NHH

INSTITUTT FOR SAMFUNNSØKONOMI

DEPARTMENT OF ECONOMICS

SAM 18 2018

ISSN: 0804-6824

August 2018

Discussion paper

Second-best fairness under limited information: The trade-off between false positives and false negatives

BY

Alexander W. Cappelen, Cornelius Cappelen AND **Bertil Tungodden**

Norges Handelshøyskole

NORWEGIAN SCHOOL OF ECONOMICS

Second-best fairness under limited information: The trade-off between false positives and false negatives *

Alexander W. Cappelen Cornelius Cappelen

Bertil Tungodden†

August 28, 2018

Abstract

In many important economic settings, limited information makes it impossible for decision makers to ensure that each individual gets what he or she deserves. Decision makers are then faced with the trade-off between giving some individuals more than they deserve, *false positives*, and giving some individuals less than they deserve, *false negatives*. We present the results from a large-scale experimental study of how people trade off these two mistakes in distributive choices. We find that a majority are more concerned with avoiding false negatives than with avoiding false positives, but we also document heterogeneity with respect to how people make this trade-off. The findings shed important light on people's attitudes to a wide range of policies by providing novel evidence on an important dimension of people's social preference.
(JEL D63)

*We have received very valuable comments and suggestions from Björn Bartling, Tore Ellingsen, Uri Gneezy, Rune Jansen Hagen, Stein Kuhnle, Axel Sonntag and Jean-Robert Tyran. We would also like to thank Sebastian Fest, Ingar Haaland and Ida Kjørholt for excellent research assistance. The project was financed by support from the Research Council of Norway through its Centres of Excellence Scheme, FAIR project no 262675, the Research Council of Norway research grant 236995, and the NORFACE Welfare State Future Program. The project was administered by FAIR - The Choice Lab, NHH Norwegian School of Economics.

†Alexander W. Cappelen (corresponding author): Norwegian School of Economics, Helleveien 30, 5045 Bergen, Norway (phone: +47 93630179, email: alexander.cappelen@nhh.no); Cornelius Cappelen: Bergen University (email: cornelius.cappelen@uib.no); Bertil Tungodden: Norwegian School of Economics (email: bertil.tungodden@nhh.no).

1 Introduction

Limited information creates second-best situations where not all optimality conditions can be satisfied (Lipsey and Lancaster, 1956; Mirrlees, 1971). An important set of such situations are those where the deservingness of an individual is conditional on certain unobserved characteristics. Decision makers must then rely upon indirect indicators of deservingness, such as group membership, which correlate only imperfectly with those characteristics that constitute the real grounds of deservingness (Atkinson and Stiglitz, 1976; Goodin, 1985). As a result, decision makers must trade off giving some individuals more than they deserve, *false positives*, against giving others less than they deserve, *false negatives*. This type of trade-off is present in all settings where people are rewarded or punished under limited information, and represents a fundamental challenge in the design and implementation of policies in the public and the private sector.

The question of how the trade-off between false positives and false negatives should be handled is a key issue in the design of public policy. This is seen, for example, in the debate regarding the social welfare system and the risk of undeserving claimants receiving benefits and in the debate on whether the government should legislate stricter immigration policies to prevent undeserving individuals from receiving residency. There seems to be a clear political divide in these debates, with people on the right of the political spectrum being more concerned with the problem of false positives than people on the left, who are more concerned with the problem of false negatives (Blake, 2012; Stateline, 2017). Such differences could reflect different beliefs about the prevalence of false positives and false negatives, but they could also reflect that there are systematic differences in the relative importance people attach to each of these two mistakes.

The trade-off between false positives and false negatives is also important for the implementation of policies when these policies leave room for discretion. Decision makers may differ in how concerned they are about avoiding false positives relative to avoiding false negatives and might therefore make different decisions even when implementing the same policies. To illustrate, judges often have limited information when deciding (after appeal) whether to grant disability benefits to an applicant. They can then make false positives by granting benefits to undeserving applicants and false negatives by not granting benefits to deserving applicants (Gruber, 2000). It turns out that the likelihood of an applicant being granted disability benefits in such cases depends on the judge (randomly) assigned to the case (Autor, Kostol, and Mogstad, 2015), which is suggestive of judges differing in how they trade off false positives and false negatives. Similar trade-offs between false positives and false negatives are found in a wide range of situations, from the implementation of government immigration policies to the implementation of company bonus policies.

We present the results from a large-scale experiment involving 4000 participants, which studies how people make trade-offs between false positives and false negatives in distributional choices. In the experiment, people act as third-party spectators and

determine payments between two groups of workers. In the first group, all workers have done the assignment they were recruited to do, but in the second group, a number of workers have falsely reported to have done the assignment, i.e. they are "cheaters".

Each spectator is randomized into one of five treatments, which vary with respect to the number of cheaters in the second group. Importantly, to capture the idea of limited information, the spectators have to treat members of the same group equally; in other words, they can only condition payment on group membership. Specifically, they can choose to give all the money to the first group of workers or distribute the money equally between the two groups. If they equalize payments between the two groups, they would implement a number of false positives by paying the cheaters in the second group. However, if they give all the money to the first group, they implement a number of false negatives by not paying those individuals in the second group who had done the assignment.

By varying the number of cheaters in the second group across treatments, we can study the causal effect of the number of cheaters on the distributive behavior of the spectators. There are two key features in this design that are important for the interpretation of the results. First, the spectators know the number of cheaters in the second group, which means that we control their beliefs about the prevalence of cheaters. Second, the spectators agree that cheaters do not deserve payment. Given these features, we can identify how the spectators trade off false positives and false negatives in their preferences.

The workers were recruited from an online labor market platform, while the spectators were recruited from the US and Norway. We recruited 1000 spectators from each country who were nationally representative on a set of observable characteristics. The two countries differ significantly with respect to redistributive policies and the design of welfare systems (Barth, Moene, and Willumsen, 2014; Landersø and Heckman, 2016). A possible explanation for these differences may be that Americans are more concerned than Norwegians with not providing support to individuals who are viewed as undeserving. It is therefore of great interest to examine whether Americans and Norwegians systematically differ in their preferences for false positives and false negatives. The present experimental design allows us to do this by studying the distributive decisions of Americans and Norwegians in a setting where they have the same beliefs about the prevalence and deservingness of the cheaters.

The experiment provides three main findings. First, we find that the willingness to equalize is significantly reduced by an increase in the number of cheaters. The average marginal effect of adding one cheater to the second group is a 17.2 percentage point reduction in the share of spectators who choose equal payment for the two groups. Second, we find strong evidence of spectators having heterogeneous and asymmetric preferences for avoiding false positives and false negatives. In both the US and Norway, the majority are strictly false-negative averse and only a small minority are strictly false-positive averse. Finally, we find a significant association between political affiliation and how the spectators trade off false positives and false negatives, with right-wing participants being more concerned with the number of cheaters in the

second group and therefore less false negative averse.

Our study provides novel evidence on an important dimension of people's social preferences that has not yet been systematically explored in the literature. Economic experiments studying social preferences have typically focused on situations with complete information, where they have studied what people consider to be the fair distribution and how people trade off self-interest with fairness (Rabin, 1993; Fehr and Schmidt, 1999; Konow, 2000; Charness and Rabin, 2002; Bolton and Ockenfels, 2006; Cappelen, Drange Hole, Sørensen, and Tungodden, 2007; Balafoutas, Kocher, Putterman, and Sutter, 2013; Cappelen, Konow, Sørensen, and Tungodden, 2013a; Bartling, Weber, and Yao, 2015). This literature has established that people may differ both in their level of selfishness and in their fariness view. The present study extends this literature by considering situations in which the decision maker has limited information and must make a trade-off between giving some individuals more than they deserve and giving some individuals less than they deserve. We show that people have different views about how this trade-off should be made, even in a setting where they agree on the deservingness criteria. A full characterization of people's social preferences, must therefore include a description of the relative weight they attach to avoiding false positives and false negatives. A recent paper by Bortolotti et al (Bortolotti, Soraperra, Sutter, and Zoller, 2017), which examines how the possibility of cheating affects distributive behavior in a setting where the participants self-report the outcome of a gamble, is the paper most closely related to the our paper. In their design, however, decision makers do not know whether there has been any cheating and they do not aim to measure the relative weight attached to false negatives and false positives.

Our results also contribute to the literature that examines the association between redistributive preferences and political affiliation (Bénabou and Ok, 2001; Bénabou and Tirole, 2006; Alesina and Angeletos, 2005; Alesina and Giuliano, 2011). We show that right-wing spectators place less weight on avoiding false negatives relative to avoiding false positives. Furthermore, we find that the distributional choice made in the experiment is strongly associated with the spectators' general attitude towards income redistribution. Our results thus suggest that preferences for how to make the trade-off between giving someone more than they deserve and giving someone less than their fair share is of fundamental importance for understanding disagreements about redistributive policies.

