

Becker, Christoph; Fehr, Dietmar; Rau, Hannes; Trautmann, Stefan T.; Xu, Yilong

Working Paper

Fairness Properties of Compensation Schemes

CESifo Working Paper, No. 11943

Provided in Cooperation with:

Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

Suggested Citation: Becker, Christoph; Fehr, Dietmar; Rau, Hannes; Trautmann, Stefan T.; Xu, Yilong (2025) : Fairness Properties of Compensation Schemes, CESifo Working Paper, No. 11943, CESifo GmbH, Munich

This Version is available at:

<https://hdl.handle.net/10419/322505>

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.

Fairness Properties of Compensation Schemes

Christoph Becker, Dietmar Fehr, Hannes Rau, Stefan T. Trautmann, Yilong Xu

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.ifo.de/en/cesifo/publications/cesifo-working-papers>

An electronic version of the paper may be downloaded

· from the SSRN website: www.SSRN.com

· from the RePEc website: www.RePEc.org

· from the CESifo website: <https://www.ifo.de/en/cesifo/publications/cesifo-working-papers>

Fairness properties of compensation schemes*

Christoph Becker¹, Dietmar Fehr², Hannes Rau³, Stefan T. Trautmann^{1,4}, and Yilong Xu⁵

¹*University of Heidelberg, Germany*

²*University of Stuttgart, Germany*

³*Karlsruhe Institute for Technology, Germany*

⁴*Tilburg University, Netherlands*

⁵*Utrecht University, Netherlands*

January 26, 2025

Abstract

How do different characteristics of pay-for-performance schemes affect fairness perceptions? In two studies, we systematically consider three major classes of incentive schemes: continuous piece rate incentives, discrete bonus schemes, and tournament incentives. We find that pay inequality has a strong negative effect on perceived fairness. Controlling for pay inequality, people consider piece rate schemes fairer than those with a discrete bonus and a tournament design in particular. Adding performance-dependent resource advantages or handicaps negatively influences perceived fairness. We find that procedural fairness judgments are an important factor influencing overall judgments and demonstrate in a third study that the latter have relevant behavioral consequences.

KEYWORDS: incentives, merit, contract design, fairness, inequality.

JEL CLASSIFICATIONS: C90, D63, J31, J41

* Trautmann (corresponding author): Alfred-Weber-Institute for Economics, University of Heidelberg, Bergheimer Str. 58, 69115 Heidelberg, Germany; Phone: +49 6221 54 2952, Fax: +49 6221 54 3592; email: trautmann@uni-hd.de. Becker: Alfred-Weber-Institute for Economics, University of Heidelberg; email: christoph.becker@awi.uni-heidelberg.de. Fehr: Institute for Economics and Law, University of Stuttgart; email: dietmar.fehr@ivr.uni-stuttgart.de. Rau: Department of Economics, KIT; email: hannes.rau@kit.edu. Xu: Utrecht School of Economics, Utrecht University; email: y.xu3@uu.nl.

1. Introduction

Pay-for-performance contracts are a key feature of any modern governance structure, and there is a long tradition in management research to investigate the optimal design of incentive contracts. However, this research has mostly focused on how bonus and tournament incentive schemes can mitigate the agency problem (Milgrom and Roberts, 1992; Prendergast, 1999; Bell et al., 2021) but less on more intangible issues, such as fairness perceptions of incentive schemes. In theoretical contributions, fairness has typically been incorporated by accounting for agents' utility of monetary consequences and some formal aspects of social comparison (e.g., inequality aversion; Kőszegi, 2014). More general factors of fairness play a minor role when optimizing incentive schemes to induce a certain level of employee effort. However, such generalized fairness perceptions in employment relations are relevant for softer governance aspects, such as organizational culture and worker morale (Greenberg, 1987; Ambrose et al., 2002; Bénabou and Tirole, 2016).

