# **ECONSTOR** Make Your Publications Visible.

A Service of

ZBW

Leibniz-Informationszentrum Wirtschaft Leibniz Information Centre for Economics

Valasek, Justin; Wang, Weijia

Working Paper Fair Institutions

CESifo Working Paper, No. 11804

**Provided in Cooperation with:** Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

*Suggested Citation:* Valasek, Justin; Wang, Weijia (2025) : Fair Institutions, CESifo Working Paper, No. 11804, CESifo GmbH, Munich

This Version is available at: https://hdl.handle.net/10419/316918

#### Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

#### Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



# WWW.ECONSTOR.EU



# **Fair Institutions**

Justin Valasek, Weijia Wang





## Impressum:

**CESifo Working Papers** ISSN 2364-1428 (electronic version) Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute Poschingerstr. 5, 81679 Munich, Germany Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de **Editor: Clemens Fuest** https://www.cesifo.org/en/wp An electronic version of the paper may be downloaded · from the SSRN website: www.SSRN.com www.RePEc.org • from the RePEc website: https://www.cesifo.org/en/wp · from the CESifo website:

## Fair Institutions.\*

Weijia Wang and Justin Valasek<sup>†‡</sup>

### April 1, 2025

#### Abstract

The experimental literature on preferences for redistribution has established that individual perceptions of what earning distributions are fair depend greatly on context. In this paper, we study an important and novel dimension of context: whether the choice to redistribute occurs before workers work and accrue earnings, or after. Contrary to the predictions of our theoretical framework, we find no evidence that spectators are less likely to equalize earnings ex ante than to equalize earnings ex post. Interestingly, our study also suggests that, relative to American subjects, Scandinavian subjects are more likely to equalize expost earnings, but we find no evidence that Scandinavian and American subjects make different choices ex ante. A follow-up analysis suggests that the latter result is largely due to Scandinavian and American subjects having similar preferences over ex ante redistribution when equalizing earnings comes at a cost to efficiency. Overall, our results suggest that context-dependent preferences for redistribution are sensitive to the relative timing of the redistribution choice.

**JEL:** C91, D63, J16 **Keywords:** Inequality, Fairness, Institutions, Experiment

<sup>\*</sup>The study is funded by FAIR, the Centre for Experimental Research on Fairness, Inequality and Rationality through grants from Telenor and the NHH Smaforsk Fund. The project is approved by the Institutional Review Board at the Norwegian School of Economics and is preregistered at the AEA Registry (AEARCTR-0012985 "Fair Institutions.")

<sup>&</sup>lt;sup>†</sup>Department of Economics, NHH Norwegian School of Economics, Helleveien 30, 5045 Bergen, Norway. Contact e-mails: weijia.wang@nhh.no justin.valasek@nhh.no.

<sup>&</sup>lt;sup>‡</sup>We are grateful to participants in the 2024 ESA Europe Conference, 2025 Ludwig Erhard ifo Conference on Institutional Economics, seminar participants at OsloMet and the NHH Micro Seminar, Alexander Cappelen, Ranveig Falch, Heidi Thyssen and Bertil Tungodden for helpful comments and suggestions.

## 1 Introduction

A large literature has emerged over the past decades that focuses on detailing individuals' moral preferences for redistribution and how these "fairness preferences" influence important economic decisions. One of the main insights that has emerged from this literature is that individuals' preferences for redistribution depend on context, and specifically the institutions and choices that led to the observed distribution of earnings—for example, whether earnings are allocated based on differences in merit, luck or talent (Cappelen et al., 2013; Konow, 2000). Despite the importance of context, suprisingly there has been little research on how preferences for redistribution depend on whether redistribution occurs through institutions that are set ex ante, or through a redistribution of earnings that is chosen ex post.

To fix ideas, imagine the following two societies. In one society, an ex ante collective decision is made about the proportion of earnings that will be redistributed before individuals make decisions about how much to work and where to invest. That is, redistribution occurs through institutions that are chosen ex ante. In the other society, however, individuals make decisions about labor and investment first, and a collective decision about the proportion of earnings that will be redistributed is made ex post. That is, redistribution occurs through an ex post adjustment of outcomes.

Redistribution in the real world likely occurs through both channels: income tax systems are generally set ex ante, while wealth and capital gain taxes can be changed after earnings have accrued and long-term investments are fixed. However, the existing experimental literature on contextdependent fairness preferences has largely considered preferences over ex post redistribution of outcomes, rather than preferences for redistributionary institutions that are fixed ex ante. Arguably, the advantage of this design is that it ensures that subjects' decisions are not influenced by incentive concerns. Given the sensitivity of fairness preferences to context, however, it is an open question whether individuals' preferences are the same in an ex ante and an ex post setting. In particular, it is unclear whether documented differences in fairness preferences across countries with different cultures and institutions (see Almås et al., 2020, 2025) depend on whether individuals make choices over ex ante institutions or ex post outcomes.<sup>1</sup>

We address this important element of context by conducting an experiment to compare preferences for redistribution when redistribution occurs through ex ante institutions, to when redistribution occurs through changing ex post outcomes. Specifically, we ask subjects who have no individual stake in the experimental outcome, i.e. "spectators," to select a distribution of earnings that will be given to two workers who either (i) will complete a real effort task or (ii) have already completed a real effort task. By comparing the choices of spectators who choose ex ante institutions to the choices of spectators who choose ex post outcomes, we can identify whether preferences for redistribution vary between the two settings. Importantly, we both consider an ex ante setting intended to minimize incentive concerns, and measure beliefs regarding the importance of incentives when choosing ex ante institutions to test for the importance of this mechanism on the observed treatment differences.

To derive predictions in our setting, we utilized a variant of the model of Andreoni et al. (2020), who study intertemporal reversions in fairness pref-

<sup>&</sup>lt;sup>1</sup>We use the term "institutions" to refer to redistributionary choices made ex ante analogous to the Northian definition of institutions, where individuals make choices given the institutions, or "rules of the game," set by society (North, 1991).

erences in a "context free" setting and find that agents equalize expected payoffs at the point they are asked to make a distributional decision.<sup>2,3</sup> In addition to a preference for equal expected earnings, our framework also considers a preference for an earnings distribution that is consistent with workers' expectations. Based on our framework, we predicted that spectators will choose less redistribution when choosing ex ante institutions relative to choosing ex post outcomes. The logic is as follows: when choosing ex post outcomes, expected payoffs are only equalized if the spectator chooses an equal distribution of earnings; when choosing ex ante institutions, however, even unequal distributions of earnings have the same expected utility for both workers since each worker has an equal chance of getting higher or lower earnings.

We test our predictions by comparing an "Institutions" treatment to an "Outcomes" treatment in two different choice environments (a two-bytwo design): a "Divide" treatment where redistribution is costless, in the sense that equalizing payoffs does not impact aggregate earnings, and an "Efficiency" treatment in which redistribution decreases aggregate payoffs. Additionally, we collect data from two different regions—the US and Scandinavia (Sweden)—representing two countries with different institutions and attitudes towards inequality and redistribution. As we discuss further below, a main focus of our study is to measure the impact of the ex ante/ex post context on cross-country comparisons.

<sup>&</sup>lt;sup>2</sup>In particular, Andreoni et al. (2020) follows previous literature on intertemporal fairness preferences and considers spectator choices over lotteries where recipients have not completed a real effort task. We follow the literature on context-dependent fairness preferences and consider spectator choices over workers' earnings. However, in our setting spectators who choose ex ante institutions are effectively choosing a lottery over payoffs, which is why we adopt the framework of Andreoni et al. (2020).

<sup>&</sup>lt;sup>3</sup>Our analysis, predictions and experimental hypotheses were preregistered in the AEA registry entry AEARCTR-0012985 "Fair Institutions."

Contrary to the predictions that we derive based on our theoretical framework that assumes that individuals have a preference for equalizing expected earnings at the time of choice, we find no evidence that subjects are more likely to equalize ex post outcomes relative to choosing equal institutions ex ante. Instead, we find some suggestive evidence for the opposite relationship—i.e. that subjects are more likely to choose equal institutions.