Finally, our study contributes to the large literature on international differences in attitudes towards inequality and redistribution (Alesina and Glaeser, 2004; Cappelen, Moene, Sørensen, and Tungodden, 2013b; Falk, Becker, Dohmen, Enke, Huffman, and Sunde, 2015; Henrich, Boyd, Bowles, Camerer, Fehr, Gintis, and McElreath, 2001; Jakiela, 2015; Luttmer and Singhal, 2011; Svallfors, 1997). To our knowledge, we provide the first comparative study of how people in different societies handle distributive situations with limited information. We do not find large differences in the behavior of the US and the Norwegian samples, which is interesting given the well-documented differences between the two countries with respect to redistributive policies and fairness views (Almås, Cappelen, and Tungodden, 2016). Our findings

thus suggest that international differences in attitudes towards inequality and redistribution do not mainly reflect differences in how people trade off false negatives and false positives.

The paper is organized as follows: Section 2 describes the experimental design, Section 3 provides a simple social preference model to guide the analysis, and Section 4 outlines the empirical strategy. Section 5 reports the results and Section 6 concludes.

2 Design

We recruited two types of participants in the experiment; *workers* and *spectators*.[1] The workers were asked to complete an assignment that required them to work continuously for 15 minutes. However, they could choose to do the assignment or to falsely report to have done it.

Before the workers decided whether to do the assignment, they were informed that they could receive a bonus payment in addition to their participation fee and that this bonus payment would be determined by a randomly selected third party. Importantly, they were told that this third party would sometimes be unable to distinguish between those who had done the assignment and those who had falsely reported to have done it. The workers thus knew that even if they did not to do the assignment they might get the same bonus payment as someone who did do the assignment.

Each spectator was randomly matched with two groups of four workers and asked to make a decision that could determine the final payment to the workers. In the first group, all workers had done the assignment (non-cheaters), but in the second group, there could be workers who had falsely reported to have done it (cheaters). The spectators were informed about the instructions given to the workers and the number of cheaters in the second group, but, importantly, they had to treat everyone in the second group equally.

The spectators could choose to distribute 16 USD in one of two ways: either distribute the money equally between the two groups (*equalize*), in which case all eight workers would each be paid 2 USD, or give all the money to the first group (*not equalize*), in which case, each worker in the first group would be paid 4 USD and each worker in the second group would be paid nothing (regardless of whether they were cheaters or not).

After the spectators made their distributive choice, they completed a non-incentivized survey that included questions about their political affiliation and attitude to redistribution, as well as standard background questions about gender, age, pre-tax household income, and education. Table 1 summarizes the main stages in the experiment.

[Table 1 about here]

[1] The complete instructions for both workers and spectators are provided in the Appendix.

2.1 Treatments

The spectators were randomly assigned to one of five treatments that differed only in the number of cheaters in the second group, which varied from zero to four. In the treatment without cheaters, the spectators were informed that all workers in the second group had done the assignment. By contrast, in the treatment with four cheaters, they were informed that none of the workers in the second group had done the assignment. In these two treatments, spectators who held the view that only individuals who did the assignment deserved to get paid, could satisfy all optimality conditions in terms of deservingness, i.e., they could give all workers what they deserved. These two treatments therefore allow us to establish whether the spectators viewed it as fair that all non-cheaters were paid the same bonus and whether they viewed it as fair that the cheaters were not paid a bonus.

The three intermediate treatments, with one, two or three cheaters in the second group, are second-best distributive situations where the spectators had to treat the deserving and undeserving individuals in the second group in the same way. They thus had to choose between implementing false positives or false negatives:

False positive: A cheater is paid a bonus.

False negative: A non-cheater is not paid a bonus.

If the spectators chose to equalize, the number of false positives was given by the number of cheaters in the second group, C; if they chose not to equalize, the number of false negatives were given by the number of non-cheaters in the second group, $4-C$. In describing the alternatives to the spectators, we highlighted this trade-off. To illustrate, in the treatment with two cheaters in the second group, the two alternatives were presented as follows:

Alternative A: Give 4 USD to four of the individuals who did the assignment and nothing to the other four individuals. This means that two individuals who did the assignment are not paid.

Alternative B: Give 2 USD to each of the eight individuals. This means that the two individuals who falsely reported to have done the assignment are paid.

In this treatment, if the spectators choose to equalize (Alternative B), they reveal a preference for implementing two false positives rather than two false negatives. An additional cheater in the second group makes equalization less attractive by increasing the number of false positives, and also makes not equalizing more attractive by decreasing the number of false negatives. By comparing the share of spectators who equalized in the different treatments we can study how the number of cheaters in the second group causally affects the spectators' willingness to equalize the bonus.

Three features of the design should be highlighted. First, the distribution of payments associated with the two alternatives is the same in all treatments, which means

that inequality aversion cannot explain any treatment effects. Second, there are no efficiency costs associated with choosing to equalize payments, which means that efficiency considerations cannot explain any treatment effects. Finally, the design rules out any differences in beliefs about the number of cheaters in the distributive situation. Hence, if the spectators view non-cheaters as deserving of a bonus and cheaters as non-deserving, the treatment effects allow us to identify how the spectators trade off false positives and false negatives.

2.2 Sample

The spectators in the experiment, 1000 from the US and 1000 from Norway, were recruited through a leading data-collection agency in each of the countries.[2] The two countries were selected because they represent the extremes among the OECD countries with respect to income inequality, with Norway being characterized by a much more compressed income distribution than the US (Atkinson, Piketty, and Saez, 2011). The US and Norway also differ dramatically with respect to redistributive policies, with Norway having a significantly higher tax level, a more generous welfare state, and more income mobility than the US (Barth et al., 2014; Landersø and Heckman, 2016). By comparing the distributive choices of spectators from the US and Norway, we can examine whether the differences in redistributive policies and inequality acceptance in the two countries correspond to a difference in how Americans and Norwegians make the trade-off between false positives and false negatives. In particular, do Norwegians have a more generous welfare state than Americans partly because they are more false negative averse?

In both countries we recruited a sample that was nationally representative (+ 18 years old) on a set of observable characteristics (age, gender and geography). Table 2 provides an overview of the background characteristics of the spectators. We observe that income distribution is much more compressed in Norway than in the US, but otherwise the distributions of the different background characteristics are relatively similar. The spectators from the US and Norway were matched with the same group of workers. They were told that the workers had been recruited via an international online marketplace, but were given no information about the nationality of the workers. The distributive situations were therefore identical for the spectators in both countries.

[Table 2 about here]

We recruited 2000 workers from the international online labor market Amazon Mechanical Turk, which specializes in recruiting workers to complete small tasks. When recruited, the workers were promised a participation fee of 2.5 USD and told that they could earn additional money. On average, the workers were paid 4.5 USD, including the participation fee.

[2]Research Now in the US (https://www.researchnow.com/about-us/) and Norstat in Norway (http://www.norstat.co.uk/). Out of the 2000 decisions made by the spectators, 250 were randomly drawn to decide payments for the 250 groups of eight workers who we recruited to the study.

3 Theoretical framework

Here we provide a simple social preference model to guide the analysis and interpretation of the results. The point of departure is a version of the spectator model in Cappelen et al. (2013a), which assumes that a spectator dislikes an individual's payment, y_i, deviating from what he or she views as the fair payment to this individual, m_i:

$$V(y;\cdot) = -\sum_{i \in N} (y_i - m_i)^2 \tag{1}$$

where y is the vector of individual incomes, $y = (y_1, .., y_n)$, where n is the cardinality of the set of individuals N. This model allows for heterogeneity in what the spectators view as fair payments, m_i. Some spectators may view all inequalities as unfair, while others may consider some inequalities, for example those reflecting differences in merit, as fair. However, the model does not allow for heterogeneity in how spectators evaluate deviations from the fair payment. Spectators who have the same fairness view would therefore make the same choice if they were placed in the same distributive situation.

Spectators may, however, disagree about how they evaluate deviations from the fair distribution: some may have a particularly strong dislike for someone getting more than what is fair (a false positive), while others may have a particularly strong dislike for someone getting less than what is fair (a false negative). We therefore introduce a more general version of (1) that allows for *asymmetry* and *heterogeneity* in how people evaluate negative and positive deviations from the fair distribution:

$$V(y;\cdot) = -\sum_{i \in N} (max[0, y_i - m_i])^2 - \beta \sum_{i \in N} (min[0, y_i - m_i])^2, \tag{2}$$

where $0 < \beta < \infty$ is the relative weight attached to false negatives versus false positives.[3] $\beta > 1$ means that the spectator places more weight on avoiding false negatives than avoiding false positives, while the opposite holds true if $\beta < 1$. With $\beta = 1$, the two mistakes are treated symmetrically and (2) is identical to the standard model in (1).