Assuming a framework in which the features of a performance contract affect fairness perceptions with subsequent downstream effects on employee behavior, pay-for-performance contracts may thus have unintended consequences: they may lead to lower productivity, more cheating, and sabotage, or less cooperation and social cohesion in the organization (Harbring et al., 2007; Card et al., 2012; Murayama and Elliot, 2012; Cohn et al. 2015; Fehr, 2018; Breza, Kaur, and Shamdasani, 2018; Fehr et al. 2020; Xu and Marandola, 2023). Therefore, from a governance perspective, understanding how different aspects of contracts affect fairness perceptions is critical to minimize the possible adverse side effects of incentive contracts. Yet, the empirical evidence on fairness perceptions of incentive contracts is exceedingly scarce.¹

¹ While there is a large literature that provides evidence that perceptions of (un)fairness influence employee behavior (e.g., Greenberg, 1987; Ambrose et al., 2002; Cohn et al. 2015; Kao et al., 2018; Fehr et al. 2020), very

We set out to systematically examine the impact of contract incentive features on fairness perceptions. In doing so, we focus on three distinct contractual types typically used to mitigate agency problems that arise through asymmetric information about employee types and behavior: *piece-rate incentives*, which pay per unit contribution, *discrete bonus schemes* paying a discrete bonus once a performance threshold is surpassed, and *tournament incentives* where the high performer receives a larger compensation than the low performer. Moreover, we consider additional contractual features that reallocate resources across agents depending on their performance. Firms and organizations widely use these contract types and features to overcome agency problems (e.g., Lazear, 2018). While their intended effect is to induce performance, they may also have unintended adverse behavioral effects driven by perceptions of the unfairness of the compensation process and outcomes. Importantly, different fairness aspects like inequality, merit, or procedural concerns may affect these perceptions in countervailing ways, making it ex-ante hard to predict how specific features of a compensation scheme affect fairness perceptions.

We provide insights into these relationships using two large-scale studies. In particular, we study how third-party observers (spectators) judge the fairness of payment schemes, focusing on how they weigh specific features of these schemes and the resulting pay inequality. That is, we are interested in impartial fairness judgments that are not biased by experience of own performance or outcomes. To provide a comprehensive account of how different contract features affect fairness judgments, we consider direct effects, exploring contractual features that are possibly regarded as inappropriate or

few papers investigate the effect of different aspects of the incentive contract on fairness perceptions (Jawahar and Stone, 2011; Gupta and Shaw, 2014; see also Cappelen et al., 2020). An exception is the experimental study by Kleinlercher and Stöckel (2018), which varies the ‘saliency’ of the underlying reasons for subjects’ final payoffs and measures the fairness perceptions of the corresponding incentive scheme.

undesirable (e.g., fierce worker competition) and indirect effects, such as the influence of contractual features on the level and distribution of monetary payoffs. In a third study, we probe the meaningfulness of these impartial fairness judgments and elicit spectators' willingness to implement a fairer payment scheme.

Observing fairness perceptions of incentive contracts in the field is complicated by three obstacles. First, observational data does not allow us to compare schemes without confounding unobserved factors of the situation or the workplace. For example, fairness perceptions are likely tainted by experience, performance, and self-serving purposes. Second, selection effects further complicate the interpretation of field data: Because people may select into occupations with specific incentive contracts, fairness judgment may not be unbiased (Dillard and Fisher, 1990; Bartling et al., 2018; Fulmer and Shaw, 2018). Third, randomized experimental variation of several contract features within a fixed organizational context is typically impossible (our study implements systematic variation in 16 conditions).

To overcome these challenges, we run experiments in which respondents provide impartial fairness judgments in different payment scheme scenarios. The scenarios are based on a repeated, time-constrained, real-effort task that we adapted from our previous work (Fehr, Rau, Trautmann, and Xu 2020).² Respondents received a description of the task and a payment scheme, along with information about a typical empirical skill distribution in the real-effort task (independent of the payment scheme) and the resulting monetary outcomes of the payment scheme. They then rate the fairness of the payment scheme based on this information.³ Across raters, we then vary the description of the

² In Fehr, Rau, Trautmann, and Xu (2020), we are interested in the impact of income inequality on social cohesion and show that unfair inequality results in a significant decline in trust and trustworthiness. We discuss this study in more detail when we present the results in Section 3.2.

