESSA.

3 Results

3.1 Matrix Task Performance

Figure 1 shows that reported performances are significantly higher in *Cheat* than in *Monitor* and indicates that manipulation by giving stakeholders the opportunity to cheat was successful ($p < 0.001$, Mann-Whitney U -test, two-sided). Average reported performance is higher in *Cheat* (13.8) than in *Monitor* (11.5). Moreover, while only 5.4% of the stakeholders in *Monitor* indicate to have solved all matrices, this is the case for 16.7% of the stakeholders in *Cheat*. This is line with the finding of previous studies on dishonesty that although some people are cheating, the assumption of people always submitting payoff-maximizing reports is empirically not valid (e.g., Mazar et al., 2008; Fischbacher and Föllmi-Heusi, 2013; Jiang, 2013; Cohn et al., 2014; Abeler et al., 2016).

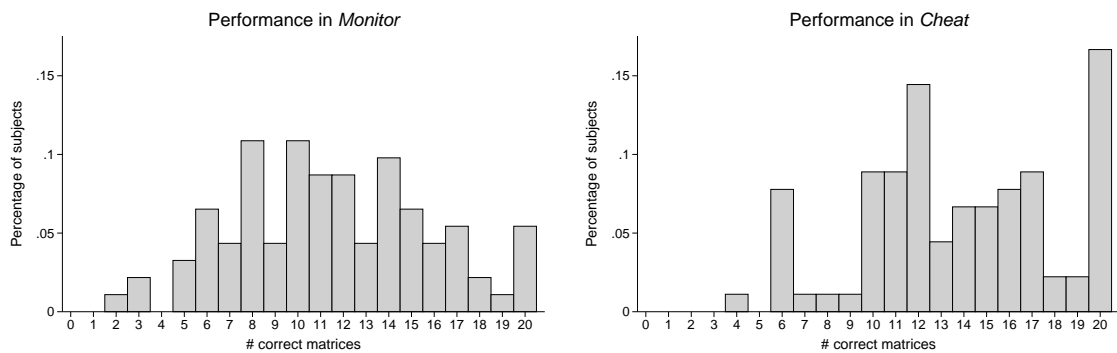


Figure 1: Distribution of reported performances

In addition, subjects' beliefs about average reported performance in their own session is significantly different across treatments ($p < 0.0001$, Mann-Whitney U -test, two-sided). On average, *belief-own-treat* is 1.9 tasks higher in *Cheat* than in *Monitor*, which is close to the actual difference of 2.3 tasks. Furthermore, subjects in *Monitor* find the income generating process of preliminary income significantly more fair than subjects in *Cheat* ($p = 0.037$, Mann-Whitney U -test, two-sided). Both of these results are further indications of successful treatment manipulation.

3.2 Inequality Acceptance and Political Preferences

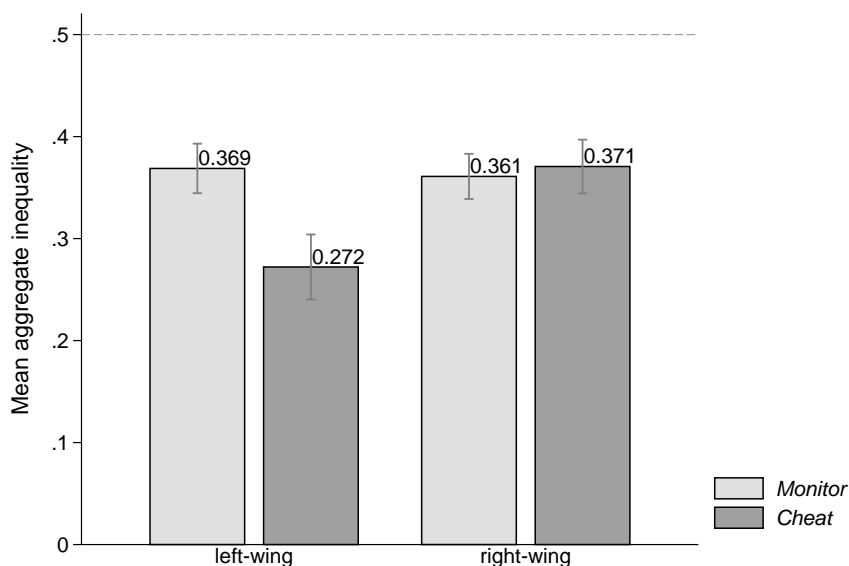
The redistribution decisions of the spectators determine the final income distribution between two stakeholders. In order to quantify the extend to which spectators are willing to accept inequalities, I use the Gini coefficient as inequality measure:

$$\text{Inequality} = \frac{|\text{Income Player A1} - \text{Income Player A2}|}{\text{Income Player A1} + \text{Income Player A2}}$$

This measure of inequality relates the absolute difference in income to the total income and is zero in cases of full equality and one if one of the two stakeholders receives the entire total income. I define aggregate inequality as the average over the Gini coefficients of the eleven possible distributions of preliminary income. Thus, aggregate inequality contains the spectator's decisions for perfectly unequal preliminary incomes of (10,0), full equality in the case of (5,5) as well as all cases in between. Using this measure yields a lower bound for the treatment effect because one can expect hardly any redistribution in cases of low inequality (e.g., (5,5)) in both treatments.

Figure 2 shows the mean aggregate inequality across the two treatments and political preferences. If a spectator never redistributes income, aggregate inequality is 0.5, which is indicated by the horizontal dashed line. Left-wing spectators implement significantly lower inequality in *Cheat* than in *Monitor* ($p = 0.027$, Mann-Whitney *U*-test, two-sided). While they eliminate 26.2% of initial inequality in *Monitor*, they reduce inequality by 45.6% in *Cheat*, implying an increase of 74% in inequality reduction. In contrast, the treatment has no significant effect for right-wing spectators ($p = 0.641$, Mann-Whitney *U*-test, two-sided). They reduce inequality by 27.8% in *Monitor* and 25.8% in *Cheat*.

Table 1 contains a series of Tobit regressions to account for using the censored dependent variable aggregate inequality. Column (1) suggests that aggregate inequality is not affected by the treatment for the pooled sample, which is due to the fact that the majority of spectators are right-wing. Column (2) confirms the result depicted in Figure 2. Inequality is significantly reduced through cheating



Notes: The figure shows aggregate inequality defined as the average Gini coefficient of all eleven redistribution decisions. The dashed line indicates aggregate inequality in case of no redistribution. Error bars indicate standard errors of the mean.

Figure 2: Aggregate inequality

opportunities when being left-wing ($p = 0.004$). However, the treatment does not affect implemented inequality of right-wing spectators. The sum of the treatment dummy and the interaction term “Cheat \times Right-wing” in column (2) indicates that the treatment effect for right-wing spectators is insignificant ($p = 0.791$, F -Test).¹² Furthermore, being right-wing has no impact on aggregate inequality in the *Monitor* treatment ($p = 0.805$). In column (3), I add as regressor the answer to whether the spectators are of the opinion that income inequality should be reduced, or enlarged in order to provide incentives for individual performance. Believing that “we need inequality” increases implemented income inequality, while the treatment effect remains significant for left-wing spectators ($p = 0.003$). Hence, differences in the opinion on inequality¹³ cannot explain that the effect of cheating opportunities depends on political preferences. The results are robust to adding a time dummy

¹²Interpreting the coefficient of an interaction term can be misleading in Tobit models (Ai and Norton, 2003). To examine this problem, I perform an alternative calculation of the interaction effect by computing the predicted values of aggregate inequality separately for left-wing and right-wing spectators in *Monitor* and *Cheat*. The respective difference in differences of these four groups’ predicted values are of the same size as the marginal effect of the interaction terms in the models of column (2) to (6) in Table 1. Thus, the bias induced by using interaction terms in a nonlinear model is negligible in my estimations.