Let us first establish what the model would predict in our experiment if there was no heterogeneity in β. Assume that all spectators agree on what is the fair distribution of payments: non-cheaters should be paid a bonus, while cheaters should not. Based on (2), it then follows that: (i) everyone equalizes in the treatment with no cheaters, (ii) no one equalizes in the treatment where all the workers in the second group are cheaters, and (iii) there exists a "switching treatment" (possibly the treatment with four

[3] It is instructive to contrast the asymmetry introduced in our model with the asymmetry in the weights attached to advantageous inequality and disadvantageous inequality in Fehr and Schmidt (1999). In (2), the asymmetry captures that a spectator may care differently about different types of mistakes, while in Fehr and Schmidt (1999), the asymmetry captures that a stakeholder may care more about inequalities to his or her own disadvantage than about inequalities to the disadvantage of others.

cheaters) defined as a treatment where all spectators equalize in the treatments with fewer cheaters while no one equalizes in this treatment and in treatments with more cheaters. Consequently, the absence of such a "switching treatment" would provide evidence of heterogeneity in β.[4]

To provide a model-based classification of how different people handle false positives and false negatives, we introduce two additional assumptions. First, we assume that the spectators focus on the trade-off between making false negatives and false positives in the second group, which is the trade-off emphasised in the description of the two alternatives. It is an inherent feature of false negatives and false positives in distributive choices that false negatives imply that more money is available to the rest of society and that false positives imply that less money is available to the rest of society. The first assumption implies that this aspect is not decisive when trading off false positives and false negatives. Second, we assume that the spectators consider the fair payment to the cheaters to be zero, $m_c = 0$, and the fair payment to the non-cheaters in the second group to be 2 USD, $m_n c = 2$, which is what these workers would have received if the payment had been distributed equally between both groups. Importantly, this implies that a false positive involves a cheater getting 2 USD more than he or she deserves, while a false negative involves a non-cheater getting 2 USD less than he or she deserves.

Given these assumptions, and letting y^k denote the payment assigned to the individuals in the second group if alternative $k = (A,B)$ is chosen, we can rewrite (2) as follows:[5]

$$V(k;\cdot) = -C(max[0,y^k - 0])^2 - \beta(4-C)(min[0,y^k - 2])^2. \tag{3}$$

Given $y_A = 0$ and $y_B = 2$, (3) implies that $V(A;\cdot) = -\beta(4-C)2^2$ and that $V(B;\cdot) = -C2^2$. A spectator is indifferent between the two alternatives when $V(A;\cdot) = V(B;\cdot)$, i.e. when $\beta = C/(4-C)$. This implies that a spectator with $\beta = 1/3$ would be indifferent between equalizing and not equalizing in the treatment with one cheater, a spectator with $\beta = 1$ would be indifferent between equalizing and not equalizing in the treatment with two cheaters, and a spectator with $\beta = 3$ would be indifferent between equalizing and not equalizing in the treatment with three cheaters. We can now introduce the following model-based classification of spectators:

- **Strictly false positive averse spectators**: $\beta \leq 1/3$.

[4]We here rule out the possibility that all spectators are indifferent between the two alternatives in one of the treatments.

[5]Note that even though the spectator makes a decision in an environment with limited information, where he or she is uncertain about the type (cheater, non-cheater) of each of the workers in the second group, there is no uncertainty in terms of the utility that the spectator gets from each of the two alternatives. This follows from the fact that there is no uncertainty about the number of cheaters in the second group, which is the relevant feature for the spectator's decision.

- **Intermediate spectators:** $1/3 < \beta < 3$.
- **Strictly false negative averse spectators:** $\beta \geq 3$.

(3) provides a justification for focusing on the number of cheaters in the analysis of whether individuals are asymmetric in their handling of positive and negative deviations from a fair distribution. The model implies that the spectator preferences are symmetric ($\beta = 1$) if he or she is indifferent between paying two cheaters and not paying two non-cheaters. Correspondingly, the spectator preferences are asymmetric ($\beta \neq 1$) if he or she prefers to pay three cheaters to avoid one non-cheater not being paid (strictly false negative averse), or vice versa (strictly false positive averse).

4 Empirical strategy

The main empirical specification used in the analysis is:

$$e_i = \alpha + \alpha_1 C_1 + \alpha_2 C_2 + \alpha_3 C_3 + \alpha_4 C_4 + \gamma \mathbf{X}_i + \varepsilon_i, \tag{4}$$

where e_i is an indicator variable taking the value one if the spectator equalizes payments, C_1, C_2, C_3 and C_4 are indicator variables for spectator i being in the treatment with one to four cheaters, respectively, and $\mathbf{X}_i$ is a vector of control variables. Although our main specification includes the control variables, we also report results for regressions without them. We report regressions for the US and Norway separately, and for the pooled sample. The regressions also provide the basis for applying the test outlined in the model section of whether the observed choice pattern shows evidence of heterogeneity in how people trade off false positives and false negatives. Specifically, we test whether we can reject the presence of a unique switching treatment, which would be the case if the level of equalization in any of the intermediate treatments differs significantly from those in both the base treatment with no cheaters and in the treatment with four cheaters.

In addition to the main specification we estimate how the spectators' choice depends on the number of cheaters in the second group, C:

$$e_i = \alpha + \beta C + \gamma \mathbf{X}_i + \varepsilon_i, \tag{5}$$

This specification gives us the average marginal effect of increasing the number of cheaters in the second group.

The estimates from the main specification (4) can also be used to study further the asymmetry in how people trade off false positives and false negatives. Using (3), and the corresponding definitions of the different types, we estimate the share of different types of spectators as follows[6]:

[6]It follows from (3) that a spectator that equalizes with k cheaters in the second group would also have equalized with $k+1$ cheaters in the second group. Since the random variation between treatments

- **The estimated share of strictly false positive averse spectators**. A strictly false positive averse spectator would choose to equalize if and only if there are no cheaters in the second group. The estimated share of these spectators is thus given by $-\alpha_1$.

- **The estimated share of intermediate spectators**. An intermediate spectator would choose to equalize when there is a minority of cheaters in the second group (less than two), but would choose not to equalize in the treatments where the majority are cheaters. The estimated share of these spectators is thus given by $\alpha_1 - \alpha_3$.

- **The estimated share of strictly false negative averse spectators**. A strictly false negative averse spectator would choose not to equalize if and only if everyone in the second group is a cheater. The estimated share of these spectators is thus given by $\alpha_3 - \alpha_4$.

The spectators who choose not to equalize in the treatment with zero cheaters in the second group and the spectators who choose to equalize in the treatment where everyone is a cheater violate the underlying assumption in (3), because they implement false positives or false negatives in situations where this is avoidable. These spectators will be referred to as "unclassified" and the share of unclassified is estimated by $1 + \alpha_4$.[7]

To study heterogeneity in the trade-off between false negatives and false positives, we use specification (5) to examine whether there are differences across subgroups. We also provide estimates of the share of the different spectator types for each subgroup. In this analysis, our main focus is on comparing spectators with different political affiliation and comparing spectators from the US and Norway.

5 Results

We first provide the main analysis of the treatment effects and the corresponding estimation of different types of spectators, before we examine the interaction between the effect of cheaters and political affiliation.

converges to zero as the number of observations goes to infinity, it follows that the estimation approach is asymptotically consistent with $\alpha_1 \geq \alpha_2 \geq \alpha_3 \geq \alpha_4$.

[7]The unclassified spectators consist of those who do not equalize when there are no cheaters in the second group $(1 - \alpha)$ and those who equalize when there are only cheaters in the second group $(\alpha + \alpha_4)$. It follows straightforwardly that this approach provides an exhaustive classification of types in the population: $(-\alpha_1) + (\alpha_1 - \alpha_3) + (\alpha_3 - \alpha_4) + (1 + \alpha_4) = 1$.

5.1 Main analysis

The focus of our analysis is on how the number of cheaters in the second group affects the distributive choices of the spectators.[8] The lower panel in Figure 1 shows the share of spectators who choose to equalize payments for each of the five treatments in the pooled data.

[Figure 1 about here]

We observe that the vast majority of spectators, 90.5 percent, chose to equalize payments when there are no cheaters in the second group. An almost equally large majority, 86.8 percent, chose not to equalize when the second group consists of only cheaters. These results show that we have succeeded in creating an environment where the spectators, with few exceptions, agree that the first best solution is that workers who did the assignment are paid and workers who did not do the assignment are not. As seen from the upper panels in Figure 1, this is the case in both the US and Norway. Differences in how the spectators choose in the intermediate treatments must therefore reflect differences in how they make the trade-off between false positives and false negatives. The choice patterns in the intermediate treatments furthermore provide strong evidence for heterogeneity in how people trade off false positives and false negatives. We do not observe a unique "switching treatment" in either of the countries; rather, we observe a gradual decrease in the willingness to equalize as the number of cheaters increases in the second group.