³ Our design shares some features of vignettes, which offer a contextually detailed description of a specific situation and systematically varies the features of interest. Vignette studies have only recently gained traction

payment scheme, considering the three classes of payment schemes discussed above: continuous *piece rate incentives*, *discrete bonus schemes* that combine a piece rate with a bonus when surpassing a performance target level, and *tournament incentives* with two workers pitted against each other. Within each class, we vary the degree of pay inequality (steepness of the incentives) and consider time advantages and handicaps as a function of earlier performance in the discrete bonus and tournament incentive schemes. Time advantages benefit successful employees by granting additional task completion time while limiting completion time for unsuccessful employees, reflecting the principle that successful employees often receive more resources or opportunities to improve future performance. Time handicaps, on the other hand, penalize high performers with reduced completion time while providing lower performers with additional time resources aiming to create a more level playing field.⁴

This setup is well-suited to address our research question because we can keep the contextual features fixed and only vary the payment scheme. Additionally, the impartial perspective guarantees that any potential stakeholder biases, experiences, and interpretations will not influence the results. The setup also avoids conflating fairness perceptions with changing final outcomes, as would be the case in third-party redistribution designs (e.g., Almås, Cappelen, and Tungodden 2020). Moreover, we can address the empirical challenges described above. First, we avoid selection problems by randomly assigning respondents to payment schemes. We also use an abstract real-effort task, which makes it less likely that uncontrolled aspects of payment schemes in real-

in economic research, for example in research on labor markets (e.g., Finseraas et al., 2016; Kübler, Schmid, and Stueber, 2018), ethical judgments (Ambuehl et al., 2015; Ambuehl and Ockenfels, 2017), and financial literacy (Samek, Kapteyn, and Gray, 2020). Earlier pioneering work include Kahneman, Knetsch, and Thaler (1986) and Dahl and Ransom (1999).

⁴ They also capture ratchet effects, where better performers are subsequently confronted with higher demands.

world labor relationships drive respondents' perceptions. Second, it also allows us to use a controlled manipulation of the different properties of the payment schemes within one conceptual framework. In other words, we can present various payment schemes holding the background situation fixed and, thus, avoid problems of omitted variable bias.

We first recruit two diverse samples of respondents from two different platforms (MTurk and Prolific). In Study 1 (MTurk), we elicit respondents' overall fairness judgments in a non-incentivized way. In Study 2 (Prolific), we elicit non-incentivized and incentivized fairness judgments using the coordination method proposed by Houser and Xiao (2011). This provides a comprehensive perspective on fairness perceptions of incentive contracts. Non-incentivized fairness judgments directly measure how individuals independently perceive different contractual features. In contrast, the incentivized fairness judgments measure a shared understanding of fairness perceptions, revealing how the target population perceives certain contractual features. Moreover, to understand how the fairness judgment depends on specific fairness concepts, in Study 2, we elicit separate judgments about the fairness of the allocation process (*procedural fairness*) and the allocation outcomes of a payment scheme (*outcome fairness*). This distinction matters, given that a payment scheme can be considered procedurally fair even if it results in unequal outcomes (Trautmann, 2023). We test if a payment scheme affects the two concepts differently and how these concepts relate to the overall fairness assessment.

We find that the different features of performance-based compensation schemes profoundly impact fairness perceptions across different samples and measures. People generally view payment schemes leading to higher inequality as unfair. Controlling for the degree of inequality, people consider payment schemes with discrete bonuses and competitive tournaments less fair than piece rates. Furthermore, time advantages or handicaps render a payment scheme less fair. This is interesting because the latter feature aims to level the playing field in subsequent interactions, facilitating more equal outcomes,

which seems desirable from a normative point of view (Roemer, 1998; Konow, 2003). However, a handicap appears at odds with fundamental conceptions of merit, and we find that it significantly reduces procedural fairness judgments. That is, merit is clearly part of a fair procedure. By decomposing fairness perceptions into procedural and outcome fairness, we find that procedural fairness judgments are consistently more positive than outcome fairness judgments in 13 of the 16 payment schemes, and they substantially influence the overall fairness judgments.

Noting clear differences in fairness judgments across payment schemes, we next demonstrate the behavioral relevance of these perceptions. In Study 3, student participants completed an online survey where they first evaluated two payment schemes and then stated their willingness to pay for implementing the fairer payment scheme in a subsequent lab experiment. We randomly varied whether they could participate in this lab experiment, i.e., they could become a partial spectator or remain impartial. We observe a substantial willingness to pay to implement the fairer payment scheme, which is correlated with the strength of the individual-level difference in fairness judgments between the two payment schemes. This holds for both the partial and the impartial spectators. These results highlight that, to the extent that pay-for-performance contracts cannot be avoided, organizations and firms may want to accommodate fairness concerns at least. They can do so by choosing contractual arrangements with fewer adverse effects on fairness perceptions. If unequal outcomes are hard to attain, they may also benefit from emphasizing procedural fairness considerations of their payment schemes.