¹³Right-wing spectators (mean = 4.95) favor inequality significantly more than left-wing spectators (mean = 3.49, $p < 0.0001$, Mann-Whitney U -test, two-sided).

Table 1: Aggregate inequality

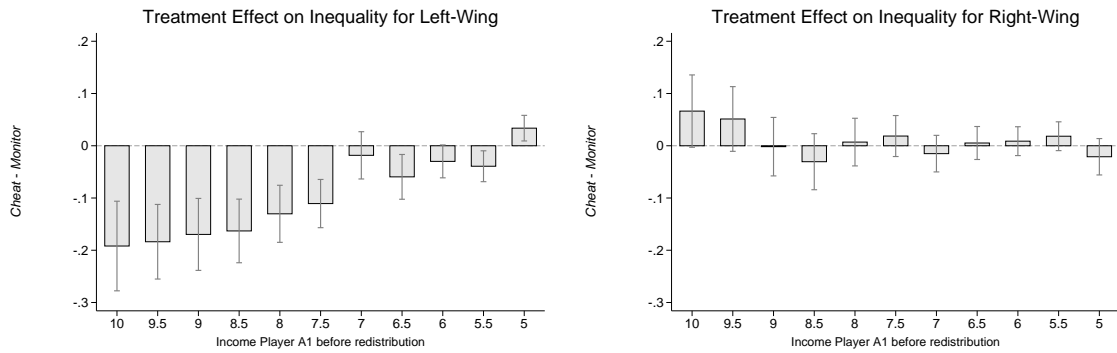
Dependent variable	Aggregate inequality					
	(1)	(2)	(3)	(4)	(5)	(6)
Cheat	-0.036 (0.025)	-0.112*** (0.038)	-0.110*** (0.037)	-0.110*** (0.040)	-0.089** (0.041)	-0.118*** (0.039)
Right-wing		-0.010 (0.042)	-0.028 (0.041)	-0.032 (0.040)	-0.006 (0.048)	-0.021 (0.048)
Cheat × Right-wing		0.120** (0.048)	0.116** (0.045)	0.115** (0.047)	0.128** (0.062)	0.164*** (0.058)
We need inequality (1 = no, 10 = yes)			0.014* (0.007)	0.013* (0.007)	0.010 (0.007)	0.012 (0.009)
Constant	0.358*** (0.010)	0.364*** (0.030)	0.316*** (0.040)	0.334*** (0.077)	0.306*** (0.075)	0.215** (0.101)
Additional controls	No	No	No	Yes	Yes	Yes
Observations	182	182	182	182	182	144
Log likelihood	1.433	4.589	6.502	7.931	9.596	21.228

Notes: Two-limit Tobit regressions regressions on aggregate inequality. Columns (4) to (6) include a binary variable for whether the experiment was conducted in 2016 or 2017 (insignificant in all specifications) and additional covariates from the questionnaire: age, gender, semester, and number of experiments so far (all insignificant in all specifications). Column (5) includes the categories “other party” and nonvoters in the definition of being left-wing. Column (6) excludes spectators who would vote for “other party” as well as nonvoters. Robust and clustered (on session level) standard errors in parentheses. Stars indicate significance on the levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

(whether the experiment was conducted in 2016 or 2017) as well as personal background characteristics in column (4). In column (5), I include nonvoters and spectators who indicate to vote for “other party” in the definition of left-wing instead of right-wing. The results are robust to this specification (with the exception that the significance of the treatment dummy decreases to $p = 0.031$). Furthermore, the results are also robust to excluding nonvoters and spectators who would vote for “other party” in column (6).¹⁴

In order to provide further evidence that left-wing spectators’ decisions are affected by the treatment *because* they suspect player A1 to be cheating, I analyze each redistribution decision separately. A spectator should doubt the performance

¹⁴In addition to investigating the Gini coefficient, I use the amount of money redistributed by spectators to study treatment differences. This is important because spectators might redistribute away from player A1 — who receives the higher preliminary income of the two stakeholders except for the case of full equality — in such a way that after redistribution player A2 has more income than player A1. For instance, a spectator in *Cheat* might determine the final income distribution to be (2,8) when preliminary incomes were (8,2), which results in the same inequality before and after redistribution. However, it is important to capture this incidence of redistribution as it might reflect punishing player A1 for potential cheating. Results on redistribution are reported in Appendix A and support the results of analyzing the Gini coefficient.



Notes: The figure shows the treatment effect on implemented inequality by subtracting inequality in the *Monitor* treatment from inequality in the *Cheat* treatment. The effect is shown separately for each preliminary level of income of player A1. The left panel displays the effect for left-wing spectators and the right panel for right-wing spectators. Error bars indicate standard errors of the mean.

Figure 3: Treatment effect on inequality

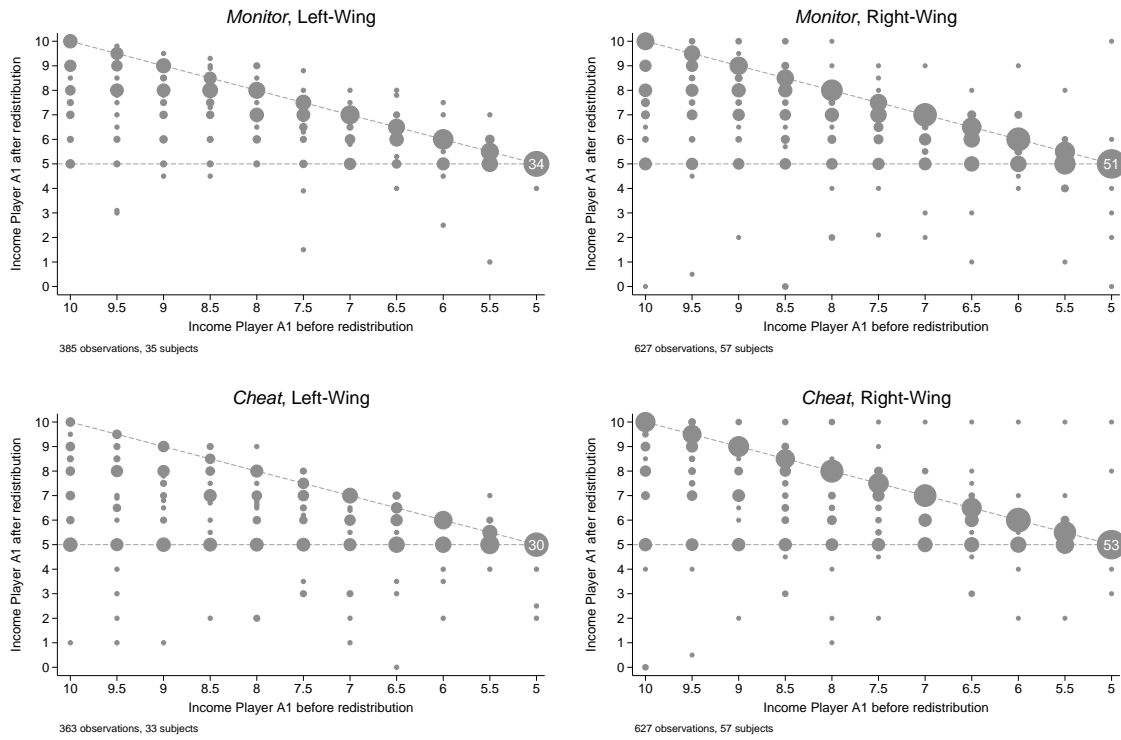
of a stakeholder if it is far above the other stakeholder's performance, which results in high income inequality.¹⁵ In contrast, there is little reason to attribute cheating to a stakeholder when income is evenly distributed and thus reported performances of both stakeholders are similar.¹⁶