Table 3 reports the corresponding regression analysis, for the US and Norway separately and for the whole sample combined. Focusing on the regression with the pooled sample and controls (column 6), we observe that the introduction of the first cheater in the second group causes a 9.3 percentage point drop in the share of spectators who choose to equalize ($p < 0.001$). Adding a second cheater further decreases the share who choose to equalize by 17.3 percentage points ($p < 0.001$); thus, about one-third of the spectators choose not to equalize when there is an equal number of cheaters and non-cheaters in the second group. This share only decreases by 0.3 percentage points when we move from two to three cheaters ($p = 0.993$). Finally, we find that introducing a fourth cheater has a dramatic effect on the share who equalize, it drops by 50.4 percentage points ($p < 0.001$). From columns 2 and 4, we observe that the treatment effects are strikingly similar in the US and in Norway, and from columns 1, 3 and 5 we observe that the results are largely the same with and without the inclusion of control variables.

Table 3 also allows us to examine whether certain background characteristics are associated with a higher or a lower willingness to equalize. We find that males and right-wing voters on average are significantly less likely to equalize ($p < 0.001$), while

[8] A large minority of the workers recruited to this study chose to falsely report that they had done the assignment (47.3 percent). This number was not reported to the spectators, who were only told the number of cheaters in the second group in their distributive situation.

low-income individuals are more likely to equalize ($p < 0.01$). We do not find any significant effect of income or education on the willingness to equalize.

[Table 3 about here]

In line with what we observed in Figure 1, we find that the level of equalization in all the intermediate treatments is significantly different from the level of equalization in the treatment with zero cheaters and in the treatment with four cheaters ($p < 0.001$). It therefore follows from the estimated treatment effects that there is no unique "switching treatment", which in turn provides evidence of heterogeneity in the weight attached to false negatives relative to false positives, β.

Table 4 reports regressions capturing the average marginal effect of one more cheater in the second group on the willingness to equalize. For the pooled sample, the estimated effect of onemore cheater in the second group is a 17.2 percentage point reduction in the share who equalize ($p < 0.001$). We observe that the effect is almost identical in the US and Norway. Based on Table 3 and Table 4, we can summarize the first set of main results.

Result 1: *We find that the spectators' willingness to equalize is significantly reduced by an increase in the number of cheaters. We also find strong evidence of heterogeneity in how they trade off false positives and false negatives.*

Using the regression results in Table 3, we estimate the share of strictly false positive averse, intermediate and strictly false negative averse spectators in the sample, as well as the share of unclassified spectators. In Figure 2, we report the estimated shares both separately for the US and Norway and for the pooled sample.

[Figure 2 about here]

In the pooled sample, reported in the upper panel in Figure 2, we estimate that the majority of the spectators, 50.5 percent (46.4 percent in the US and 54.8 percent in Norway), are strictly false negative averse. By contrast, only 9.3 percent of the spectators (9.0 percent in the US and 9.6 percent in Norway) are strictly false positive averse. The share of spectators who are willing to accept three false positives in order to avoid one false negative is thus more than five times as high as the share of spectators who are willing to accept three false negatives in order to avoid one false positive. The share of intermediate spectators is 17.6 percent (21.3 percent in the US and 13.7 percent in Norway), while 22.7 percent of the spectators (23.4 percent in the US and 22.0 in Norway) are unclassified.

Result 2: *We find strong evidence of spectators, both in the US and Norway, having asymmetric preferences for avoiding false positives and false negatives, where the majority are strictly false negative averse and only a small minority are strictly false positive averse.*

The absence of large differences in the behavior of the US and the Norwegian sample is interesting given the well-documented differences between the two countries with respect to redistributive policies and fairness views (Almås et al., 2016). Our findings suggest that these institutional differences do not reflect different views on how to trade off false negatives and false positives.

5.2 Political affiliation

We now turn to an analysis of how distributional choices in the experiment are associated with the political affiliation of the spectators. From column (2) in Table 5, we observe that on average, the share of right-wing spectators who choose to equalize is 9.1 percentage points lower than that for the other spectators. From column (4), which includes an interaction between political affiliation and the number of cheaters, we see that this difference reflects the right-wing spectators' stronger response to the number of cheaters in the second group ($p = 0.035$).[9] From column (6), we furthermore observe that the association between political affiliation and the effect of cheaters is particularly strong in the US, but this country difference is not significant ($p = 0.202$).[10]

Figure 3 reports the classification of spectators for both right-wing and not right-wing spectators in the US and Norway separately and for the full sample. For the full sample, we observe that the share of strictly false negative averse spectators is significantly smaller for the right-wing than for the non-right-wing voters (0.41 vs. 0.54, $p = 0.033$). The difference is somewhat larger in the Norwegian (0.43 vs. 0.59, $p = 0.07$) than in the US sample (0.40 vs. 0.49, $p = 0.321$), but the pattern is similar in both countries.

[Figure 3 about here]

Result 3: *We find a significant political difference in how the spectators trade off false positives and false negatives, with right-wing participants being less false negative averse.*

This result provides evidence of political disagreement partly being about how to trade off false positives and false negatives in second best situations. Such disagreements may prevail even if there is political agreement about what would have been the first-best solution to the distributive problem.

[9] In the Appendix, Table A1, we report the heterogeneity analysis for the other background variables, where we only observe a significant interaction effect for age, with older people being more sensitive to the presence of cheaters.

[10] Table A2 in the Appendix shows that the interaction between political affiliation and the effect of cheaters is also large and significant when we only consider the three intermediate treatments.

6 Conclusion

Some of the most difficult choices people face in their personal and professional lives are choices in which moral mistakes are unavoidable because they have limited information. An important type of such situations involves distributive situations in which it is difficult to distinguish between deserving and undeserving individuals in a group. Decision makers then face a trade-off between not giving to those who are deserving and giving to those who are undeserving. How people make this trade-off constitutes a potentially important source of moral disagreement. The spectator design employed in this study is particularly well suited to identify this important, yet unexplored dimension of people's social preferences: the spectators had no personal stake in the decision and it is therefore reasonable to assume that their choices reflect their social preferences. Furthermore, because the spectators largely agreed on who is deserving and there is no uncertainty about the number of cheaters, differences in their choices identify differences in how they make the trade-off between false positives and false negatives.

We find that spectators, in both the US and Norway, have a stronger dislike for not giving to the deserving (false negatives) than they have for giving to the undeserving (false positives). In particular, our estimates suggest that about half the sample is strictly false negative averse, while only a small minority are strictly false positive averse. However, we also find important heterogeneities with respect to the relative weight attached to the importance of avoiding false positives and false negatives, with right-wing voters being less likely to be strictly false negative averse.

At the end of the experiment, all spectators were asked to state the extent to which they agreed with the statement that the state should help reduce income inequality in society. Table A3 in the Appendix reports regressions on this measure of general support for redistribution in society; we find that it is strongly associated with the choice the spectators made in the experiment. Spectators who equalized in the experiment are significantly more likely to agree with the statement ($p < 0.001$) and this holds also when we control for political affiliation ($p = 0.011$). This association suggests that attitudes towards redistribution in society partly reflect the relative importance attached to false positives and false negatives.

We have studied the trade-off between false positives and false negatives in the context of distributional choices. Similar trade-offs are also present in many other contexts. Importantly, with limited information, any judicial system involves the risk of convicting the innocent and acquitting the guilty. The standard of evidence directly affects the probability of making these mistakes. The legal literature typically argues that it is worse to implement a false negative than a false positive; i.e., it is worse to convict an innocent than to acquit a guilty (Volokh, 1997). There is also an emerging literature studying punishment in public goods and dictator games (Ambrus and Greiner, 2012; Dickson, Gordon, and Huber, 2009; Fehr and Gächter, 2000; Rizzolli and Stanca, 2012; Markussen, Putterman, and Tyran, 2016). An important question in these studies is how decision makers handle the risk of punishing those who do not

deserve punishment. An interesting question for future research is therefore whether people's trade-off between false positives and false negatives in different contexts is associated across different domains.

A key aspect of our design is that we control the spectators' beliefs about the prevalence of false positives and false negatives. However, our findings suggest that such beliefs will be crucial for attitudes to redistributive policies. Future research should therefore also examine heterogeneity in beliefs about the prevalence of false negatives and false positives. A better understanding of the preferences and beliefs people have regarding false positives and false negatives may be of great importance for understanding both individual behavior and institutional design in the private and the public sector.