2. Hypotheses and General Study Design

Our study design aims to test a set of hypotheses regarding the fairness effects of various compensation scheme features. In particular, we hypothesize that, all else being equal, a piece rate scheme will be perceived as fairer than a discrete bonus scheme, which will be

perceived as fairer than a competitive tournament. That is because the mapping of individual performance on reward is closest in the case of piece rate. In discrete bonus schemes, a person may fall short by a bit and forgo the bonus. In contrast, in tournament schemes, performance is benchmarked against another worker, meaning that a strong performance may potentially go unrewarded if the other worker performs even better. Conditional on the type of payment scheme, we also expect that steeper incentives inducing more inequality are perceived as less fair because of inequality aversion.

We also test for the effects of performance-induced resource allocation regarding a time advantage or handicap. The impact of a time advantage or handicap depends on whether considerations of merit or equal opportunity are more compelling in the current context. While very common (giving more resources to potentially more effective agents), we expect resource bonuses to be perceived as unfair because they make the playing field less level. Based on previous research (Dong et al., 2024), we predict that a handicap will also be perceived as unfair. Although they improve equality of opportunity, they also imply an ex-post negative reward for a good performance. This may counter basic principles of merit.

We test these hypotheses in Studies 1 and 2, which are based on the same design and scenarios described in this section. However, they differ concerning the collected fairness judgments, the incentivization, and the subject pools. We describe these aspects separately for each study in Sections 3 and 4.

2.1. Task Scenario

In all conditions, respondents first received the same general description of an incentivized real-effort task, including a screenshot of the task. This task consisted of a screen showing 48 slider bars. The initial position of each slider was at the far left of a bar and had to be moved to the middle position of a bar within a given time frame (setup

adapted from Gill and Prowse, 2012). Moreover, respondents learned that two workers were completing the real-effort task, that the task was repeated four times, and that the workers were paid depending on their success in the task, according to a payment scheme described to them later. It was made clear that, independently of the payment scheme, both workers in a group knew the other worker's performance and earnings after each task repetition (relative performance feedback; Eriksson et al., 2009). Respondents received a description of the specific payment scheme on the next screen. We implemented sixteen payment schemes, which are explained in detail below. Each respondent saw precisely one of these payment scheme scenarios to avoid concerns that ratings are blended with relative comparisons across schemes and order effects. Since we randomly assigned respondents to conditions and have substantial sample sizes, we expect that any individual differences regarding the potential comparison standard of fairness judgments average out in our data.

After reading the task description and the description of the payment scheme, respondents answered four questions testing their understanding of the payment scheme. These questions related to core aspects of the payment scheme, namely the expected earnings for a high-skilled versus a low-skilled worker and the time available for the slider task (which varies across conditions and workers). The questions ensured that respondents understood the task, the payment scheme, and its payoff implications for high and low-skilled workers. Below, we discuss how we treat respondents who failed the comprehension check for each study separately. After answering the comprehension questions, respondents judged the payment scheme's fairness properties. Finally, we ask respondents to fill out a demographic questionnaire (see section Sociodemographic Questionnaire in the Online Supplement).

2.2. Payment schemes

We base the task and payment schemes on a laboratory experiment with real payments (Fehr et al., 2020). We consider three general payment schemes (piece rate, discrete bonus, and tournament) and vary them along two margins. First, we vary the extent of the resulting pay inequality from low to high. Second, we manipulate specific characteristics of the schemes by introducing time advantages or handicaps for successful workers.

We consider 16 payment schemes (see the Online Supplement for the exact wording of all conditions). Table 1 displays the implications of each payment scheme on earnings for the average, the top 10%, and the bottom 10% of the workers. We calibrated these numbers based on workers' performance distribution in Fehr et al. (2020), in which the top 10% correctly placed 29 sliders per round, the bottom 10% set 11 sliders, and the average worker 21 sliders. Note that we apply this same skill distribution in calculating resulting distributions for all 16 schemes; that is, we do not make additional assumptions regarding a potential indirect effect caused by the incentives provided.⁵ We designed the incentive contracts such that an average-skilled worker would earn approximately the same payment in all conditions. Therefore, we calculate the Gini coefficient based on the top and bottom percentile payoffs. Note that we do not consider the potential effects of time advantages or handicaps (described below) on later-round earnings in the calculation of the Gini because these effects depend on the specific performance of both workers. However, *ceteris paribus*, time advantages induce larger inequality by giving more time to already successful workers, while handicaps reduce inequality by giving less time to successful workers.