Figure 3 shows the treatment effect depending on preliminary income distribution for left-wing (left panel) and right-wing spectators (right panel). The very left and the very right bar represent the treatment effect when the income distribution before redistribution is (10,0) and (5,5) respectively.¹⁷ The treatment has a significantly negative effect on the implemented inequality of left-wing spectators for high pre-redistribution inequality. Two-sided Mann-Whitney *U*-tests indicate a reduction in inequality if player A1 has a preliminary income of at least 7.5 (*p*-values do not exceed 0.0334), while there is no significant treatment effect otherwise (with the exception of player A1 having €6.5 before redistribution, $p = 0.072$). These results suggest that left-wing spectators suspect player A1 to cheat when initial inequality is high and therefore reduce inequality in these cases. There is no significant effect

¹⁵This holds true even more when only one of the two stakeholders can cheat. To analyze the impact of asymmetric cheating opportunities within a pair of stakeholders, I ran an additional treatment where only one of the two stakeholders could misreport the own performance. The results of this treatment are reported in Appendix B.

¹⁶In this case, both stakeholders could be cheating. However, there is no possibility to redistribute away from a suspicious potential cheater to another presumably more honest stakeholder because it is not possible to detect cheating.

¹⁷Levels of inequality for both treatments are shown in Figure A.2 in the Appendix.



Notes: The figure shows the spectators' redistribution decisions depending on treatment (upper panels show *Monitor* and lower panels *Cheat*) and political preferences (left panels show left-wing and right panels right-wing spectators). Numbers in circles in case of no redistribution when the preliminary income of player A1 is €5 indicate the number of observations and serve as a benchmark for the remaining circles.

Figure 4: Overview redistribution decisions

of cheating opportunities for any of the preliminary income distributions when being right-wing.

Despite analyzing implemented inequality, the question remains which kind of redistribution decisions drive the treatment effect for left-wing spectators. Therefore, Figure 4 depicts player A1's income after redistribution contingent on his income before redistribution. Circles on the downward-sloping line indicate cases of no redistribution, circles on the horizontal line cases of redistribution resulting in full equality, and circles between the two lines are associated with redistribution away from player A1 such that the ranking of incomes is maintained. The few circles above the downward-sloping line represent "negative redistribution", which leaves the pair of stakeholders with higher inequality after redistribution at the expense of player A2. Circles below the horizontal line indicate "overredistribution", where player A1 receives a lower income than player A2.

Supporting the previous findings, systematic treatment differences can be inferred from Figure 4 only for left-wing spectators. Two-sided Fisher's exact tests reveal that the fraction of full equality is significantly higher in *Cheat* (lower left panel) than in *Monitor* (upper left panel) if player A1 has a preliminary income between 10 and 7.5, or 6.5 (five comparisons are significant at the 5% and two at the 10% level). This is in line with the result depicted in Figure 3 that left-wing spectators only react to cheating opportunities by implementing a lower inequality when the preliminary income distribution is unequal. In addition, there is a higher fraction of left-wing spectators in *Cheat* than in *Monitor* that always implement full equality independent of preliminary incomes (21.2% vs. 5.7%; $p = 0.079$, Fisher's exact test, two-sided). These two results suggest that the treatment difference for left-wing spectators is driven by an increase in redistribution decisions that result in full equality when cheating opportunities are present.

The analysis of this section shows that left-wing spectators are less willing to accept inequalities when cheating is possible. Looking at their decisions contingent on the income distribution before spectators can redistribute suggests that this is the case because they suspect the stakeholder with higher initial income of cheating. As a consequence, left-wing spectators implement more often a perfectly equal income distribution between the two stakeholders. The interaction between cheating opportunities and being left-wing cannot be explained by differences in their opinion on income inequality. In the next section, I therefore investigate whether different beliefs or norms about cheating across left-wing and right-wing spectators can account for this finding.

3.3 Beliefs and Norms about Cheating

Apart from preferences, there are two alternative explanations for why treatment differences depend on political color. (i) Beliefs about cheating might interact with being left-wing. If in the *Cheat* treatment left-wing spectators believe to a larger extent that stakeholders are cheating than right-wing spectators, this might account for the treatment effect. (ii) Norms about cheating might differ between left-wing and

right-wing spectators. If right-wing spectators find it more acceptable to cheat when there is an opportunity to do so than left-wing spectators, this could also explain the results.

Beliefs about cheating can be inferred from the three different measures of beliefs in the *Cheat* treatment. First, subjects were asked to guess the average reported performance in their own session (*belief-own-treat*) and the average reported performance in the *Monitor* treatment (*belief-other-treat*). Subtracting the latter from the former one indicates the spectator's belief to which extent stakeholders cheated on average. This difference is not significantly different between left-wing (mean = 2.65) and right-wing (mean = 2.02) spectators ($p = 0.389$, Mann-Whitney *U*-test, two-sided). Second, I also compare *belief-own-treat* between the two groups because spectators' answers to the belief about the *Monitor* treatment (*belief-other-treat*) could suffer from self-serving ex-post rationalization of their redistribution decisions. *Belief-own-treat* is not significantly different between left-wing (mean = 10.88) and right-wing spectators (mean = 11.21; $p = 0.666$, Mann-Whitney *U*-test, two-sided). Third, subjects in *Cheat* were asked to guess the fraction of dishonest stakeholders (*belief-frac-cheat*). Again, beliefs do not differ between left-wing (58%) and right-wing spectators (57%; $p = 0.807$, Mann-Whitney *U*-test, two-sided). Thus, although right-wing spectators believe to the same extent as left-wing spectators that stakeholders are cheating, they are not willing to redistribute more in *Cheat* than in *Monitor*.