References

Alesina, Alberto and George-Marios Angeletos (2005). "Fairness and redistribution," *American Economic Review*, 95(4): 960–980.

Alesina, Alberto and Paola Giuliano (2011). "Preferences for redistribution," in Jess Benhabib, Alberto Bisin, and Matthew O Jackson (eds.), "Handbook of Social Economics," volume 1, chapter 4, Amsterdam, NL: Elsevier, pp. 99–131.

Alesina, Alberto and Edward Ludwig Glaeser (2004). *Fighting poverty in the US and Europe: A world of difference*, Oxford, UK: Oxford University Press.

Almås, Ingvild, Alexander W. Cappelen, and Bertil Tungodden (2016). "Cutthroat capitalism versus cuddly socialism: Are Americans more meritocratic and efficiency-seeking than Scandinavians?" mimeo.

Ambrus, Attila and Ben Greiner (2012). "Imperfect public monitoring with costly punishment: An experimental study," *American Economic Review*, 102(7): 3317–32.

Atkinson, Anthony, Thomas Piketty, and Emanuel Saez (2011). "Top incomes in the long run of history," *Journal of Economic Literature*, 49(1): 3–71.

Atkinson, Anthony. B. and Joseph E. Stiglitz (1976). "The design of tax structure: direct versus indirect taxation," *Journal of Public Economics*, 6: 55–75.

Autor, David, Andreas Ravndal Kostol, and Magne Mogstad (2015). "Disability benefits, consumption insurance, and household labor supply," MIT Working Paper.

Balafoutas, Loukas, Martin G. Kocher, Louis Putterman, and Matthias Sutter (2013). "Equality, equity and incentives: An experiment," *European Economic Review*, 60: 32–51.

Barth, Erling, Karl O Moene, and Fredrik Willumsen (2014). "The Scandinavian model - an interpretation," *Journal of Public Economics*, 117: 60–72.

Bartling, Björn, Roberto A. Weber, and Lan Yao (2015). "Do markets erode social responsibility?" *Quarterly Journal of Economics*, 130(1): 219–266.

Bénabou, Roland and Efe A Ok (2001). "Social mobility and the demand for redistribution: The POUM hypothesis." *Quarterly Journal of Economics*, 116(2).

Bénabou, Roland and Jean Tirole (2006). "Belief in a just world and redistributive politics," *Quarterly Journal of Economics*, 12(2): 699–746.

Blake, John (2012). "The return of the 'welfare queen"," cnn.com.

Bolton, Gary E. and Axel Ockenfels (2006). "Inequality aversion, efficiency, and maximin preferences in simple distribution experiments: Comment," *American Economic Review*, 96(5): 1906–1911.

Bortolotti, Stefania, Ivan Soraperra, Matthias Sutter, and Claudia Zoller (2017). "Too lucky to be true. fairness views under the shadow of cheating," Mimeo.

Cappelen, Alexander W., Astri Drange Hole, Erik Ø. Sørensen, and Bertil Tungodden (2007). "The pluralism of fairness ideals: An experimental approach," *American Economic Review*, 97(3): 818–827.

Cappelen, Alexander W., James Konow, Erik Ø. Sørensen, and Bertil Tungodden (2013a). "Just luck: An experimental study of risk taking and fairness," *American Economic Review*, 103(3): 1398–1413.

Cappelen, Alexander W., Karl O. Moene, Erik Ø. Sørensen, and Bertil Tungodden (2013b). "Needs versus entitlements: An international fairness experiment," *Journal of the European Economic Association*, 11(3): 574–598.

Charness, Gary and Matthew Rabin (2002). "Understanding social preferences with simple tests," *Quarterly Journal of Economics*, 117(3): 817–869.

Dickson, Eric S., C. Gordon, Sanford, and Gregory A. Huber (2009). "Enforcement and compliance in an uncertain world: an experimental investigation," *The Journal of Politics*, 71: 1357–1378.

Falk, Armin, Anke Becker, Thomas Dohmen, Benjamin Enke, David Huffman, and Uwe Sunde (2015). "The nature and predictive power of preferences: Global evidence," IZA Discussion Paper. Bonn: Institute of Labor Economics.

Fehr, Ernst and Simon Gächter (2000). "Cooperation and punishment in public goods experiments," *American Economic Review*, 90(4): 980–994.

Fehr, Ernst and Klaus M. Schmidt (1999). "A theory of fairness, competition and cooperation," *Quarterly Journal of Economics*, 114(3): 817–868.

Goodin, Robert E. (1985). "Erring on the side of kindness in social welfare policy," *Policy Science*, 18: 141–156.

Gruber, Jonathan (2000). "Disability insurance benefits and labor supply," *Journal of Political Economy*, 108(6): 1162–183.

Henrich, Joseph, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr, Herbert Gintis, and Richard McElreath (2001). "In search of homo economicus: Behavioral experiments in 15 small-scale societies," *American Economic Review*, 91 (2): 73–78.

Jakiela, Pamela (2015). "How fair shares compare: Experimental evidence from two cultures," *Journal of Economic Behavior & Organization*, 118: 40–54.

Konow, James (2000). "Fair Shares: Accountability and Cognitive Dissonance in Allocation Decisions," *American Economic Review*, 90(4): 1072–1091.

Landersø, Rasmus and James J. Heckman (2016). "The scandinavian fantasy: The sources of intergenerational mobility in denmark and the u.s." NBER Working Paper No. 22465.

Lipsey, R. G. and Kelvin Lancaster (1956). "The general theory of second best," *The Review of Economic Studies*, 24(1): 11–32.

Luttmer, Erzo FP and Monica Singhal (2011). "Culture, context, and the taste for redistribution," *American Economic Journal: Economic Policy*, 3(1): 157–179.

Markussen, Thomas, Louis Putterman, and Jean-Robert Tyran (2016). "Judicial error and cooperation," *Europan Economic Review*, 89: 372–388.

Mirrlees, James A. (1971). "An exploration in the theory of optimum income taxation," *Review of Economic Studies*, 38(2): 175–208.

Rabin, Matthew (1993). "Incorporating fairness into game theory and economics," *American Economic Review*, 83(5): 1281–1302.

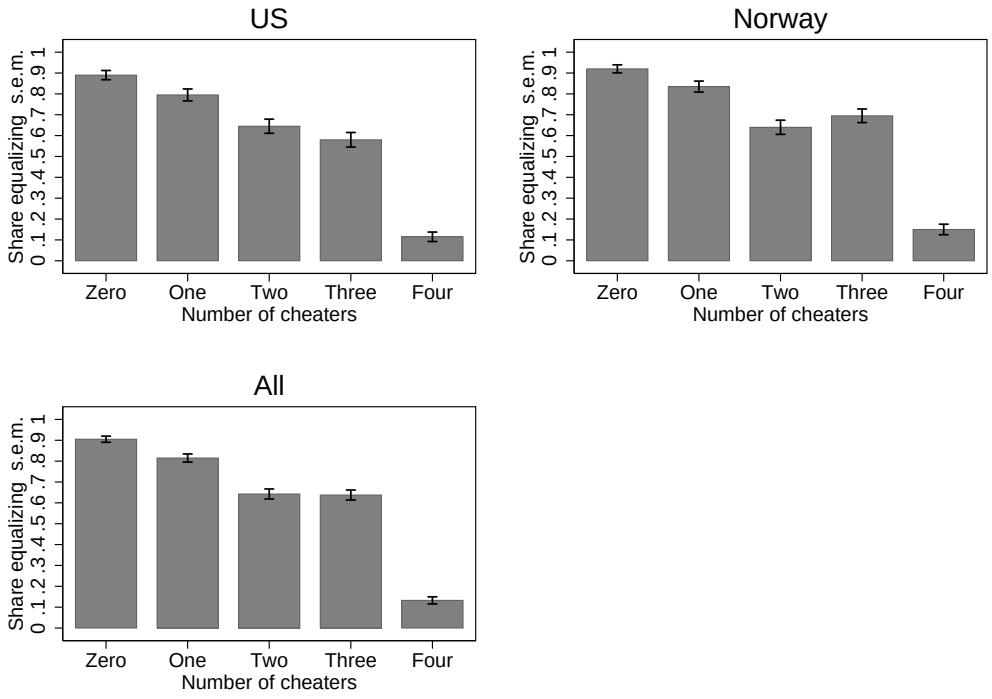
Rizzolli, Matteo and Luca Stanca (2012). "Judicial errors and crime deterrence:theory and experimental evidence," *Journal of Law and Economics*, 55: 311–338.

Stateline (2017). "What happens when states go hunting for welfare fraud," Huffingtonpost.com.

Svallfors, Stefan (1997). "Worlds of welfare and attitudes to redistribution: A comparison of eight western nations," *European Sociological Review*, 13(3): 283–304.