We explore whether this result could be due to spectators choosing ex ante earnings distributions to impact workers' incentives. That is, a natural asymmetry of comparing preferences over ex ante institutions and ex post outcomes is that changing institutions can change incentives, while changing expost outcomes cannot. This raises the question of whether the observed differences between the Institutions and Outcomes treatments are due to differences in moral preferences over the two contexts, or due to a concern for incentives in the Institutions treatments. Notably, our design aims to minimize the perceived impact of incentives by informing spectators that all workers complete a real effort task and are given a baseline payment for the task—importantly, there is no intensive margin to incentivize. However, in a post experiment questionnaire, subjects report that they believe that equalizing payments would increase the probability that workers follow through and complete the task. That is, spectators report that they believe that ex ante institutions could impact incentives on the extensive margin.

Therefore, in a follow-up analysis, we explore whether our findings could be explained by subjects equalizing earnings to provide incentives to complete the real effort task in the Institutions treatments. Additionally, we find that beliefs about the incentive effect of equal payoffs are correlated with spectators' choices at the individual level. However, this correlation is virtually identical in the Institutions treatments and the Outcomes treatments. This suggests that incentive concerns on the extensive margin are unlikely to drive the observed treatment effects: given that the spectator's decision is made after workers complete the real-effort task in the Outcomes treatment, their choice cannot impact incentives. Therefore, given the equal estimated impact of beliefs in both treatments, concerns over incentive effects of earnings are unlikely to have caused the observed treatment differences.

As we discuss in detail following our presentation of the empirical results, this finding can instead be rationalized by a failure of our theoretical assumption that subjects place an equal weight on workers' expectations in both contexts. That is, our design follows a standard approach for eliciting fairness preferences (see Almås et al., 2020, 2025) where spectators are presented with an initial unequal distribution of earnings and are asked whether they would like to change the distribution. As shown in Charité et al. (2022), initial earnings may serve as reference points for the workers and, in turn, the spectator may have a preference for choosing an earnings distribution that close to this reference point.

While our theoretical framework assumes that spectators place an equal weight on the reference point in both settings, arguably the salience of the initial earnings distribution as a reference point may differ between our treatments. In the Outcomes treatments, spectators make a choice after workers have been informed about the initial earnings distribution. In the Institutions treatments, however, spectators make a choice before workers are informed about the initial earnings distribution. This difference could cause spectators to place a lower weight on the (unequal) reference point and equalizing earnings more often in the Institutions treatments, which could rationalize our findings.

Next, as mentioned above, a central focus of our study is to contribute to the knowledge of comparative preferences for redistribution across countries with different cultures and institutions. We focus on the US and Scandinavia since, as highlighted in previous research and in the public debate, the US and Scandinavia are at opposite ends of the spectrum among liberal democracies when it comes to redistributive policies and income inequality (see for example Acemoglu et al., 2017; Almås et al., 2020; Fochesato and Bowles, 2015). Therefore, understanding the root of these differences is an important element for understanding why redistributive institutions and income inequality vary across countries.

Our study replicates previous findings, e.g. Almås et al. (2020), that Scandinavian subjects redistribute more than US subjects when choosing earnings ex post. Surprisingly, however, we find that there is no statistical difference between the level of redistribution chosen by Scandinavian and US subjects when they choose ex ante institutions, and the point estimates of the average level of implemented inequality for the two populations are quite similar. This suggests that the difference between preferences for redistribution between US and Scandinavians is more pronounced when considering ex post redistribution of earnings relative to ex ante institutions.<sup>4</sup>

<sup>&</sup>lt;sup>4</sup>Although we note that the difference between US and Scandinavian subjects in the Outcomes treatments is not statistically different from the difference between US and Scandinavian subjects in the Institutions treatments.

A follow-up analysis suggests a more nuanced picture of the comparison between Scandinavian and US subjects in the Institutions treatments, depending on whether redistribution is costless (Divide treatment) or costly (Efficiency treatment). Specifically, we find that Scandinavian subjects redistribute more than US subjects in all treatments other than the Institutions-Efficiency treatment, where they redistribute *less* than US subjects. While unexpected, we note that conceptually this is consistent with the findings of Almås et al. (2020), who show that while Scandinavian subjects are more egalitarian overall, they are more sensitive to efficiency costs than US subjects—a result that may have been amplified in our Institutions treatment due to the decreased salience of the intermediate earnings distribution.

Our findings relate to several strands of the literature on fairness preferences. First, our study brings together a strand of the literature that has studied the importance of context, (such as effort, luck and agency) on fairness preferences (Akbaş et al., 2019; Almås et al., 2020; Bhattacharya and Mollerstrom, 2022; Cappelen et al., 2013; Konow, 2000) and a strand of the literature that has studied fairness preferences over lotteries and the timing of spectator choices in a context-free setting (Andreoni et al., 2020; Bolton et al., 2005; Trautmann and van de Kuilen, 2016). Here, we show that timing can be an important element of context, and that it may interact with other elements of context in unexpected ways that influence the elicitation and comparison of fairness preferences across treatments.

Our study also contributes to the study of cross-national comparisons of preferences (Almås et al., 2020, 2025; Falk et al., 2018) and shows that comparisons may differ depending on whether choices are made over ex ante institutions or ex post outcomes. As detailed in Breitmoser and Vorjohann (2024); Charité et al. (2022), a potential mechanism for this difference may be that the timing influences the saliency of beliefs about others' expectations about earnings, which in turn influences what earnings distributions individuals view as fair.<sup>5</sup> Lastly, our research also speaks to the literature that has distinguished between a preference for equality of outcomes and equality of opportunities (Alesina and La Ferrara, 2005; Alesina et al., 2018; Saito, 2013)—our results suggest that in a country with a strong norm of equality (Sweden), individuals may be more comfortable trading inequality of efficiency ex ante, when each worker has an equal opportunity of earning the higher income.

## 2 Framework and Experiment Overview

Before detailing the experiment, we introduce a simple framework that follows our experimental setup to formally introduce our research questions and experimental hypotheses. There are three agents: one spectator and two workers, j = A, B. The workers complete an assignment, and are paid according a vector of payoffs  $(x'_A, x'_B)$ , that assigns payoff  $x'_A$  to j = A and  $x'_B$  to j = B. The spectator selects the vector of payoffs  $(x'_A, x'_B)$  either ex ante or ex post from the set of available payoff distributions, X, with an intermediate payoff distribution—a default—of  $(\bar{x}_A, \bar{x}_B)$ . That is, we consider two different choice settings, Outcomes and Institutions, which vary the *timing* of the spectator's choice and the information available to the workers, as summarized below.

 $<sup>{}^{5}</sup>$ In a companion paper, Valasek et al. (2024), we consider the impact of the default earnings profile on spectator preferences. In this paper, we focus on the timing of the choice on spectator preferences.

| Outcomes Setting                                  | Institutions Setting                              |  |  |
|---|---|--|--|
| 1. Workers $\rightarrow$ randomized role $A/B$ .  | 1. Spectator chooses $(x'_A, x'_B)$ from X.       |  |  |
| 2. Workers informed of $(\bar{x}_A, \bar{x}_B)$ . | 2. Workers $\rightarrow$ randomized role $A/B$ .  |  |  |
| 3. Workers complete a real effort task.           | 3. Workers informed of $(\bar{x}_A, \bar{x}_B)$ . |  |  |
| 4. Spectator chooses $(x'_A, x'_B)$ from X.       | 4. Workers informed about $(x'_A, x'_B)$ .        |  |  |
| 5. Workers informed about $(x'_A, x'_B)$ .        | 5. Workers complete a real effort task.           |  |  |
| 6. Workers receive payoffs $x'_A$ and $x'_B$ .    | 6. Workers receive payoffs $x'_A$ and $x'_B$ .    |  |  |

In the Institutions setting the spectator chooses a payment distribution before the workers are randomized into roles, and workers are informed of the payment before completing the task, which corresponds to the institutional definition of setting the "rules of the game" prior to workers completing the real effort task. In the Outcomes setting, the workers have imperfect information about payoffs when completing the real effort task.