It has been shown that norms (behavior that people perceive as appropriate) have predictive power for subjects' actual behavior (e.g., Krupka and Weber, 2013). In addition, several studies in the economics literature use actual behavior to identify norms (e.g., Camerer and Fehr, 2004; Fehr and Fischbacher, 2004). Therefore, I use actual cheating behavior of the stakeholders as a proxy for the spectators' norms about cheating. Performances in the *Monitor* treatment show that left-wing and right-wing stakeholders are equally able to work on the matrix task ($p = 0.977$, Mann-Whitney *U*-test, two-sided). While their average performance in *Monitor* is 11.53 and 11.5 tasks respectively, in *Cheat*, left-wing stakeholders report to have solved 14.35 tasks

and right-wing stakeholders 13.08 tasks. This difference between the two groups is not significant ($p = 0.283$, Mann-Whitney U -test, two-sided). If anything, left-wing stakeholders tend to cheat more than right-wing stakeholders. However, there is no significant evidence that the norm about cheating depends on political preferences. In particular, looking at actual behavior suggests that right-wing stakeholders do not find it more acceptable to cheat than left-wing stakeholders and I assume that the same holds true for right-wing and left-wing spectators.¹⁸

Finding no differences in beliefs and norms about cheating between left-wing and right-wing spectators suggests that the difference in their choices reflects a difference in preferences. Right-wing spectators are reluctant to take away money from a stakeholder due to potential cheating if they do not know whether this stakeholder actually cheated. In contrast, left-wing spectators are willing to redistribute more if they believe that someone has cheated — even without being able to detect cheating.

4 Concluding Remarks

The sources of inequality largely influence what people consider to be a fair distribution of income and wealth within a society. Assuming that fiscal imbalances will rise in most western countries due to demographic trends, these redistributive preferences will be particularly relevant for designing welfare policies in the future and thus constitute an important issue in public economics (Kuziemko et al., 2015). In this paper, I focus on cheating as a potential source of unequal outcomes. I find large differences in how to deal with these inequalities depending on political preferences. Supporters of left-wing parties substantially redistribute incomes when cheating is possible, while supporters of right-wing parties refrain from redistribution. As a consequence, cheating opportunities — which receive increasing public attention through recent revelations about fraudulent behavior — amplify the disagreement over redistributive policies between the political left and right.

A deeper look into my data reveals that left-wing spectators' redistribution

¹⁸Since roles were randomly assigned in the experiment, norms of stakeholders and spectators should not systematically differ from each other.

decisions are only affected by potential misreporting in cases of large income differences. This provides strong evidence that left-wing spectators suspect a “rich” stakeholder to be cheating. Furthermore, my results suggest that both beliefs and norms about cheating do not depend on political color. Hence, right-wingers refrain from redistribution although they believe that stakeholders are cheating and although they are themselves reluctant to cheat to the full extent. This shows that right-wing spectators hesitate to redistribute if they do not know for sure that high relative income was acquired by dishonest means, while left-wingers are less concerned about this.

These findings might help to understand the political debate about how to tackle tax evasion and can inform politicians about the consequences of preventing fraudulent behavior. An example for fighting tax evasion are a couple of German federal states which bought tax CDs that contain information about German tax dodgers’ Swiss bank accounts. While the Social Democrats advocate the potentially illegal purchases from whistleblowers, the Conservatives object such measures.¹⁹ One reason for these different strategies might be that tax CDs purchases raises attention to potential cheating, which is, according to my findings, beneficial for left-wing parties with regard to justifying redistributive policies.

Further implications might be drawn concerning the different extent of redistributive policies between Europe and the United States. Europeans prefer substantially more redistribution than U.S. Americans (Almås et al., 2016), which can be partly explained by differences in beliefs about whether luck or effort determine inequalities (Alesina and Angeletos, 2005). As U.S. Americans seem to be far less skeptical towards the very rich, perceived cheating opportunities might also contribute to explain cross-country evidence on redistributive policies. Therefore, exploring how redistributive preferences are affected by potential cheating in the United States as opposed to Europe is a fruitful avenue for further research.

¹⁹See, e.g., <http://www.spiegel.de/international/germany/german-authorities-investigate-ubs-in-relation-to-tax-evasion-a-849366.html>, last accessed on March 5, 2018.

References

- Abeler, J., Nosenzo, D. and Raymond, C. (2016). "Preferences for truth-telling." *CESifo Working Paper Series*, No. 6087.
- Ai, C. and Norton, E. C. (2003). "Interaction terms in logit and probit models." *Economics Letters*, 80(1), 123–129.
- Alesina, A. and Angeletos, G.-M. (2005). "Fairness and redistribution." *American Economic Review*, 95 (4), 960–980.
- Alesina, A. and La Ferrara, E. (2005). "Preferences for redistribution in the land of opportunities." *Journal of Public Economics*, 89(5), 897–931.
- Alesina, A. and Giuliano, P. (2011). "Preferences for redistribution." In *Handbook of Social Economics*, edited by J. Benhabib, A. Bisin and M. O. Jackson, 1, 93–132. Amsterdam: North-Holland.
- Almås, I., A. W. Cappelen and Tungodden, B. (2016). "Cutthroat capitalism versus cuddly socialism: Are Americans more meritocratic and efficiency-seeking than Scandinavians?" *Discussion Paper 18/2016*, NHH Dept. of Economics.
- Bortolotti, S., Soraperra, I., Sutter, M. and Zoller, C. (2017). "Too lucky to be true: Fairness views under the shadow of cheating." *CESifo Working Paper Series*, No. 6563.
- Brandts, J. and Charness, G. (2011). "The strategy versus the direct-response method: A first survey of experimental comparisons." *Experimental Economics*, 14(3), 375–398.
- Camerer, C. F. and Fehr, E. (2004). "Measuring social norms and preferences using experimental games: A guide for social scientists." In *Foundations of human sociality: Economic experiments and ethnographic evidence from fifteen small-scale societies*, edited by J. Henrich, R. Boyd, S. Bowles, C. Camerer, E. Fehr and H. Gintis, 97, 55–95. Oxford University Press.
- Cappelen, A. W., Cappelen, C. and Tungodden, B. (2017). "False positives and false negatives in distributive choices." *Working Paper*, mimeo.

- Cappelen, A. W., Fest, S., Sørensen, E. Ø. and Tungodden, B. (2016). "Choice and personal responsibility: What is a morally relevant choice?" *Discussion Paper* 27/2014, NHH Dept. of Economics.
- Cappelen, A. W., Konow, J., Sørensen, E. Ø. and Tungodden, B. (2013). "Just luck: An experimental study of risk-taking and fairness." *American Economic Review*, 103(4), 1398–1413.
- Cappelen, A. W., Sørensen, E. Ø. and Tungodden, B. (2010). "Responsibility for what? Fairness and individual responsibility." *European Economic Review*, 54(3), 429–441.
- Cohn, A., Fehr, E. and Maréchal, M. A. (2014). "Business culture and dishonesty in the banking industry." *Nature*, 516(7529), 86–89.
- Conrads, J. and Lotz, S. (2015). "The effect of communication channels on dishonest behavior." *Journal of Behavioral and Experimental Economics*, 58, 88–93.
- Fehr, E. and Fischbacher, U. (2004). "Third-party punishment and social norms." *Evolution and Human Behavior*, 25(2), 63–87.
- Fehr, E., Naef, M. and Schmidt, K. M. (2006). "Inequality aversion, efficiency, and maximin preferences in simple distribution experiments: Comment." *American Economic Review*, 96(5), 1912–1917.
- Fehr, E. and Schmidt, K. M. (1999). "A theory of fairness, competition, and cooperation." *Quarterly Journal of Economics*, 114(3), 817–868.
- Fischbacher, U. (2007). "z-Tree: Zurich toolbox for ready-made economic experiments." *Experimental Economics*, 10(2), 171–178.
- Fischbacher, U. and Föllmi-Heusi, F. (2013). "Lies in disguise — an experimental study on cheating." *Journal of the European Economic Association*, 11(3), 525–547.
- Fong, C. (2001). "Social preferences, self-interest, and the demand for redistribution." *Journal of Public Economics*, 82(2), 225–246.