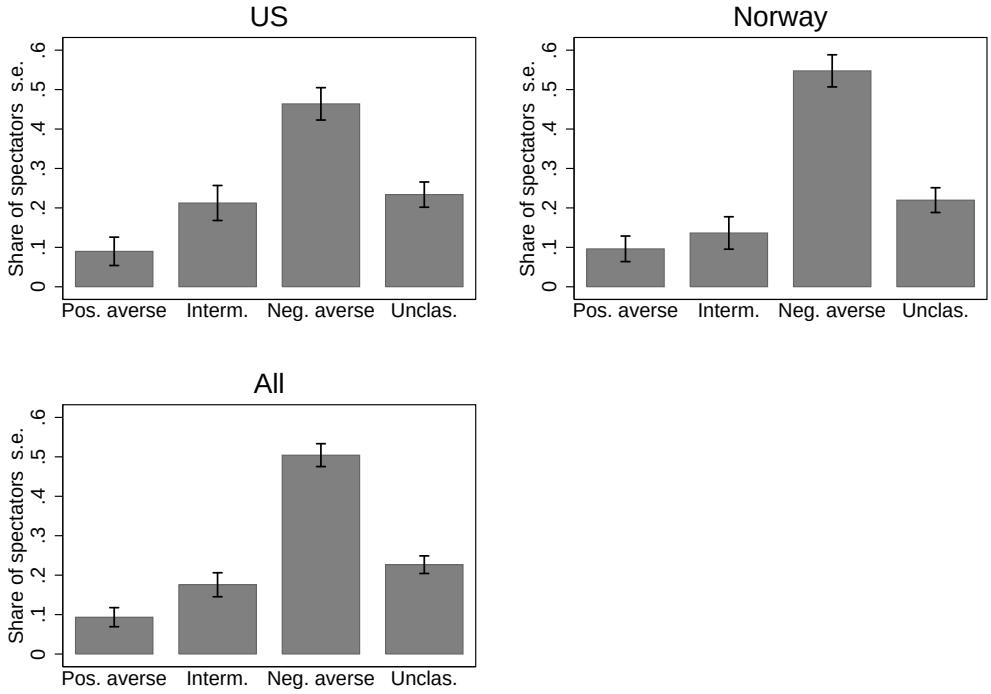
Volokh, Alexander (1997). "*n* guilty men," *University of Pennsylvania Law Review*, 146(1): 173–216.

Figure 1: Share who equalizes by treatment



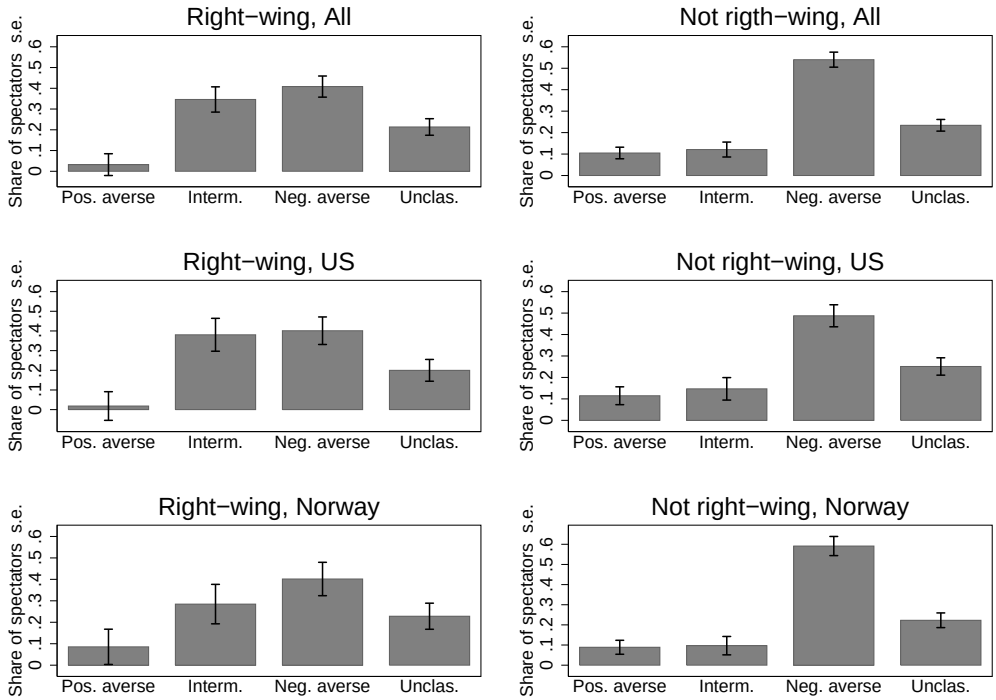
Note: The figure shows the share of spectators who choose to equalize in each of the five treatments. The upper panels show these shares for the US and Norway separately, while the lower panel shows these shares for the pooled sample. The treatments are indicated with the number of cheaters in the second group (where four cheaters imply that everyone has cheated in the second group). The standard errors are indicated by the bars.

Figure 2: Classification by type



Note: The upper panels report the estimated share of spectators who are classified as strictly false positive averse ("Pos. averse"), intermediate ("Interm."), strictly false negative averse ("Neg. averse"), or unclassified ("Unclas.") for each country. The lower panel reports the same shares for the pooled sample. All estimates are based on the regressions reported in Table 3 (columns (2), (4), and (6)). The standard errors are indicated by the bars.

Figure 3: Classification by political affiliation



Note: The figure reports the estimated share of spectators who are classified as strictly false positive averse ("Pos. averse"), intermediate ("Interm."), strictly false negative averse ("Neg. averse"), or unclassified ("Unclas.") by political affiliation. A participant is classified as right-wing if voting for the Republican Party in the US or one of two right-wing parties in Norway ("Høyre" or "Fremskrittspartiet"). The upper two panels report these shares for the full sample, while the middle and bottom panels report the same shares for the US and Norwegian sample, respectively. Estimates are based on the regression in Table 3 (columns (2), (4), and (6)), using right-wing and not right-wing samples separately). The standard errors are indicated by the bars.

Table 1: Sequence of events in the experiment

Stage of experiment
1. **Work stage:** Workers choose whether to do an assignment or to falsely report to have done an assignment.
2. **Matching stage:** Workers are matched in sets of eight with a maximum of four cheaters. Each set is divided into two groups of four: the first group has no cheaters and only consists of workers who did the assignment and the second group consists of either zero, one, two, three or four workers who falsely reported to have done the assignment.
3. **Distribution stage:** Each spectator is randomly matched to a set of workers and asked to decide whether to divide the money equally between the two groups or to give all the money to the first group.
4. **Payment stage:** The workers are paid according to a spectator decision.

Note: The table provides an overview of the main stages of the experiment.

Table 2: Descriptive statistics - background variables for the spectator sample

	United States	**Norway**
Female (share)	0.46	0.50
Age (year)		
Median	44	53
p10	23	27
p90	67	72
Education (share)		
High school or less	0.34	0.37
College	0.37	0.33
High education	0.22	0.30
Income (USD)		
Median	75.000	87.000
p10	25.000	34.000
p90	175.000	155.00
Right-wing (share)	0.30	0.26
Number of participants	1000	1000

Note: The table displays the descriptive statistics for the background variables of the spectator sample. The income variable is yearly household income in USD and given in standard categories where we use the mid-point of the category. A person is classified as right-wing if he or she would have voted for the Republican Party in the US or one of two right-wing parties in Norway (Høyre and Fremskrittspartiet).