Note that spectator decisions do not impact their own monetary payoffs, but spectators have complete information of the setting. Therefore, spectator decisions will only depend on their preferences over the distribution of workers' payoffs. First, we know from previous studies that most subjects will either (i) equalize payoffs, or (ii) select the default payoffs. Accordingly, we will model spectator preferences as consisting of two competing components: a preference for equalizing workers' expected utility, and a preference for leaving the default distribution of payoffs unchanged.

First, to account for a preference for equalizing payoffs we apply the model of spectator preferences introduced by Andreoni et al. (2020), who study intertemporal reversions in distributional preferences based on an expected-utility framework. According to the model, subjects follow a "naive deontological heuristic" which can be understood as a myopic preference for equality: spectators derive utility from equalizing workers' *expected payoffs* at the time when the spectator makes their choice.

Specifically, the spectator's preferences for equality are represented by the following symmetric function:

$$W(EU_j, EU_{j'}) = W(EU_{j'}, EU_j) \text{ for } j \neq j', \tag{1}$$

where  $W(\cdot, \cdot)$  is strictly quasi-concave, and the utility functions of the workers,  $u_j(x_j) = x_j$ , are linear.

Second, as documented in previous research, some spectators display a preference for choosing an earnings distribution that is consistent with workers' expectations and selecting the default payment distribution (see Almås et al., 2020; Charité et al., 2022). To account for this preference, we model spectators as also deriving heterogeneous utility from choosing the default payoff distribution,  $(\bar{x}_A, \bar{x}_B)$ . That is, we model preferences according to the following utility function:

$$u_i(EU_j, EU_{j'}, x'_A, x'_B) = W(EU_j, EU_{j'}) + \beta_i \mathbb{1}(x'_A = \bar{x}_A, x'_B = \bar{x}_B).$$
(2)

Note that we assume that  $\beta_i \in [0, \infty)$  is constant across the Institutions and Outcomes settings. Accordingly, this model of spectator preferences will allow us to make relative predictions on the "cutoff" level of  $\beta$ ,  $\beta^c$ : if  $\beta_i \geq \beta^c$  spectators will choose the default and if  $\beta_i < \beta^c$  spectators will change worker payoffs to be more equal.

In our experiment, we consider two different sets of payoff distribu-

tions,  $X^D \equiv \{(3,3), (4,2), (5,1), (6,0)\}$  and  $X^E \equiv \{(2,2), (4,1), (6,0)\}$ , corresponding to a "Divide" and "Efficiency" treatment respectively. In both cases the default is set to (6,0)  $((\bar{x}_A, \bar{x}_B) = (6,0))$ .

Our model of spectator preferences implies the following results. First, comparing across the Institutions and the Outcomes settings, Proposition 1 details that spectators' relative preference for equalizing payoffs is lower under the Institutions setting given both  $X^D$  and  $X^E$ .

**Proposition 1** For both  $X^D$  and  $X^E$  with  $(\bar{x}_A, \bar{x}_B) = (6, 0)$ ,  $\beta^c$  is lower in the Institutions setting than in the Outcomes setting.

Take  $(x_A^*, x_B^*)$  to be the point that maximizes  $W(EU_j, EU_{j'})$ . At  $\beta^c$ , the utility of the spectator is the same under  $(x_A^*, x_B^*)$  and  $(\bar{x}_A, \bar{x}_B)$ . That is:

$$W(EU_j, EU_{j'}|\bar{x}_A, \bar{x}_B) + \beta^c = W(EU_j, EU_{j'}|x_A^*, x_B^*),$$
(3)

which shows that  $\beta^c$  is increasing in the difference in  $W(\cdot, \cdot)$  at  $(x_A^*, x_B^*)$ and  $(\bar{x}_A, \bar{x}_B)$ .

Note that in the Institutions setting, the expected utility of the workers is the same under any payoff distribution since the spectator's choice of  $(x'_A, x'_B)$  comes before workers are randomized into the roles of worker A/B. For example, since both workers have an equal chance of selecting the role of worker A,  $EU_j = EU_{j'} = 3$  even for  $(x'_A, x'_B) = (6, 0)$ . However, in the Outcomes setting, where roles are set,  $EU_A = 6$  and  $EU_B = 0$  for  $(x'_A, x'_B) = (6, 0)$ . This implies that  $\beta^c = 0$  in the Institutions setting for both  $X^D$  and  $X^E$ , and that  $\beta^c \ge 0$  in the Institutions setting. This proves the result.<sup>6</sup>

Next we compare the relative impact of the Institutions setting with  $X^D$  and  $X^E$ . Note that we have shown above that  $\beta^c$  is the same in the Institutions setting for both  $X^D$  and  $X^E$ . Therefore, the comparison of the relative impact of institutions will depend on the comparison of  $\beta^c$  in the Outcomes setting between  $X^D$  and  $X^E$ .

**Proposition 2** Given  $(\bar{x}_A, \bar{x}_B) = (6, 0)$ , the difference in  $\beta^c$  between the Institutions and Outcomes setting is higher with  $X^D$  than with  $X^E$ .

We provide a short formal proof of the proposition in the Appendix (Section A.1). Intuitively, however, the result follows from the fact that with  $X^E$ , choosing a more equal distribution of payoffs is "costly" to the spectator in two margins: i.e. it results in a utility loss of  $\beta_i$  and reduces the aggregate payoffs to the workers. Since the spectator's utility is increasing in both the equality and the size of aggregate worker payoffs, the spectator will choose a less equal distribution of worker payoffs with  $X^E$  relative to  $X^D$ .

We will draw upon both of the above results when detailing our experimental hypotheses, following an overview of the experiment.

## 2.1 Overview of the Experiment

Our experiment employs a between-individual design that closely mirrors the setup of Almås et al. (2020). Each spectator was randomly assigned to one of the five treatments. For the purpose of comparability, all treatments

<sup>&</sup>lt;sup>6</sup>In the Outcomes setting  $\beta^c > 0$  when  $(x_A^*, x_B^*) \neq (\bar{x}_A, \bar{x}_B)$ . Note that  $(x_A^*, x_B^*) = (3,3)$  for  $X^D$ , which implies  $\beta^c > 0$  for the Outcomes-Divide treatment introduced below.

followed the same basic structure. We collect data from the five different treatments split between two population—a representative US sample and a representative Scandinavian (Swedish) sample.

For each treatment, we gathered data from roughly 250 US subjects and 250 Scandinavian subjects for a total of 2,631 subjects. The spectator data were collected by a professional survey firm and the English-language questionnaires used in the treatments are provided in Section B of the Appendix.<sup>7</sup> All details of the data collection and our empirical analysis and hypotheses were pre-registed in the AEA registry (AEARCTR-0012985).

In each treatment, subjects were asked to make a decision about the distribution of earnings for two workers who either had completed, or who would complete, a real effort task. In all treatments, subjects were free to choose any distribution of bonuses, which were paid on top of a fixed payment of 2 USD; e.g. if the subject chose (4, 2), then one worker will receive a total payment of 6 for the assignment, and the other worker will receive a payment of 4.

As outlined above, in the "Outcomes" treatments, subjects were asked to make their distributional choices after the workers had been sorted into roles and completed the real effort task. In the "Institutions" treatments, subjects were asked to make their distributional choices prior to workers being sorted into roles and this choice was revealed to the workers before they completed the real effort task.