- Frohlich, N., Oppenheimer, J., Bond, P. and Boschman, I. (1984). "Beyond economic man: Altruism, egalitarianism, and difference maximizing." *Journal of Conflict Resolution*, 28(1), 3–24.
- Gneezy, U. (2005). "Deception: The role of consequences." *American Economic Review*, 95(1), 384–394.
- Greiner, B. (2015). "Subject pool recruitment procedures: Organizing experiments with ORSEE." *Journal of the Economic Science Association*, 1(1), 114–125.
- Houser, D., List, J. A., Piovesan, M., Samek, A. and Winter, J. (2016). "Dishonesty: From parents to children." *European Economic Review*, 82, 242–254.
- Jiang, T. (2013). "Cheating in mind games: The subtlety of rules matters." *Journal of Economic Behavior & Organization*, 93, 328–336.
- Kirchler, E., Maciejovsky, B. and Schneider, F. (2003). "Everyday representations of tax avoidance, tax evasion, and tax flight: Do legal differences matter?" *Journal of Economic Psychology*, 24(4), 535–553.
- Konow, J. (2000). "Fair shares: Accountability and cognitive dissonance in allocation decisions." *American Economic Review*, 90(4), 1072–1091.
- Krupka, E. L. and Weber, R. A. (2013). "Identifying social norms using coordination games: Why does dictator game sharing vary?" *Journal of the European Economic Association*, 11(3), 495–524.
- Kuziemko, I., M. I. Norton, E. Saez and Stantcheva, S. (2015). "How elastic are preferences for redistribution? Evidence from randomized survey experiments." *American Economic Review*, 105(4), 1478–1508.
- Mazar, N., Amir, O. and Ariely, D. (2008). "The dishonesty of honest people: A theory of self-concept maintenance." *Journal of Marketing Research*, 45(6), 633–644.
- Möllerström, J., Reme, B.-A. and Sørensen, E. Ø. (2015). "Luck, choice and responsibility — an experimental study of fairness views." *Journal of Public Economics*, 131, 33–40.

- Pigors, M. and Rockenbach, B. (2016). "The competitive advantage of honesty." *European Economic Review*, 89, 407–424.
- Piketty, T. (1995). "Social mobility and redistributive politics." *Quarterly Journal of Economics*, 110(3), 551–584.
- Rabin, M. (1993). "Incorporating fairness into game theory and economics." *American Economic Review*, 83(5), 1281–1302.
- Shalvi, S. and De Dreu, C. K. (2014). "Oxytocin promotes group-serving dishonesty." *Proceedings of the National Academy of Sciences*, 111(15), 5503–5507.
- Van Lange, P. A., Bekkers, R., Chirumbolo, A. and Leone, L. (2012). "Are conservatives less likely to be prosocial than liberals? From games to ideology, political preferences and voting." *European Journal of Personality*, 26(5), 461–473.
- Zucman, G. (2014). "Taxing across borders: Tracking personal wealth and corporate profits." *Journal of Economic Perspectives*, 28(4), 121–148.

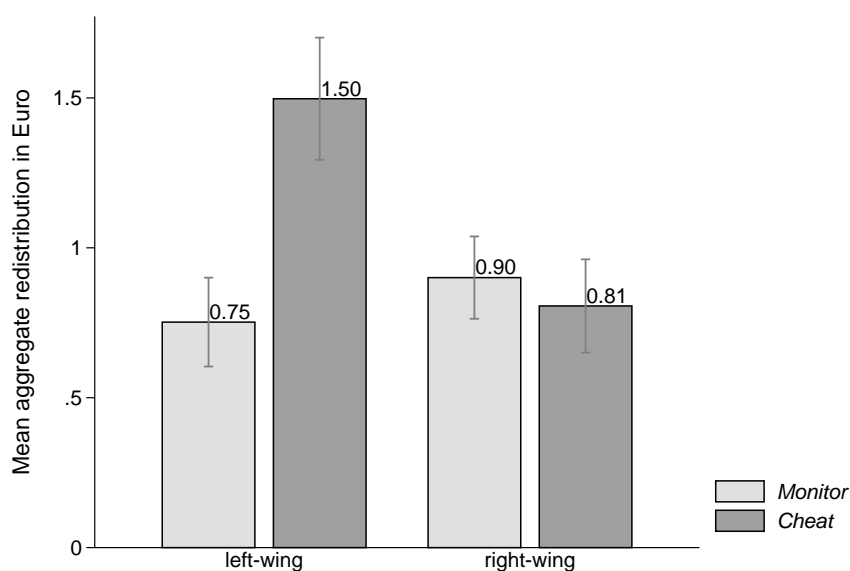
Appendix

A Supplementary Results

Table A.1: Voting behavior and political left-right score of spectators & 2017 German federal election results

Political Party	Experiment (in %)	2017 federal election (in %)	left-right scale
CDU/CSU	28.57	25.02	5.73
SPD	14.84	15.58	4.04
Die Grünen	14.29	7.02	4.08
Die Linke	6.04	6.79	3.45
AFD	3.30	8.17	7.67
FDP	9.89	9.60	5.56
Die Piraten	2.20	0.28	3.75
Other party	6.04	3.51	4.73
Would not go to the election	14.84	24.03	4.30

Notes: The parties are the Christian Democrats (CDU/CSU), the Social Democrats (SPD), the Green Party (Die Grünen), the socialist party (Die Linke), the right-wing populist Alternative for Germany (AFD), the libertarian party (FDP), and the Pirate Party Germany (Die Piraten). Results of the 2017 German federal election are based on the “second vote” and calculated without rejected votes. The scale for political attitudes ranges from 1 (left) to 10 (right).



Notes: The figure shows aggregate redistribution in € defined as the average redistribution of all eleven redistribution decisions. Error bars indicate standard errors of the mean.