Table 3: Effect of the treatments

	US		Norway		All	
	(1)	(2)	(3)	(4)	(5)	(6)
One cheater	-0.095***	-0.090**	-0.085***	-0.096**	-0.090***	-0.093***
	(0.036)	(0.041)	(0.033)	(0.039)	(0.024)	(0.028)
Two cheaters	-0.245***	-0.246***	-0.280***	-0.287***	-0.262***	-0.266***
	(0.041)	(0.041)	(0.039)	(0.039)	(0.028)	(0.028)
Three cheaters	-0.310***	-0.302***	-0.225***	-0.233***	-0.267***	-0.269***
	(0.041)	(0.041)	(0.038)	(0.039)	(0.028)	(0.028)
Four cheaters	-0.775***	-0.766***	-0.770***	-0.780***	-0.773***	-0.773***
	(0.032)	(0.041)	(0.032)	(0.039)	(0.022)	(0.028)
Male		-0.086***		-0.042*		-0.061***
		(0.026)		(0.025)		(0.018)
Low age		0.001		0.052**		0.021
		(0.026)		(0.025)		(0.018)
Low income		0.022		0.067**		0.047**
		(0.028)		(0.026)		(0.019)
Low education		0.023		-0.060**		-0.017
		(0.028)		(0.026)		(0.019)
Right-wing		-0.091***		-0.083***		-0.089***
		(0.028)		(0.029)		(0.020)
Constant	0.890***	0.937***	0.920***	0.944***	0.905***	0.941***
	(0.022)	(0.039)	(0.019)	(0.036)	(0.015)	(0.026)
Observations	1000	1000	1000	1000	2000	2000
R^2	0.301	0.319	0.315	0.335	0.306	0.322

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: The table reports OLS regressions on an indicator for whether the spectator chooses to equalize. "One cheater", "Two cheaters", "Three cheaters", and "Four cheaters" are indicators for the spectator being in the treatment with one, two, three or four cheaters in the second group respectively. "Male" is an indicator for being male, "Low age" is an indicator for being below median age, "Low income" is an indicator for reporting a household income of less than 60.000 USD in the US (35.4 percent of the respondents in the US) and less than 600.000 NOK in Norway (35.7 percent of the respondents in Norway), and "Low education" is an indicator for not having completed a degree beyond high school (36.8 percent of the sample in the US and 33.5 percent of the sample in Norway). A participant is classified as "Right-wing" if voting for the Republican Party in the US or one of two right-wing parties in Norway ("Høyre" or "Fremskrittspartiet).

Table 4: Effect of cheaters

	US		Norway		All	
	(1)	(2)	(3)	(4)	(5)	(6)
Cheaters	-0.177***	-0.174***	-0.168***	-0.170***	-0.172***	-0.172***
	(0.008)	(0.009)	(0.008)	(0.009)	(0.006)	(0.006)
Male		-0.077***		-0.041		-0.055***
		(0.027)		(0.026)		(0.019)
Low age		0.004		0.051*		0.022
		(0.027)		(0.026)		(0.019)
Low income		0.028		0.077***		0.055***
		(0.029)		(0.027)		(0.020)
Low education		0.019		-0.060**		-0.020
		(0.029)		(0.027)		(0.020)
Right-wing		-0.098***		-0.078***		-0.091***
		(0.029)		(0.030)		(0.021)
Constant	0.958***	1.001***	0.984***	1.000***	0.971***	1.000***
	(0.020)	(0.035)	(0.018)	(0.033)	(0.013)	(0.021)
Observations	1000	1000	1000	1000	2000	2000
R^2	0.261	0.279	0.247	0.268	0.254	0.270

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: The table reports OLS regressions on an indicator for whether the spectator chooses to equalize (for the US sample in columns (1)-(2), for the Norwegian sample in columns (3)-(4) and for the pooled sample in columns (5)-(6)). ”Cheater” is the number of cheaters in the second group. “Male” is an indicator for being male, “Low age” is an indicator for being below median age, “Low income” is an indicator for reporting a household income of less than 60.000 USD in the US (35.4 percent of the respondents in the US) and less than 600.000 NOK in Norway (35.7 percent of the respondents in Norway), and “Low education” is an indicator for not having completed a degree beyond high school (36.8 percent of the sample in the US and 33.5 percent of the sample in Norway). A participant is classified as ”Right-wing” if voting for the Republican Party in the US or one of two right-wing parties in Norway (”Høyre” or ”Fremskrittspartiet).

Table 5: Heterogeneity - political affiliation

	(1)	(2)	(3)	(4)	(5)	(6)
Cheaters	-0.172***	-0.172***	-0.164***	-0.165***	-0.163***	-0.162***
	(0.006)	(0.006)	(0.007)	(0.007)	(0.010)	(0.010)
Right-wing	-0.102***	-0.091***	-0.050	-0.039	-0.027	-0.014
	(0.021)	(0.021)	(0.031)	(0.031)	(0.045)	(0.045)
Cheaters*Right-wing			-0.026**	-0.026**	-0.039**	-0.040**
			(0.012)	(0.012)	(0.017)	(0.017)
Norway					0.039	0.050
					(0.030)	(0.030)
Cheaters*Norway					-0.002	-0.005
					(0.014)	(0.014)
Right-wing*Norway					-0.048	-0.049
					(0.063)	(0.062)
Cheaters*Right-wing*Norway					0.030	0.031
					(0.025)	(0.025)
Constant	0.999***	1.000***	0.984***	0.986***	0.964***	0.961***
	(0.014)	(0.021)	(0.015)	(0.022)	(0.023)	(0.027)
Controls	No	Yes	No	Yes	No	Yes
Observations	2000	2000	2000	2000	2000	2000
R^2	0.263	0.270	0.264	0.271	0.266	0.274

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: The table reports OLS regressions on an indicator for whether the spectator chooses to equalize. ”Cheaters” is the number of cheaters in the second group and a participant is classified as ”Right-wing” if voting for the Republican Party in the US or one of two right-wing parties in Norway (”Høyre” or ”Fremskrittspartiet). ”Norway” is an indicator for the participant being in the Norwegian sample. Controls included, but not reported, are indicators for gender, age, education and income.

7 Appendix

Table A1: Heterogeneity - other background characteristics

	(1)	(2)	(3)	(4)	(5)
Cheaters	-0.186***	-0.180***	-0.173***	-0.177***	-0.202***
	(0.008)	(0.007)	(0.007)	(0.007)	(0.011)
Cheaters*Low age	0.029**				0.031***
	(0.011)				(0.012)
Cheaters*Male		0.015			0.018
		(0.011)			(0.012)
Cheaters*Low education				0.013	0.013
				(0.012)	(0.012)
Cheaters*Low income			0.001		0.002
			(0.012)		(0.012)
Constant	0.975***	0.991***	0.977***	0.986***	1.033***
	(0.023)	(0.024)	(0.025)	(0.024)	(0.026)
Controls	Yes	Yes	Yes	Yes	Yes
Observations	2000	2000	2000	2000	2000
R^2	0.271	0.273	0.272	0.273	0.266

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: The table reports OLS regressions on an indicator for whether the spectator chooses to equalize for the three intermediate treatments. "Cheater" is the number of cheaters in the second group. "Low age" is an indicator for being below median age, "Male" is an indicator for being male, "Low education" is an indicator for not having completed a degree beyond high school (36.8 percent of the sample in the US and 33.5 percent of the sample in Norway), and "Low income" is an indicator for reporting a household income of less than 60.000 USD in the US (35.4 percent of the respondents in the US) and less than 600.000 NOK in Norway (35.7 percent of the respondents in Norway). Controls included, but not reported, are indicators for nationality, political affiliation, gender, age, education and income.

Table A2: Heterogeneity - political affiliation - intermediate treatments

	(1)	(2)	(3)	(4)	(5)	(6)
Cheaters	-0.086***	-0.087***	-0.062***	-0.062***	-0.074***	-0.073***
	(0.015)	(0.015)	(0.018)	(0.017)	(0.027)	(0.026)
Right-wing	-0.120***	-0.103***	0.062	0.097	0.123	0.159
	(0.030)	(0.030)	(0.075)	(0.074)	(0.101)	(0.100)
Cheaters*Right-wing			-0.089**	-0.098***	-0.116**	-0.121**
			(0.036)	(0.035)	(0.049)	(0.049)
Norway					0.004	0.018
					(0.072)	(0.071)
Cheaters*Norway					0.024	0.023
					(0.035)	(0.035)
Norway*Cheaters					-0.132	-0.130
					(0.152)	(0.151)
Cheaters*Right-wing*Norway					0.057	0.050
					(0.072)	(0.071)
Constant	0.902***	0.910***	0.856***	0.859***	0.853***	0.846***
	(0.032)	(0.038)	(0.036)	(0.041)	(0.055)	(0.058)
Controls	No	Yes	No	Yes	No	Yes
Observations	1200	1200	1200	1200	1200	1200
R^2	0.038	0.053	0.043	0.059	0.048	0.064

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: The table reports OLS regressions on an indicator for whether the spectator chooses to equalize for the three intermediate treatments. ”Cheaters” is the number of cheaters in the second group. ”Norway” is an indicator for the participant being in the in the Norwegian sample. A participant is classified as ”Right-wing” if voting for the Republican Party in the US or one of two right-wing parties in Norway (”Høyre” or ”Fremskrittspartiet). Controls included, but not reported, are indicators for gender, age, education and income.

Table A3: Support for redistribution

	(1)	(2)	(3)
Equalize	0.335***	0.207**	0.207**
	(0.086)	(0.081)	(0.081)
Right-wing		-1.292***	-1.292***
		(0.093)	(0.093)
Constant	4.576***	5.022***	5.022***
	(0.069)	(0.069)	(0.069)
Controls	No	No	Yes
Observations	2000	2000	2000
R^2	0.008	0.108	0.108

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: The table reports OLS regressions on the response to the following: a "To what extent do you agree or disagree with the following statement: The state should help reduce income inequalities in society", where respondents answered on a 7-point scale from strongly disagree to strongly agree. "Equalize" is an indicator for choosing to equalize in the experiment. A participant is classified as "Right-wing" if voting for the Republican Party in the US or one of two right-wing parties in Norway ("Høyre" or "Fremskrittspartiet). Controls included, but not reported are indicators for gender, age, education and income.