To capture the difference between redistribution through ex post outcomes and redistribution through ex ante institutions, the text of the two baseline treatments differ in the following ways: (1) the Outcome treat-

<sup>&</sup>lt;sup>7</sup>The workers in the experiment were recruited from the international online market place Amazon Mechanical Turk (mTurk).

ments uses past tense ("were recruited...to conduct an assignment"), while the Institution treatments uses future tense ("will be recruited...to conduct an assignment"); (2) distributional choices in the Outcomes treatment are between "Worker A" and "Worker B," while in the Institution treatments "one worker is paid [X] USD and the other worker is paid [Y] USD"; and (3) it is made clear to the spectators that working will be informed about the payment scheme prior to completing the assignment in the Institution treatments ("Prior to completing the assignment the workers will be informed about the third person's decision on the division of their earnings").

Note that, for simplicity and for comparability with previous studies on spectator choices, we do not allow workers to make a choice on the intensive margin, such as their level of effort. Workers were required to complete the same real effort task regardless of the chosen payment scheme. However, in the experiment, workers may always choose to exit the experiment, which means that spectators may anticipate an impact of their distributional decision on the extensive margin. Therefore, we gathered subjects' beliefs on the impact of the payment scheme on worker exit in an ex post experimental questionnaire (we discuss this evidence in Section 3.2).

Within each treatment, Outcomes and Institutions, we run two different treatments to elicit distributional preferences. In the "Divide" treatment, subjects choose bonus payments from the set  $\{(3,3), (4,2), (5,1), (6,0)\}$ ; that is, the sum of the bonus payments is constant. In the "Efficiency" treatment, subjects choose bonus payments from the set  $\{(2,2), (4,1), (6,0)\}$ ; that is, the sum of bonus payments is higher for more unequal distributions.

Lastly, following the example of Almås et al. (2020) and Almås et al. (2025), subjects were asked if they wanted to "change the earnings" from

an intermediate distribution (i.e. a default) of (6,0), in the Outcome-Divide, Outcome-Efficiency, Institutions-Divide and Institution-Efficiency treatments. However, we also run fifth treatment "No-Default Institutions-Divide," which is identical to the Institutions treatment except for the fact that subjects are not primed with any intermediate distribution, and were simply asked to select an alternative among the payment distributions. We discuss this treatment in more detail in Section 3.1 following our main results.<sup>8</sup>

### 2.2 Experimental Hypotheses

We test the following three hypotheses. First, Proposition 1 implies that subjects are more likely to leave the default distribution, (6, 0), unchanged when in the Institutions treatment, leading to a higher level of implemented inequality.

**Hypothesis 1** Subjects implement more inequality when they decide on ex ante institutions compared to ex-post outcomes.

Second, Proposition 2 shows that the difference in implemented inequality between the Institutions treatment and the Outcomes treatment should be higher in the Divide treatment relative to the Efficiency treatment.

**Hypothesis 2** The difference in implemented inequality between the Institution-Divide and the Outcomes-Divide treatments is higher than the difference between the Institutions-Efficiency treatment and the Outcomes-Efficiency treatment.

 $<sup>^{8}</sup>$ Defaults and their impact on the elicitation of fairness preferences is also the topic of our companion paper, Valasek et al. (2024).

Lastly, we are interested in whether the gap between implemented inequality in the US and Scandinavia that has been highlighted in previous research for the Outcomes treatment (see Almås et al., 2020) is replicated in the Institutions treatment, or if the different in fairness preferences between two countries with different cultures and institutions changes in the institutional context. Deriving a precise prediction here using our theoretical model would require a strong assumption on the distributions of  $\beta_i$  in the two populations. Therefore, we do not make an ex ante prediction and test a two-sided hypothesis.

**Hypothesis 3** The difference between implemented inequality in the US and Scandinavia is the same when deciding on ex ante institutions compared to deciding on ex-post outcomes.

## 2.3 Empirical Strategy

In the analysis, we follow Almås et al. (2020) and test our predictions using the following measure of the implemented inequality of spectator ito provide a common measure across treatments:

$$g_i = \frac{|Income \ Worker \ A_i - Income \ Worker \ B_i|}{Total \ Income} \in [0, 1], \qquad (4)$$

where *Income Worker*  $A_i$  is the income allocated to the worker with the \$6 USD initial earnings and *Income Worker*  $B_i$  is the income allocated to the worker without initial earnings. This inequality measure is equivalent to the Gini coefficient in a two-person situation and is equal to one if the spectator decides to transfer nothing to the worker without initial earnings and zero if the spectator decides to equalize the incomes between the two workers. As secondary (preregistered) measure, we also report on the number of subjects who choose a non-egalitarian payment distribution in our data visualization, but do not run any statistical tests on this measure.

We use the following empirical specification, Model 1, to study the treatment effects of institutions on implemented inequality pooled across US and Scandinavian subjects:

$$g_{i} = \alpha + \beta_{1} Institutions_{i} + \beta_{2} Efficiency_{i} + \beta_{3} Institutions_{i} \times Efficiency_{i} + \gamma \mathbf{X}_{i} + \epsilon_{i},$$
(5)

where  $Institutions_i$  is an indicator variable for being in the Institutions-Divide or Institutions-Efficiency treatment,  $Efficiency_i$  is an indicator variable for being in either of the Efficiency treatments, and  $\mathbf{X}_i$  is a vector of control variables.

In Model 1, Outcomes-Division is the base treatment and the estimated value of  $\beta_1$  corresponds to the causal effect of the Institutions treatment on the level of implemented inequality and thus provides a test for Hypothesis 1. The coefficient on the interaction of Institutions and Efficiency,  $\beta_3$ , corresponds with an additional effect of institutions in the Efficiency treatment and thus provides a test for Hypothesis 2. As preregistered, we will report the coefficient estimates both with and without control variables. We use the following variables for individual background control (detailed in the Appendix): age, gender, income, political orientation, and education.

Next, to compare country differences we use the following empirical

specification, Model 2, pooled across the Divide/Efficiency treatments:

$$g_{i} = \alpha + \delta_{0} Institutions_{i} + \delta_{1} S candinavia_{i} + \delta_{2} Institutions_{i} \times S candinavia_{i} + \gamma \mathbf{X}_{i} + \epsilon_{i}.$$
(6)

The estimation of this model will allow us to see if there is an effect of being Scandinavian in the Outcomes treatments  $(\delta_1)$  and/or an effect of being Scandinavian in the Institutions treatments  $(\delta_1 + \delta_2)$ , and to test whether these two estimates are statistically different from each other  $(\delta_2)$ . As above, we also present the estimation of Model 2 with and without controls.

## 3 Results

We begin with a descriptive analysis of data, beginning with a comparison of the Outcomes and Institutions treatments in Figure 1. Focusing on the treatment effect on the implemented Gini, we immediately see that we do not find any evidence that redistribution is lower under the Institutions treatments—in fact, the figures suggest that instead the opposite may be true. Moreover, while we see less redistribution overall under the Efficiency treatment, the difference between Outcomes and Institutions seems comparable in the Divide and Efficiency treatments. The same seems to be true when we consider the impact on the number of individuals who choose a non-egalitarian earnings distribution.



Outcomes

Institutions

Figure 1: Average implemented inequality and proportion Non-Egalitarian by treatment

(a) Average Gini (b) Proportion non-egalitarian

E

Outcomes

Institutions

Next we consider the relative difference in implemented inequality in the US versus Scandinavia (Sweden). As mentioned above, previous papers have documented a lower implemented inequality in Scandinavia that, intuitively, mirrors the different level of redistribution and redistributive institutions across the US and Scandinavia (see Almås et al., 2020). Therefore, one of our main motivations for conducting this experiment is to provide evidence on whether the gap in implemented inequality in Outcomes between a country with a culture of more/less redistribution is replicated in Institutions.

We illustrate the data visually in in Figure 2. Looking first at the Outcomes treatment, the figure shows a smaller average implemented inequality and a higher proportion of subjects choosing egalitarian payments in Scandinavia versus the US. In the Institutions treatments, however, we see only marginal differences in the implemented inequality and the proportion of egalitarian choices between Scandinavia and the US. This suggests that the documented differences in fairness preferences between the US and Scandinavia may indeed depend on context and whether the comparison is between preferences over institutions or between ex post redistribution.