Figure A.1: Aggregate redistribution

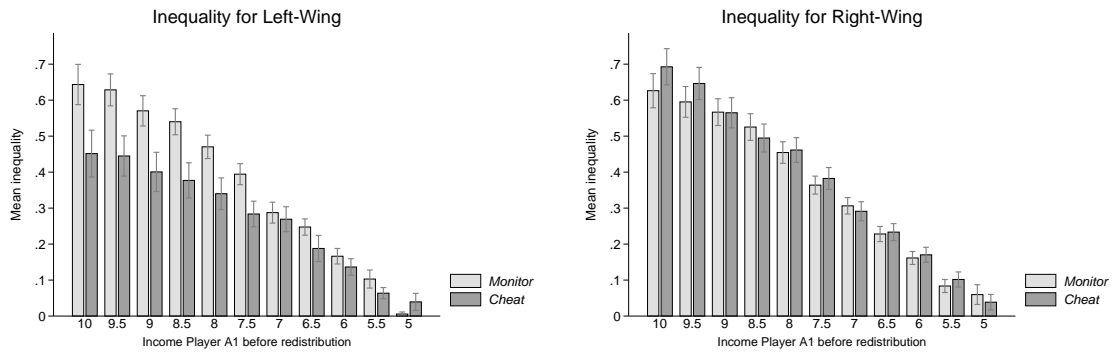
Figure A.1 is equivalent to Figure 2 for redistribution instead of inequality. As for inequality, aggregate redistribution is defined as the average over all eleven decisions of the spectator.²⁰ Left-wing spectators react to the treatment even stronger when looking at redistribution instead of inequality. They redistribute twice as much in the *Cheat* treatment than in the *Monitor* treatment, which is highly significant ($p < 0.01$, Mann-Whitney *U*-test, two-sided). There is no significant effect of cheating opportunities for right-wing spectators ($p = 0.433$, Mann-Whitney *U*-test, two-sided). Regression results for aggregate redistribution can be found in Table A.2 and are in line with the findings on inequality in Table 1.

Table A.2: Aggregate redistribution

Dependent variable	Aggregate redistribution in €					
	(1)	(2)	(3)	(4)	(5)	(6)
Cheat	0.215 (0.142)	0.745*** (0.246)	0.737*** (0.240)	0.758*** (0.252)	0.478* (0.265)	0.795*** (0.260)
Right-wing		0.148 (0.236)	0.233 (0.237)	0.247 (0.231)	0.010 (0.320)	0.210 (0.300)
Cheat × Right-wing		-0.840*** (0.307)	-0.818*** (0.284)	-0.820*** (0.280)	-0.593 (0.374)	-0.930*** (0.341)
We need inequality (1 = no, 10 = yes)			-0.065** (0.029)	-0.064** (0.028)	-0.052 (0.034)	-0.077* (0.045)
Constant	0.844*** (0.078)	0.752*** (0.165)	0.983*** (0.176)	0.800* (0.478)	0.959** (0.457)	1.268** (0.520)
Additional controls	No	No	No	Yes	Yes	Yes
Observations	182	182	182	182	182	144
Log likelihood	-274.643	-270.166	-268.661	-264.546	-265.249	-202.152

Notes: Two-limit Tobit regressions regressions on aggregate redistribution (lower limit: -82.5, upper limit: 82.5). Columns (4) to (6) include a binary variable for whether the experiment was conducted in 2016 or 2017 (insignificant in all specifications) and additional covariates from the questionnaire: age, gender, semester, and number of experiments so far. Column (5) includes the categories “other party” and nonvoters in the definition of being left-wing. Column (6) excludes spectators who would vote for “other party” as well as nonvoters. Robust and clustered (on session level) standard errors in parentheses. Stars indicate significance on the levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

²⁰Decisions in which the spectator redistributes away from player A2 are also taken into account. Consequently, I use the absolute values of money redistributed to calculate the aggregate redistribution of one spectator.



Notes: The figure shows inequality separately for each level of preliminary income of player A1. Error bars indicate standard errors of the mean.

Figure A.2: Inequality by preliminary income

Figure A.2 shows inequality depending on preliminary income distribution and treatment for left-wing (left panel) and right-wing spectators (right panel). The two very left and very right bars represent inequality when the income distribution before redistribution is (10,0) and (5,5) respectively. Inequality is not significantly different between left-wing and right-wing spectators in any of the redistribution decisions in the *Monitor* treatment (Mann-Whitney U -tests, two-sided).

B The *Mixed* Treatment

I conducted an additional treatment to find out how unequal cheating opportunities affect inequality acceptance. This is motivated by “real-world” examples where only some people are able to cheat, while it is much harder or even impossible for others to report untruthfully. Consider, for instance, the cases of tax evasion, doping in sports, and faking educational achievements. Only individuals who are subject to tax can evade taxes, and only the “rich” might have the means to do this large-scale. Only professional athletes have access to well known doping doctors. And only people belonging to the higher education system have the opportunity to plagiarize a PhD thesis. Therefore, I implement a treatment called *Mixed*, where only one of the two stakeholders is able to cheat. While in the *Cheat* treatment it was unclear whether both or only one of the two stakeholders cheated, even in cases with high preliminary inequality, it is clearer that the stakeholder with a much higher income is presumably

cheating in the *Mixed* treatment. Furthermore, spectators might perceive it as unfair that only one of the two stakeholders has the opportunity to cheat. For these reasons, inequality should be even more reduced in *Mixed* than in *Cheat*.

Procedural Details. Eight sessions of the *Mixed* treatment were conducted in May and June 2017 with a total of 172 subjects. Subjects earned €12.94 on average, including a €5 show-up fee. The exercise sheet was collected only from one of the two stakeholders. After working on the matrix task, it was publicly announced that subjects will next be informed about their role and whether their exercise sheet will be collected (see Appendix D.2). Hence, all participants, including the spectators, knew that the stakeholders were aware of their cheating opportunities before stating their performance and that only one of the two stakeholders could cheat.

Results. In the *Mixed* treatment, reported performance of subjects that were monitored (mean of 11 tasks) do not significantly differ from those that were able to cheat (12.63; $p = 0.158$, Mann-Whitney U -test, two-sided). Since reported performance of the monitored stakeholders in *Mixed* is generated under the exactly same conditions as of stakeholders in *Monitor*, these observations can be pooled and compared with potential cheaters in *Mixed*. Again, performances do not significantly differ ($p = 0.176$, Mann-Whitney U -test, two-sided).²¹ In line with stakeholders' actual performances, beliefs of spectators about reported performance in their own session (*belief-own-treat*) do not significantly differ between *Monitor* (9.35 on average) and *Mixed* (9.31; $p = 0.832$, Mann-Whitney U -test, two-sided). In addition, asking spectators in *Mixed* about the average performance of stakeholders in the *Cheat* treatment (*belief-other-treat* in the *Mixed* treatment) reveals that spectators in *Mixed* believe reported performance to be higher in *Cheat* (12.87) than in *Mixed* ($p < 0.0001$, Wilcoxon signed rank test, two-sided). Moreover, spectators' beliefs about the ratio of cheaters to those stakeholders who are able to cheat (*belief-frac-cheat*) are higher in *Cheat* (57%) than in *Mixed* (50%; $p = 0.036$, Mann-Whitney U -test, two-sided).²² Hence, spectators seem to

²¹Based on these observations (135 subjects that cannot cheat with a mean of 11.35 tasks solved as baseline and 43 potential cheaters), I calculate the minimal detectable difference of a two-sided Mann-Whitney U -test for the 5% significance level. A treatment difference of at least 2.25 tasks is detected with a statistical power of 80%. Thus, the sample is large enough to detect treatment differences in performance of a similar size as implied by the difference between *Monitor* and *Cheat* (2.29 tasks).

²²*Belief-frac-cheat* is not incentivized and likely to be overstated due to demand effects by explicitly

anticipate stakeholders (missing) dishonest behavior as there is no clear evidence that spectators in *Mixed* believe stakeholders to be cheating.