INSTRUCTIONS TO SPECTATORS

(Treatment 3)

Question 1

In contrast to traditional survey questions that concern hypothetical situations, we now ask you to make a choice that could have consequences for a real life situation.

A few days ago, we recruited people via an international online market place and gave them the opportunity to complete an assignment. The assignment was a simple task where the participants were required to work continuously for a certain period of time.

Everyone also got the opportunity to falsely report that they had done the assignment without actually having done it. Those who made this choice did not do any other work.

We want you to decide how to distribute 16 USD between 8 of the recruited individuals. Your decision may be selected to determine the payments to the 8 individuals; it thus could have real life consequences.

6 of the individuals did the assignment, and 2 falsely reported to have done the assignment. You can choose between two ways of distributing the money. Please mark below which alternative you prefer:

Alternative A: Give 4 USD to 4 of the individuals who did the assignment and nothing to the other 4 individuals. This means that 2 individuals who did the assignment are not paid.

Alternative B: Give 2 USD to each of the 8 individuals. This means that the 2 individuals who falsely reported to have done the assignment are paid.

INSTRUCTIONS TO WORKERS

General instructions:

You now take part in a research project. Please, carefully read and follow all instructions. Note that we retrieved your worker ID automatically when you followed the link to this page. We will use it to assign payments and therefore you do **not** need to enter any confirmation code after you have finished the study. You will remain anonymous throughout the study.

You will be paid a fixed participation fee of 2.5 USD and may in addition earn a bonus. If you have any questions regarding this study, you may contact thechoicelab@nhh.no.

A choice

We now want you to make a choice between the two following alternatives.

A. Do a 15 minutes word unscrambling assignment. Your performance will not be measured as there is no right or wrong answer, but we expect you to work continuously on the assignment.
B. Report to have done the 15 minutes word unscrambling assignment without doing it.

Your fixed participation fee does not depend on whether you choose A or B.

Your bonus payment may depend on whether you choose A or B. Your bonus payment is determined by a randomly selected third person. This person will have a sum of money to distribute among you and other participants in this study, and will not be able to distinguish between some of those who have done the assignment and those who have only reported to have done the assignment. You may therefore get paid a bonus both if you choose A and if you choose B.

Below we want you to indicate your choice

Issued in the series Discussion Papers 2017

2017

01/17 January, **Agnar Sandmo**, "Should the marginal tax rate be negative? Ragnar Frisch on the socially optimal amount of work"

02/17 February, **Luca Picariello**, "Organizational Design with Portable Skills"

03/17 March, **Kurt R. Brekke**, Tor Helge Holmås, Karin Monstad og Odd Rune Straume, "Competition and physician behaviour: Does the competitive environment affect the propensity to issue sickness certificates?"

04/17 March, **Mathias Ekström**, "Seasonal Social Preferences".

05/17 April, Orazio Attanasio, Agnes Kovacs, and **Krisztina Molnar**: "Euler Equations, Subjective Expectations and Income Shocks"

06/17 April, **Alexander W. Cappelen**, Karl Ove Moene, Siv-Elisabeth Skjelbred, and **Bertil Tungodden**, "The merit primacy effect"

07/17 May, **Jan Tore Klovland**, "Navigating through torpedo attacks and enemy raiders: Merchant shipping and freight rates during World War I"

08/17 May, **Alexander W. Cappelen**, Gary Charness, **Mathias Ekström**, Uri Gneezy, and **Bertil Tungodden: "**Exercise Improves Academic Performance"

09/17 June, **Astrid Kunze**, "The gender wage gap in developed countries"

10/17 June, **Kristina M. Bott, Alexander W. Cappelen, Erik Ø. Sørensen** and **Bertil Tungodden**, "You've got mail: A randomized field experiment on tax evasion"

11/17 August, Marco Pagano and **Luca Picariello**, "Talent Discovery, Layoff Risk and Unemployment Insurance"

12/17 August, Ingrid Kristine Folgerø, **Torfinn Harding** and Benjamin S. Westby, «Going fast or going green? Evidence from environmental speed limits in Norway"

13/17 August, **Chang-Koo Chi**, Pauli Murto, and Juuso Välimäki, "All-pay auctions with affiliated values"

14/17 August, **Helge Sandvig Thorsen**, "The effect of school consolidation on student achievement".

15/17 September, Arild Sæther, "Samuel Pufendorf and Ludvig Holberg on Political Economy".

16/17 September**, Chang-Koo Chi,** Pauli Murto, and Juuso Välimäki**, "**War of attrition with affiliated values".

17/17 September, **Aline Bütikofer** and Giovanni Peri, "The Effects of Cognitive and Noncognitive Skills on Migration Decisions"

18/17 October, **Øivind Schøyen**, "What limits the powerful in imposing the morality of their authority?

19/17 October, **Charlotte Ringdal** and **Ingrid Hoem Sjursen**, "Household bargaining and spending on children: Experimental evidence from Tanzania"

20/17 December, **Fred Schroyen** and Karl Ove Aarbu, "Attitudes towards large income risk in welfare states: an international comparison"

21/17 December, **Alexander W. Cappelen, Ranveig Falch**, and **Bertil Tungodden**, "The Boy Crisis: Experimental Evidence on the Acceptance of Males Falling Behind

2018

01/18 January, Øystein Foros, Mai Nguyen-Ones, and **Frode Steen**, "Evidence on consumer behavior and firm performance in gasoline retailing"

02/18 January, **Agnar Sandmo**, "A fundamental externality in the Labour Market? Ragnar Frisch on the socially optimal amount of work"

03/18 February, Pierre Dubois and **Morten Sæthre, "**On the Effect of Parallel Trade on Manufacturers' and Retailers' Profits in the Pharmaceutical Sector"

04/18 March, **Aline Bütikofer**, Julie Riise, and Meghan Skira, "The Impact of Paid Maternity Leave on Maternal Health"

05/18 March, **Kjetil Bjorvatn** and **Bertil Tungodden**, "Empowering the disabled through savings groups: Experimental evidence from Uganda"

06/18 April, Mai Nguyen-Ones and **Frode Steen,** "Measuring Market Power in Gasoline Retailing: A Market- or Station Phenomenon?

07/18 April, **Chang Koo Chi** and Trond Olsen, "Relational incentive contracts and performance"

08/18 April, **Björn Bartling**, **Alexander W. Cappelen, Mathias Ekström, Erik Ø. Sørensen,** and **Bertil Tungodden**, «Fairness in Winner-Take-All Markets»

09/18 April, **Aline Bütikofer, Sissel Jensen**, and **Kjell G. Salvanes**, «The Role of Parenthood on the Gender Gap among Top Earners»

10/18 May, **Mathias Ekström**, "The (un)compromise effect"

11/18 May, Yilong Xu**, Xiaogeng Xu**, and Steven Tucker**,** «Ambiguity Attitudes in the Loss Domain: Decisions for Self versus Others»

12/18 June, **Øivind A. Nilsen**, Per Marius Pettersen, and Joakim Bratlie, "Time-Dependency in producers' price adjustments: Evidence from micro panel data"

13/18 June, **Øivind A. Nilsen**, Arvid Raknerud, and Diana-Cristina Iancu, "Public R&D support and firms' performance. A Panel Data Study"

14/18 June, Manudeep Bhuller, Gordon B. Dahl, **Katrine V. Løken**, and Magne Mogstad: «Incarceration, Recidivism, and Employment"

15/18 August, Manudeep Bhuller, Gordon B. Dahl, **Katrine V. Løken**, and Magne Mogstad: «Incarceration Spillovers in Criminal and Family Networks"

16/18 August, Pedro Carneiro, Kai Liu, and **Kjell G. Salvanes**: "The Supply of Skill and Endogenous Technical Change: Evidence From a College Expansion Reform"

17/18 August, **Chang Koo Chi**, "An analysis of the two-bidder all-pay auction with common values"

18/18 August**, Alexander W. Cappelen,** Cornelius Cappelen**,** and **Bertil Tungodden, "**Second-best fairness under limited information: The trade-off between false positives and false negatives"

NHH

Norges Handelshøyskole

Norwegian School of Economics

NHH
Helleveien 30
NO-5045 Bergen
Norway

Tlf/Tel: +47 55 95 90 00
Faks/Fax: +47 55 95 91 00
nhh.postmottak@nhh.no
www.nhh.no