Turing to the preregistered estimation, we summarize the estimation of Models 1 and 2 in the first two columns of Table 1. The coefficient estimates with controls are similar and are listed in the Appendix. Column 3 reports on the same estimation as Model 1, but uses data from the alternative, Nodefault Institutions-Divide treatment that we discuss in Section 3.1 below. Columns 4 and 5 (Models 1a and 1b) are expost, follow-up analyses that further explore the data and that are discussed in detail in, respectively, Sections 3.3 and 3.2.<sup>9</sup>

 $<sup>^{9}</sup>$ Note that Table 1 presents p-value cutoffs (10%, 5% and 1%) from standard, unadjusted two-sided t-tests for all coefficient estimates for the purpose of comparison—we will discuss the statistical tests of our preregistered hypotheses in detail below.

|                                 | (1)      | (2)      | (3)       | (4)                  | (5)                  |
|---------------------------------|----------|----------|-----------|----------------------|----------------------|
|                                 | Model 1  | Model 2  | Model 2a  | Model $1a^{\dagger}$ | Model $1b^{\dagger}$ |
| Institutions                    | -0.0267  | -0.0347* | -0.0655** | -0.0269              | -0.0240              |
|                                 | (0.0209) | (0.0210) | (0.0256)  | (0.0208)             | (0.0227)             |
| Efficiency                      | 0.0496** | -        | -         | 0.0494**             | 0.0495**             |
|                                 | (0.0209) |          |           | (0.0209)             | (0.0208)             |
| Inst×Eff.                       | 0.00793  | -        | -         | -0.0258              | 0.00795              |
|                                 | (0.0297) |          |           | (0.0341)             | (0.0295)             |
| Scandinavian                    | -        | -0.0352* | -0.0279   | -0.0392**            | -                    |
|                                 |          | (0.0210) | (0.0253)  | (0.0171)             |                      |
| $Inst \times Scand.$            | -        | 0.0262   | -0.000218 | -                    | -                    |
|                                 |          | (0.0298) | (0.0362)  |                      |                      |
| $Inst \times Eff \times Scand.$ | -        | -        | -         | 0.0689**             | -                    |
|                                 |          |          |           | (0.0343)             |                      |
| Beliefs                         | -        | -        | -         | -                    | 0.00114***           |
|                                 |          |          |           |                      | (0.000292)           |
| $Inst. \times Beliefs$          | -        | -        | -         | -                    | 0.000171             |
|                                 |          |          |           |                      | (0.000411)           |
| Constant                        | 0.170*** | 0.212*** | 0.184***  | 0.190***             | 0.197***             |
|                                 | (0.0145) | (0.0148) | (0.0179)  | (0.0169)             | (0.0159)             |
| Ν                               | 2109     | 2109     | 1072      | 2109                 | 2109                 |

Table 1: Results: Estimated impact on implemented inequality (Gini)

Standard errors in parentheses; p-values: \* p < 0.10, \*\* p < 0.05, \*\*\* p < 0.01

The empirical results in Table 1 largely correspond to the descriptive analysis. First, note that the coefficient estimates on the Institutions variable,  $\beta_1$  in our empirical model, are consistently negative; i.e. the point estimate implies that on subjects average implement *less* inequality under Institutions. Therefore, given our one-sided hypothesis of a positive effect, we find no statistical evidence in support of Hypothesis 1, as summarized in Result 1.

**Result 1** We find no evidence that spectators are less inequality averse under Institutions relative to Outcomes in the Divide treatment (i.e. there is no evidence that  $\beta_1 > 0$ ).

Next, note that the model suggests a positive impact of Efficiency on implemented inequality; i.e. subjects on average implement more inequality with a cost of redistribution. However, interestingly, given the nonsignificant coefficient estimate on the interaction of Institutions and Efficiency in Model 1, we find no evidence of a different impact of Institutions in the Efficiency treatment relative to the Divide treatment.

**Result 2** We find no evidence that there is an additional impact of Institutions in the Efficiency treatment relative to the Divide treatment (i.e. there is no evidence that  $\beta_3 > 0$ ).

Regarding the comparison between the Scandinavian and US samples, we find that the Scandinavian dummy is significant in the Outcomes treatment, but not in the Institutions treatment. However, we emphasize that while there is a difference in the statistical significance of the Scandinavian dummy in the Outcomes and Institutions treatments, the two coefficients are not statically different from each other. **Result 3** We find a weakly significant difference in implemented inequality between the Scandinavian sample and US sample in the Outcomes treatments, while the coefficient estimate in the Institutions treatments is smaller and statistically insignificant ( $\delta_1 = -0.035$  is significant at the 10% level, while  $\delta_1 + \delta_2 = -0.009$  is not significantly different from 0). However, the two coefficient estimates are not statistically different from each other ( $\delta_2$ is not significantly different from 0).

### 3.1 No-Default Institutions Treatment

There are multiple ways to think about institutional choice. First, societies regularly face a collective decision of whether to change existing institutions; e.g. whether to change the existing system of income taxation. While such decisions are taken through a political process, individual preferences over institutions, and the fairness principles individuals apply in their reasoning, can help explain institutional choice. Our stylized experiment above speaks to individual preferences for changing institutions with an established default: in both the Institutions treatments and Outcomes treatments subjects are told that workers will be informed about an intermediate distribution of earnings—the default—before they are informed about a final distribution of earnings.

However, another relevant context for fairness preferences are the preferences for institutions in a default free setting, analogous to the Rawlsian notion of an "original position," where individuals choose institutions without knowledge of social and historical circumstance (Rawls, 1971). To capture this notion, we also consider an alternative version of the Institutions treatment in which subjects were not primed with a default distribution (No-Default Institutions treatment). That is, subjects were simply asked to choose their preferred payoff distribution without any reference to an intermediate distribution.

For the No-Default Institutions treatment, we only gather data on the Divide treatment and therefore only compare this treatment to our Outcomes-Divide treatment. The estimate of our preregistered analysis are reported in Table 1 in column 3 (Model 2a). The coefficient estimates are comparable to the estimates of Model 2 in our main analysis, although the coefficient point estimate on the Institutions dummy is nearly double in Model 2a. This suggests that both the ex ante choice and the lack of a default may have caused subjects to select more equal earnings distributions.

## **3.2** Incentives and Reference Points

As discussed above, our experimental design precludes the payoff distribution having an impact on the workers' intensive margin, since all workers complete the real-effort task and are paid according to the chosen distribution regardless of their level of effort. However, in the Institutions treatment, it is possible that spectators *believe* that the payoff distribution they select may have an impact on the extensive margin—i.e. whether workers drop out of the experiment and therefore fail to complete the assignment. In contrast, spectator choices cannot have an impact on the extensive margin in the Outcomes treatment, since spectator choices are made after the workers complete the real effort task.

To gather data on spectators' beliefs, we asked them the following two questions in a post-experiment questionnaire: -Out of 100 workers, how many do you think would <u>quit</u> the job after learning that they have a 50 percent chance to earn 6 USD in addition from the assignment, and a 50 percent chance to earn nothing in addition?

-Out of 100 workers, how many do you think would <u>quit</u> the job after learning that they would earn 3 USD in addition from the assignment?

By subtracting the answer to the second question from the answer to the first question, we get a measure of the spectator's beliefs about the incentive effect of choosing a distribution of (3,3) instead of a lottery with an equal chance of a payoff of 6 and 0—we label this measure "Beliefs".

We find that subjects, on average, report a Belief that the payoff distribution has a significant effect on the probability that the worker completes the task: US subjects believe an equal distribution increases the number of workers that complete the task by 26.4, while Scandinavian subjects believe the number increases by 18.3.

This finding suggests the following narrative: any differences in Institutions versus Outcomes could be due to subjects choosing earnings distributions in the Institutions treatment to impact incentives. In fact, Beliefs are correlated with payoff distributions at the individual level in the Institutions treatment: in the Institutions treatment, the correlation between Beliefs and the implemented Gini are 0.179 for the US sample, and 0.095 for the Scandinavian sample.