As a consequence of failed treatment manipulation in the *Mixed* treatment, implemented aggregate inequality does neither differ for right-wing ($p = 0.467$) nor left-wing spectators ($p = 0.670$) between *Monitor* and *Mixed* (Mann-Whitney *U*-tests, two-sided). In summary, the null hypothesis of equal performance between stakeholders that are monitored and those who are not cannot be rejected. The low or nonexistent occurrence of cheating might be a result of stakeholders finding it unfair that only one of them can cheat. When designing this treatment, it was difficult to predict this finding since this is, to the best of my knowledge, the first treatment where only one of two subjects who are otherwise in exactly the same position can cheat.²³ In line with this, spectators in *Mixed* seem to anticipate stakeholders' behavior as I find no evidence that they believe stakeholders to be cheating. Given this, it is not surprising that I do not find a treatment effect in *Mixed*.

asking for the fraction of cheaters. Therefore, I refrain from interpreting the size of this belief but only compare the difference across treatments.

²³This conclusion is drawn from comparing the *Mixed* treatment to treatments of 72 papers analyzed in a meta-study by Abeler et al. (2016).

C Screenshots of Decision Screens

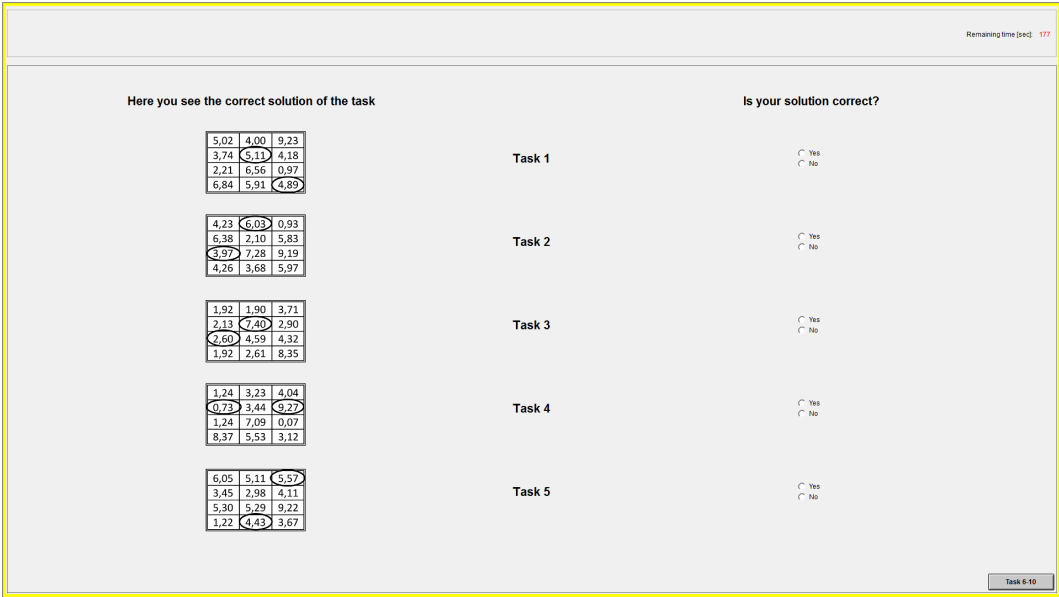


Figure C.1: Screenshot of one of the stakeholder’s decision screens

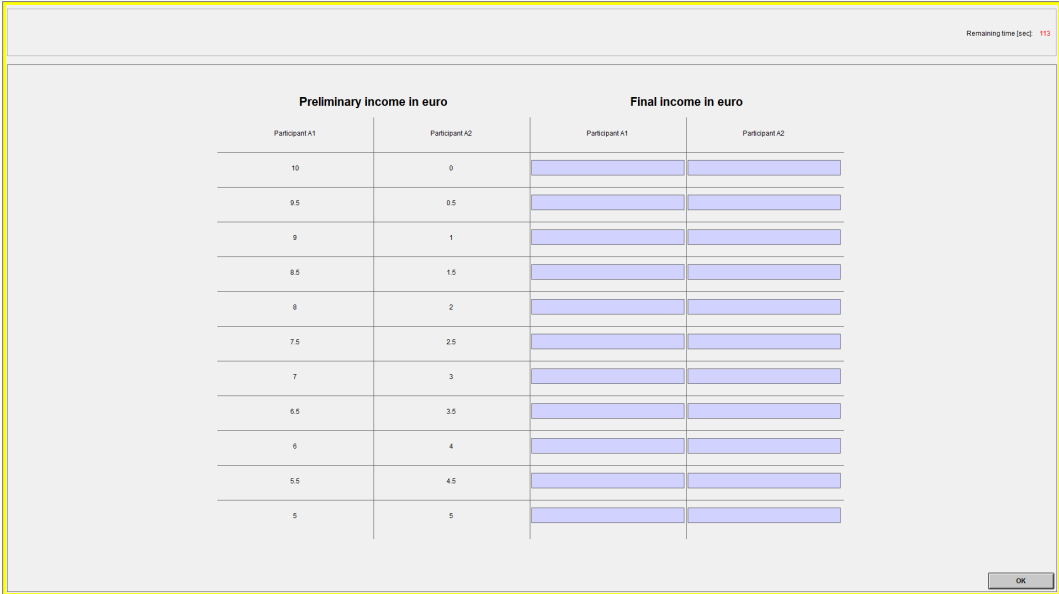


Figure C.2: Screenshot of the spectator’s decision screen

D Instructions

D.1 General Instructions at the Beginning of the Experiment

[In paper form]²⁴

Welcome to the experiment and thank you for your participation!

Please do not talk to other participants of the experiment from now on.

General information on the procedure

This experiment serves to investigate economic decision making behavior. You can earn money, which will be paid to you individually and in cash after the experiment has ended.

If you have any questions after reading these instructions or during the experiment, please raise your hand or press the red button on your keyboard. We will then come to you and answer your question in private.

During the experiment, you and the other participants will make decisions. Your own decisions as well as the decisions of other participants can determine your payoffs. These payoffs are determined according to the rules which are explained in the following.

Payment

At the end of the experiment, you will receive in cash the money that you have earned during the experiment and additional 5 euro for showing up in time. Therefore, we will call every participant based on his seat number, i.e., none of the other participants gets to know your payment, and also you will not get to know the payments of other participants.

Anonymity

Data from this experiment will be analyzed anonymously, i.e., we will never link your

²⁴The instructions were translated from German. The original version is available upon request.

name to the data of the experiment. At the end of the experiment, you have to sign a receipt confirming that you received your payment. This receipt serves accounting purposes only.

Assignment of Roles

There are two different roles in this experiment: A and B. The role of each participant is determined randomly. Whether you are participant A or B will be communicated to you at a later point in time on your screen.

Your Task

At the beginning of the experiment, we will hand out an exercise sheet, which we will place upside down on your desk. Please turn the sheet over only when you are asked to do so. There are 20 tasks on the exercise, which are numbered top down. It does not matter on which task you work first.

Each task consists of a box containing 12 numbers. Here is an example:

1,69	1,82	2,91
4,67	3,81	3,05
5,82	5,06	4,28
6,36	6,19	4,57

Example

Only two numbers in the box add up to 10.00. It is your task to find these two numbers and to circle them. In the following, you see the correct solution for the example.

1,69	1,82	2,91
4,67	3,81	3,05
5,82	5,06	4,28
6,36	6,19	4,57

Correct solution of the example

You have 6 minutes to work on the tasks. For your guidance, a clock will display the remaining time on your screen. After the 6 minutes expired, please put down your pen. Subsequently, we will collect your pens.

Income of Participant A

After timeout, participant A compares the numbers which he marked to the solution. You will receive more information on this after working on the task. Participant A receives 1 point for every correct solution.

Two participants A are randomly assigned to each other in order to determine their incomes. 10 euro will be split up among these two participants. Income is proportional to the number of points achieved in the preceding task and rounded to 50 cents.

Example 1: One participant A achieved 6 points and the other participant A 4 points. Thus, both participants A have achieved 10 points in total. The participant A with 6 points therefore receives an income of 6.00 euro ($(6 \text{ points} / 10 \text{ points}) \times 10 \text{ euro} = 6 \text{ euro}$). The participant A with 4 points receives an income of 4 Euro ($(4 \text{ points} / 10 \text{ points}) \times 10 \text{ euro} = 4.00 \text{ euro}$).

Example 2: One participant A achieved 12 points and the other participant A 7 points. Thus, both participants A have achieved 19 points in total. The participant A with 12 points therefore receives an income of 6.50 euro ($(12 \text{ points} / 19 \text{ points}) \times 10 \text{ euro} = 6.32$, rounded to 50 cents). The participant A with 7 points receives an income of 3.50 euro ($(7 \text{ points} / 19 \text{ points}) \times 10 \text{ euro} = 3.68 \text{ euro}$, rounded to 50 cents).

Income of Participant B

Participant B can distribute the incomes which participants A earned. Participants B will receive detailed information about this on their screen later on.

Further Procedures

We will soon hand out an exercise sheet to each of you. Please leave this sheet upside down until we announce the beginning of the task. After you have finished the task, the computer determines whether you are participant A or B. You will be informed about this on your screen. The assignment of roles is random. During the experiment, you will receive further information on your screen.

D.2 Treatment Variation after the Matrix Task

[On the screens of the participants]

Entering of Solutions

All participants have now worked on the task.

Participants A will soon have 3 minutes to compare the numbers which he marked to the solution as follows:

On the left hand side of the screen, participant A sees the correct solution of the tasks.

On the right hand side of the screen, participant A should indicate whether the correct solution corresponds to the numbers that he marked on his exercise sheet. After timeout, participants A automatically proceed to the next screen and can no longer compare solutions.

On the next screen, you will see an example of the screen on which participant A compares his solutions with the correct solutions. This screen will be displayed to you for 15 seconds. You do neither have to indicate something nor click on OK. The tasks on the following screen are examples. Please click now on OK.

[new screen]

[*Monitor* treatment only]

Important: After all participants A have compared their solutions to the correct solutions, we collect the exercise sheets of all participants. We then verify that participant A did not make any mistake when comparing his solutions to the correct ones. If participant A made a mistake, we will correct the number of points of participant A. You will soon be informed whether you are participant A or B. Please click now on OK.

[*Cheat* treatment only]

Important: You will receive your payment in the room next door. There is also a shredder in this room. At the end of the experiment, you will shred your exercise sheet and afterwards receive your payment. This ensures that we cannot trace back

your solutions. You will soon be informed whether you are participant A or B. Please click now on OK.

[*Mixed* treatment only, see Appendix B]

Important: After all participants A have compared their solutions to the correct solutions, we collect the exercise sheets of **one** of the two matched participants A. We then verify that this participant A did not make any mistake when comparing his solutions to the correct ones. If this participant A made a mistake, we will correct the number of points of this participant A. We **do not** collect the exercise sheet of the other participant A. You will receive your payment in the room next door. There is also a shredder in this room. If we do not collect your exercise sheet, please put it in the envelope which you find on your desk. **All** participants have to seal their envelopes and shred it before receiving their payoff. This ensures that we cannot trace back the solutions of the participants whose exercise sheets we do not collect. You will soon be informed whether you are participant A or B, and whether we will collect your exercise sheet. Please click now on OK.

D.3 Instructions for Participant B

[On screens of participants B while participants A compare their solutions to correct solutions; not read aloud by experimenter]

Your Decisions

As participant B you will be randomly assigned to 2 participants A among whom 10 euro will be split up as described in the instructions. We call this income, which they receive for working on the task, *preliminary income* of participants A. You will now determine the *final income* of participants A for *every possible* distribution of the preliminary incomes.

You see the table where you will enter the final incomes further below. [table without possibility to enter something is shown at the bottom of the screen] Final incomes of participants A, which you determine, must add up to 10 euro. You may enter final

incomes in 10 cents steps. In order to assign final incomes unambiguously, we call the two participants A “participant A1” and “participant A2”.

After the two participants A compared their solutions to the correct solutions, preliminary incomes will be determined. This income then corresponds to one row in the table: e.g., 7 euro for participant A1 and 3 euro for participant A2. Participants A will receive the final incomes that you enter in the same row on the right side of the table (e.g., next to 7 euro for participant A1 and 3 euro for participant A2). Hence, each of your decisions can be decisive for the payoffs of the participants A!

[new screen]

Implementation of Your Decision

At the end of the experiment, the computer will randomly determine whether your decision or the decision of another participant B will be implemented. The probability for your decision to be implemented is 50%.

Your Income

As participant B you receive a fixed income of 10 euro independent of your decision.

You will make your decisions on the next screen.

D.4 Incentivized Belief Elicitation

[On screens of participants; not read aloud by experimenter; text in *Monitor*]

Assessment

You will provide two assessments in the following. You will receive details hereto on the next two screens. You are paid for the accuracy of your assessments.

However, only one of the two assessments is paid. At the end of the experiment, the computer will randomly determine which of the two assessments is paid. The probability that assessment 1 or assessment 2 is paid is 50% respectively.

[new screen]

Assessment 1

All participants A now have compared their solutions to the correct ones. Please assess how many points participants A achieved on average in the task at the beginning of the experiment.

You are paid for the accuracy of your assessment. If your assessment deviates less than 1 point from the actual average, you will receive **2 euro** additional to your remaining income from the experiment. If your assessment deviates at least 1 point but less than 2 points from the actual average, you will receive **1 euro**. Larger deviations are not paid.

How many points did the participants A achieve on average? (integers only)

[new screen]

Assessment 2

We now ask you to provide an assessment for a similar experiment. In the other experiment, participants worked on exactly the same task on which you have worked in the current experiment.

Important

In contrast to the current experiment, exercise sheets were not collected and it was not verified that the solutions had been compared without making a mistake. Participants were aware of this before comparing the solutions.

Please assess how many points participants A achieved on average in the other experiment.

As before, you are paid for the accuracy of your assessment. If your assessment deviates less than 1 point from the actual average, you will receive **2 euro** additional to your remaining income from the experiment. If your assessment deviates at least 1 point but less than 2 points from the actual average, you will receive **1 euro**. Larger deviations are not paid.

How many points did the participants A achieve on average in the other experiment? (integers only)