However, according to this narrative, we would not expect to see a similar correlation in the Outcomes treatment, since the payoffs are selected ex post and therefore cannot influence the workers' choice to complete the assignment. Contradicting this narrative, however, we find that in the Outcomes treatment the correlation between Beliefs and the implemented Gini are 0.177 for the US sample, and 0.066 for the Scandinavian sample i.e. quite similar to the correlations under the Institutions treatment.

For further (suggestive) evidence, we add Beliefs and an interaction between Beliefs and Institutions to Model 1 in our empirical analysis (Model 1b in Table 1). While beliefs about the incentive effect of payoffs are statistically significant in the model estimation, the coefficient estimate on the interaction of Beliefs and Institutions is both economically and statistically insignificant. That is, we find no evidence that Beliefs have a higher impact on the implemented Gini in the Institutions treatment relative the Outcomes treatment, suggesting that beliefs about incentive effects are not driving the observed comparisons between the Institutions and Outcomes treatment in Models 1 and 2.

In summary, our empirical findings do not support our ex ante hypothesis that our findings can be rationalized by spectators using a "naive deontological heuristic" of maximizing worker expected utility at the time of choice (Andreoni et al., 2020), and therefore choose less redistribution when choosing ex ante institutions. However, as we document here, this result is also does not appear to be due to spectators choosing payoff distributions to influence worker effort in the Institutions treatment. Instead, as mentioned in the introduction, based on our findings we think the most likely explanation is that our assumption that  $\beta_i$  is constant across contexts is incorrect.

That is, while both our Institutions and Outcomes treatments feature intermediate payment distributions that may serve as reference points for the workers and spectators (see Charité et al., 2022). However, in the Institutions treatment, workers are informed about the spectator's final distribution of payoffs prior to completing the assignment, while in the Outcomes treatment, they are informed about the default prior to the assignment, and the spectator's choice only after they complete the real effort task. Intuitively, spectators may therefore believe that workers are less likely to form a reference point based on the default in the Institutions treatment, violating the assumption in our model that  $\beta_i$  is constant across settings.

In fact, the importance of the default as a reference point is also highlighted in our No-Default Institutions treatment (Section 3.1), in which we implement a design without a default payment scheme and ask spectators to make a choice over ex ante institutions: when we compare the No-Default Institutions treatment to the Outcomes treatment we see a much larger increase in the level of implemented redistribution (see column 3 in Table 1) relative to the increase in the Institutions treatment that includes a default.

## 3.3 Follow-up analysis: Scandinavia vs the US

A natural follow-up to our findings is to analyze what treatments are driving the observed lack of a statistically or economically significant difference between the US and Scandinavian sample when subjects choose institutions. That is, given the empirical results above and Result 3 in particular, we find it worthwhile to further explore the difference between the Scandinavian and US sample, disaggregated by Outcomes/Efficiency.

We begin with a visual inspection of the data illustrated in Figure 3.

First, in the Divide treatment, we see the expected relationship: the average implemented inequality is lower in Scandinavia relative to the US for both the Outcomes, Institutions and No-Default Institutions treatment. Surprisingly, however, in the Institutions-Efficiency treatment, the implemented inequality is *higher* on average in the Scandinavian sample relative to the US sample. That is, the averages presented in Figure 3 suggest that the lack of a significant difference between the Scandinavian and US sample in the Institutions treatment is largely due to the Institutions treatment having a different relative impact in the Divide and Efficiency treatment.





Based on the descriptive analysis in Figure 3, we run a follow-up analysis, Model 1a in Table 1, to explore whether there is a different relationship between implemented inequality and Scandinavian/US residence in the Institutions-Efficiency treatment. Specifically, in Model 1a we include a level effect for being Scandinavian on the implemented Gini (Scandinavian dummy) and an interaction effect of being Scandinavian in the Institutions-Efficiency treatment (Inst×Eff×Scand.)—we find this model to be the most minimal deviation from Model 1 that can explore whether there is a differential effect for being Scandinavian in the Institutions-Efficiency treatment, relative to the other treatments.

As seen in Table 1, in Model 1a the coefficient estimate on the Scandinavian Dummy is -0.0392, while the coefficient estimate on interaction of Institutions, Efficiency and Scandinavian is 0.0689—this is not a preregistered analysis, but both coefficients are statistically significant at the 5% level using standard tests. That is, our follow-up analysis suggests that Scandinavian subjects redistribute more than US subjects in all treatments other than the Institutions-Efficiency treatment, where they redistribute *less* than US subjects. While unexpected, conceptually this is consistent with the findings of Almås et al. (2020), who show that while Scandinavian subjects are more egalitarian overall, they are more sensitive to efficiency costs than US subjects.

## 4 Conclusion

In this paper, we study whether individuals' preferences for redistribution are sensitive to the timing of the choice and whether redistribution occurs through ex ante institutions or through changing outcomes ex post. We do not find any evidence that subjects choose less redistribution ex ante, and find suggestive evidence that they choose more redistribution ex ante. Moreover, we find that this result does not appear to be driven by perceived incentive effects of the ex ante choice, and suggest that this effect may instead be due to subjects placing a different relative weight on the intermediate distribution of income in the two contexts.

Our results also suggest that the comparison of preferences for redis-

tribution between countries with different cultural and institutional backgrounds may be highly sensitive to the context. Namely, we find that while Scandinavian subjects are more likely to redistribute ex post earnings than American subjects, there is no observed difference between the choices that Scandinavian and American subjects take when they make choices over ex ante institutions. Moreover, there is suggestive evidence that this finding is a result of an unexpected difference in how Scandinavian subjects respond to the Institutions treatment in the Divide and Efficiency setting.

Overall this suggests that measures of preferences for redistribution, and in particular comparisons between countries, are sensitive to the timing of redistribution. Therefore, it may be important to tailor the method used for eliciting preferences for redistribution to the exact research question, and to use a method that measures preferences for ex ante institutions when this is the more relevant metric. Additionally, given the apparent interaction between ex ante institutions and the efficiency of redistribution in our Scandinavian sample, it is an open question whether other factors studied in the literature, such as effort and talent, also have a different impact depending on whether subjects make choices over institutions or outcomes.

Lastly, we highlight the interesting finding that beliefs over incentive effects are correlated with spectators' choices in both the Institutions and Outcomes treatment, even though there cannot be an impact of incentives on worker behavior in the Outcomes treatment. Our experiment does not establish whether this is a causal relationship; however, we find it worthwhile to highlight this finding, since it may point to a new mechanism explaining some of the variation in measured fairness preferences between subjects. Further study—e.g., an experiment that provides information to spectators about the impact of payoffs on worker behavior—is needed to determine whether beliefs about incentive effects are a causal channel, and whether beliefs over incentive effects help determine what spectators view as fair.

## References

- Acemoglu, D., Robinson, J. A., and Verdier, T. (2017). Asymmetric growth and institutions in an interdependent world. *Journal of Political Econ*omy, 125(5):1245–1305.
- Akbaş, M., Ariely, D., and Yuksel, S. (2019). When is inequality fair? an experiment on the effect of procedural justice and agency. *Journal of Economic Behavior Organization*, 161:114–127.
- Alesina, A. and La Ferrara, E. (2005). Preferences for redistribution in the land of opportunities. *Journal of Public Economics*, 89(5):897–931.
- Alesina, A., Stantcheva, S., and Teso, E. (2018). Intergenerational mobility and preferences for redistribution. *American Economic Review*, 108(2):521â54.
- Almås, I., Cappelen, A. W., and Tungodden, B. (2020). Cutthroat capitalism versus cuddly socialism: Are americans more meritocratic and efficiency-seeking than scandinavians? *Journal of Political Economy*, 128(5):1753–1788.
- Almås, I., Cappelen, A. W., Tungodden, B., and Sørensen, E. (2025). Fairness across the world. *Mimeo*.
- Andreoni, J., Aydin, D., Barton, B., Bernheim, B. D., and Naecker, J. (2020). When fair isn't fair: Understanding choice reversals involving social preferences. *Journal of Political Economy*, 128(5):1673–1711.
- Bhattacharya, P. and Mollerstrom, J. (2022). Lucky to work. *GMU Work*ing Paper in Economics No. 22-46.
- Bolton, G. E., Brandts, J., and Ockenfels, A. (2005). Fair procedures: Evidence from games involving lotteries. *The Economic Journal*, 115(506):1054–1076.
- Breitmoser, Y. and Vorjohann, P. (2024). Fairness-based Altruism. Mimeo.
- Cappelen, A., Konow, J., Sørensen, E., and Tungodden, B. (2013). Just luck: An experimental study of risk-taking and fairness. *American Eco*nomic Review, 103(4):1398–1413.
- Charité, J., Fisman, R., Kuziemko, I., and Zhang, K. (2022). Reference points and redistributive preferences: Experimental evidence. *Journal of Public Economics*, 216.

- Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D., and Sunde, U. (2018). Global evidence on economic preferences. *The Quarterly Journal* of Economics, 133(4):1645–1692.
- Fochesato, M. and Bowles, S. (2015). Nordic exceptionalism? social democratic egalitarianism in world-historic perspective. *Journal of Public Economics*, 127:30–44. The Nordic Model.
- Konow, J. (2000). Fair shares: Accountability and cognitive dissonance in allocation decisions. American Economic Review, 90(4):1072â1091.
- North, D. C. (1991). Institutions. *Journal of Economic Perspectives*, 5(1):97â112.
- Rawls, J. (1971). A Theory of Justice. Belknap Press.
- Saito, K. (2013). Social preferences under risk: Equality of opportunity versus equality of outcome. *American Economic Review*, 103(7):3084â3101.
- Trautmann, S. T. and van de Kuilen, G. (2016). Process fairness, outcome fairness, and dynamic consistency: Experimental evidence for risk and ambiguity. *Journal of Risk and Uncertainty*, 53(2):75–88.
- Valasek, J., Vorjohann, P., and Wang, W. (2024). Fairness preferences, inequality acceptance and default effects. *Mimeo*.

## A Appendix

## A.1 Proof of Proposition 2

*Proof:* Since we establish that  $\beta^c$  is equal to zero for the Institutions setting for both  $X^D$  and  $X^E$  above, we only need to establish that  $\beta^c$  is larger with  $X^D$  than with  $X^E$ . Equation 3 shows us that this is true as long as  $W(EU_j, EU_{j'}|x_A^*, x_B^*)$  is higher with  $X^D$  than with  $X^E$ .

Note that  $(x_A^*, x_B^*) = (3, 3)$  with  $X^D$ , since aggregate payoffs are constant for all distributions in  $X^D$ . With  $X^E$ , however,  $(x_A^*, x_B^*)$  could be equal to any distribution in  $X^E$  depending on the functional form of  $W(\cdot, \cdot)$ . That is, the point that maximizes  $W(\cdot, \cdot)$  could be the point that maximizes aggregate payoffs, (6, 0), or that maximizes equality (2, 2), or any point inbetween. However, regardless of which point in  $X^E$  maximizes  $W(\cdot, \cdot)$ ,  $W(EU_j, EU_{j'}|x_A^*, x_B^*) < W(3, 3)$  since either aggregate payoffs and equality are weakly lower for  $(x_A^*, x_B^*)$  than for (3, 3), with one of the two being strictly lower. This shows that  $W(EU_j, EU_{j'}|x_A^*, x_B^*)$  is always higher with  $X^D$  than with  $X^E$ .

### A.2 Empirical results with controls

The following self-reported control variables were collected in a questionnaire following the main experiment:

- Age: In years.
- Gender: Coded as a dummies for responding Female and Other.
- **Income:** Coded as a dummy for having above the median income within each country.
- Political orientation: Coded 1-5 on Left-Right spectrum.
- Education: Coded as a dummy for having bachelor degree education or higher.

As preregistered, Table A.1 presents the estimates of the same model as Table 1, but with the addition of these individual control variables. The coefficient estimates are largely comparable with and without controls.

|                          |                | ( - )          | ( - )           | (                         | ()                                    |
|--------------------------|----------------|----------------|-----------------|---------------------------|---------------------------------------|
|                          | (1)<br>Madal 1 | (2)<br>Madal 2 | (3)<br>Madal 2a | (4)<br>Madal 1a           | (5)<br>Model 1h                       |
| Institutions             |                | 0.0414**       | 0.0750***       | $\frac{1000011a}{0.0306}$ | 1000000000000000000000000000000000000 |
| Institutions             | (0.0206)       | (0.0207)       | (0.0759)        | (0.0206)                  | (0.0274)                              |
|                          | (0.0200)       | (0.0201)       | (0.0202)        | (0.0200)                  | (0.0221)                              |
| Efficiency               | $0.0477^{**}$  | -              | -               | $0.0470^{**}$             | $0.0470^{**}$                         |
|                          | (0.0207)       |                |                 | (0.0207)                  | (0.0206)                              |
| $Inst \times Eff.$       | 0.00791        | -              | -               | -0.0274                   | 0.00797                               |
|                          | (0.0294)       |                |                 | (0.0337)                  | (0.0291)                              |
| Scandinavian             | -              | $-0.0372^{*}$  | -0.0302         | -0.0391**                 |                                       |
|                          |                | (0.0208)       | (0.0248)        | (0.0170)                  |                                       |
| Inst×Scand.              | -              | 0.0318         | 0.00648         | -                         | _                                     |
|                          |                | (0.0294)       | (0.0354)        |                           |                                       |
| Inst×Eff×Scand           | _              | _              | _               | 0 0717**                  | _                                     |
| 111507 (2011) (50001101) |                |                |                 | (0.0339)                  |                                       |
| Poliofa                  |                |                |                 | ()                        | 0 00106***                            |
| Deneis                   | -              | -              | -               | -                         | (0.00100)                             |
|                          |                |                |                 |                           | (0.000205)                            |
| Inst.×Beliefs            | -              | -              | -               | -                         | 0.000186                              |
|                          |                |                |                 |                           | (0.000406)                            |
| Age                      | -0.00264***    | -0.00263***    | -0.00356***     | -0.00266***               | -0.00248***                           |
|                          | (0.000544)     | (0.000545)     | (0.000658)      | (0.000544)                | (0.000541)                            |
| Gender Female            | $0.0569^{***}$ | $0.0561^{***}$ | $0.0472^{***}$  | $0.0565^{***}$            | $0.0545^{***}$                        |
|                          | (0.0148)       | (0.0148)       | (0.0180)        | (0.0148)                  | (0.0147)                              |
| Gender Other             | -0.126         | -0.150         | -0.105          | -0.120                    | -0.111                                |
|                          | (0.195)        | (0.195)        | (0.168)         | (0.195)                   | (0.193)                               |
| High Income              | -0.0173        | -0.0155        | -0.0487**       | -0.0155                   | -0.0180                               |
|                          | (0.0157)       | (0.0158)       | (0.0195)        | (0.0158)                  | (0.0156)                              |
| Dolitical                | 0.0330***      | 0.0249***      | 0.0971***       | 0.03322***                | 0.0221***                             |
| I UIIIICAI               | (0.0330)       | (0.0342)       | (0.0271)        | (0.0333)                  | (0.0331)                              |
|                          | (0.00152)      | (0.00134)      | (0.00555)       | (0.00152)                 | (0.00101)                             |
| Higher Ed.               | -0.0142        | -0.0164        | -0.00584        | -0.0136                   | -0.0136                               |
|                          | (0.0101)       | (0.0101)       | (0.0196)        | (0.0161)                  | (0.0100)                              |
| Constant                 | 0.169***       | 0.206***       | 0.260***        | 0.186***                  | 0.187***                              |
|                          | (0.0361)       | (0.0360)       | (0.0435)        | (0.0369)                  | (0.0365)                              |
| IN                       | 2109           | 2109           | 1072            | 2109                      | 2109                                  |

Table A.1: Estimated impact on implemented inequality (Gini) with controls

Standard errors in parentheses; p-values: \* p < 0.10, \*\* p < 0.05, \*\*\* p < 0.01

## A.3 Instructions: USA

### Treatment 1: Outcomes $\times$ Divide

In contrast to traditional survey questions that are about hypothetical situations, we now ask you to make a choice that has consequences for a real life situation. A few days ago two individuals, let us call them worker A and worker B, were recruited via an international online market place to conduct an assignment.

Worker A and worker B were each offered a participation compensation of 2 USD regardless of what they were paid for completing the assignment. After they had completed the assignment, they were told that it was randomly decided that one of them would earn an additional 6 USD for the work on the assignment while the other would not earn anything additional for the work on the assignment. However, they were also told that a third person could change how the additional earnings would be divided between the two of them and thus determine how much they were paid for the assignment.

You are the third person and we now want you to choose whether to change the earnings for the assignment between worker A and worker B. Your decision is completely anonymous. The workers will receive the payment that you choose for the assignment within a few days, but will not receive any further information.

Worker A was randomly selected to earn 6 USD for the assignment, thus worker B earned nothing for the assignment. Please state which of the following alternatives you choose:

I do not change the earnings:

• worker A is paid 6 USD and worker B is paid 0 USD.

- worker A is paid 5 USD and worker B is paid 1 USD.
- worker A is paid 4 USD and worker B is paid 2 USD.
- worker A is paid 3 USD and worker B is paid 3 USD.

## Treatment 2: Institutions $\times$ Divide

In contrast to traditional survey questions that are about hypothetical situations, we now ask you to make a choice that has consequences for a real life situation. In a few days two individuals, let us call them worker A and worker B, will be recruited via an international online market place to conduct an assignment.

Worker A and worker B will each be offered a participation compensation of 2 USD regardless of what they are paid for completing the assignment. Before completing the assignment, they will be told that their earnings from the assignment will be randomly determined and that one worker would earn 6 USD for the assignment and the other worker would earn nothing for the assignment. They will also be told that a third person was given the opportunity to change how the additional earnings would be divided between the two of them and thus determine how much they will be paid for the assignment. Prior to completing the assignment the workers will be informed about the third person's decision on the division of their earnings.

You are the third person and we now want you to choose how the earnings will be divided between the two workers. Your decision will be completely anonymous. The workers will receive the payment according to your choice for the assignment within a few days, but will not receive any further information.

Please state which of the following alternatives you choose:

I do not change the earnings:

• one worker is paid 6 USD and the other worker is paid 0 USD.

- one worker is paid 5 USD and the other worker is paid 1 USD.
- one worker is paid 4 USD and the other worker is paid 2 USD.
- one worker is paid 3 USD and the other worker is paid 3 USD.

#### Treatment 3: Outcomes $\times$ Efficiency

In contrast to traditional survey questions that are about hypothetical situations, we now ask you to make a choice that has consequences for a real life situation. A few days ago two individuals, let us call them worker A and worker B, were recruited via an international online market place to conduct an assignment.

Worker A and worker B were each offered a participation compensation of 2 USD regardless of what they were paid for completing the assignment. After they had completed the assignment, they were told that it was randomly decided that one of them would earn an additional 6 USD for the work on the assignment while the other would not earn anything additional for the work on the assignment. However, they were also told that a third person could change how the additional earnings would be divided between the two of them and thus determine how much they were paid for the assignment.

You are the third person and we now want you to choose whether to change the earnings for the assignment between worker A and worker B. Your decision is completely anonymous. The workers will receive the payment that you choose for the assignment within a few days, but will not receive any further information.

Worker A was randomly selected to earn 6 USD for the assignment, thus worker B earned nothing for the assignment. If you choose to change the earnings, allocating an additional 1 USD to worker B will reduce worker A's earnings by 2 USD. Please state which of the following alternatives you choose:

I do not change the earnings:

• worker A is paid 6 USD and worker B is paid 0 USD.

- worker A is paid 4 USD and worker B is paid 1 USD.
- worker A is paid 2 USD and worker B is paid 2 USD.

#### Treatment 4: Institutions × Efficiency

In contrast to traditional survey questions that are about hypothetical situations, we now ask you to make a choice that has consequences for a real life situation. In a few days two individuals, let us call them worker A and worker B, will be recruited via an international online market place to conduct an assignment.

Worker A and worker B will each be offered a participation compensation of 2 USD regardless of what they are paid for completing the assignment. Before completing the assignment, they will be told that their earnings from the assignment will be randomly determined and that one worker would earn 6 USD for the assignment and the other worker would earn nothing for the assignment. They will also be told that a third person was given the opportunity to change how the additional earnings would be divided between the two of them and thus determine how much they will be paid for the assignment. Prior to completing the assignment the workers will be informed about the third person's decision on the division of their earnings.

You are the third person and we now want you to choose how the earnings will be divided between the two workers. Your decision will be completely anonymous. The workers will receive the payment according to your choice for the assignment within a few days, but will not receive any further information.

Note that allocating an additional 1 USD to the second worker will reduce first worker's earnings by 2 USD. Please state which of the following alternatives you choose:

I do not change the earnings:

• one worker is paid 6 USD and the other worker is paid 0 USD.

- one worker is paid 4 USD and the other worker is paid 1 USD.
- one worker is paid 2 USD and the other worker is paid 2 USD.

## Treatment 5: No-Default Institutions x Divide

In contrast to traditional survey questions that are about hypothetical situations, we now ask you to make a choice that has consequences for a real life situation. In a few days two individuals will be recruited via an international online market place to conduct an assignment.

The workers will each be offered a participation compensation of 2 USD regardless of what they are paid for completing the assignment. Before completing the assignment, the workers will be told that a third person chose how the earnings for completing the assignment would be divided between the two of them, and they will be informed about the third person's choice.

You are the third person and we now want you to choose how the earnings will be divided between the two workers. Your decision will be completely anonymous. The workers will receive the payment according to your choice for the assignment within a few days, but will not receive any further information.

Please state which of the following alternatives you choose:

- one worker is randomly selected to be paid 6 USD and the other worker is paid 0 USD.
- one worker is randomly selected to be paid 5 USD and the other worker is paid 1 USD.
- one worker is randomly selected to be paid 4 USD and the other worker is paid 2 USD.
- both workers are paid 3 USD.

## **B** Survey and background questions

## Survey - Risk Attitude

**Own risk preference** To what extent do you agree or disagree with the statement: "You are generally willing to take risks."

- Strongly agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Strongly disagree

Belief about worker's risk attitude Out of 100 workers, how many do you think would quit the job after learning that they have a 50 percent chance to earn 6  $\overline{\text{USD}}$  in addition from the assignment, and a 50 percent chance to earn nothing in addition?

• 0 to 100

Out of 100 workers, how many do you think would <u>quit</u> the job after learning that they would earn 3 USD in addition from the assignment?

• 0 to 100

## Survey - Policy View

View on inequality Do you think income differences between rich and poor people are a problem?

- Not a problem at all
- A small problem
- A problem
- A serious problem
- A very serious problem

**Belief about source of inequality** To what extent do you think that differences in income are caused by differences in peopleas effort over their lifetime or rather by luck?

- Only luck
- Mainly luck
- Equally important
- Mainly effort
- Only effort

**Support for redistribution** To what extent do you agree or disagree with the statement: *The national government should aim to reduce the economic differences between the rich and the poor.* 

- Strongly agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Strongly disagree

**Political standing** How would you describe your attitude on economic policy?

- Very left-wing
- Left-wing
- Moderate
- Right-wing
- Very right-wing

## **Background Controls**

What is your age?

What is your gender?

What is the highest level of education you have completed?

Is your annual income above or below \$48,000 / 317,000kr?