⁵ Participants may well consider the possibility of such indirect effects as part of their fairness judgments. We did not observe such incentive effects though in the data of Fehr et al. (2020).

2.2.1. Piece Rates

The first four conditions are individual (no interaction) piece-rate schemes. The piece rates differ concerning the steepness of the incentives. The first scheme has a safe base payment and a low piece rate, the second has no base payment and a more significant piece rate, and the third has an entrance fee and an even larger piece rate. The fourth scheme has a non-linear, exponential mapping of the number of sliders correctly placed to payments, punishing low and average performance and rewarding high performance. The time to complete each round of sliders is identical and fixed for all piece rate conditions at 120 seconds.

2.2.2. Discrete Bonus Incentives

We have six conditions involving discrete bonus incentives with a small piece rate incentive and a fixed bonus if a worker reaches a performance target. We included these conditions because such bonuses based on performance targets are essential in real-world payment schemes (Oyer, 2000). For example, in many companies, employees receive yearly bonuses for reaching specific performance targets. Moreover, the discrete bonus conditions form an essential intermediate step to the tournament incentives discussed below. In the tournament, the other worker's performance provides a threshold at which a worker's earnings discontinuously change, depending on whether she performed better or worse than the other worker. This happens in the discrete bonus schemes at the exogenously set performance target level. The six discrete bonus schemes differ regarding the bonus size at the performance target (low vs. high bonus). We calibrated the payment schemes so that for an average worker who attains the target in two of the four rounds of work, the discrete bonus schemes pay approximately the same earnings as the piece rate schemes. However, the payoff difference between high and low performers is amplified in the high bonus condition due to the discrete nature of the bonus contract.

In the first two discrete bonus conditions, each round of sliders is set to a fixed time

limit of 120 seconds. In the other four conditions, involving either high or low bonus levels, we introduce a time advantage or handicap in later rounds based on performance in earlier rounds. Specifically, workers who surpass the target level receive an additional six seconds to complete the next round, while those who fall short have six seconds deducted (time advantage). Conversely, a time handicap reduces the time for high-performing workers by six seconds and increases it by the same amount for lower-performing workers. The idea behind these conditions is that successful workers may receive more resources or confront higher requirements (related to the ratchet effect and higher demands on good performers) for future tasks.

2.2.3. Tournament Incentives

The last six conditions in Table 1 involve competitive tournament incentives for the two workers in a group. The better-performing worker receives a larger prize each round than the worse-performing worker. In half of the six conditions, there is a low difference; in the other half, there is a high difference between the winner and loser prizes. Like piece rates and discrete bonus schemes, tournament incentives lead to income differences between skilled and unskilled workers. However, they also induce a situation where both workers are directly pitted against each other, and income differences are potentially very salient. Moreover, even a good performance may not lead to a high reward, if the matched worker performed even better.

Like the discrete bonus conditions, two tournament conditions involve a fixed time limit of 120 seconds. Four conditions involve the low or high tournament prize differences but add either a time advantage or handicap in later rounds for the winner of earlier rounds. Specifically, with a time advantage (handicap), round winners gain (lose) 6 seconds of completion time in the next round and vice versa for round losers. Again, the idea behind these conditions is that successful workers may receive more or fewer resources than less successful workers for future tasks.

3. Study 1

3.1. Sample and Fairness Assessments

3.1.1 Respondents

We conducted Study 1 using the online labor market Amazon Mechanical Turk (MTurk). We posted our study on MTurk, including a short description of the task, the requirements, and the expected payment for completing the task. After accepting our task, respondents were redirected to our survey programmed in oTree (Chen et al., 2016). To address concerns of response quality, we added a simple CAPTCHA (adding two numbers) for an initial bot screen and, in addition, required that respondents *pass all* four comprehension questions about the payment scheme to proceed to the fairness judgments.

We recruited 2,431 U.S. residents in the fall of 2019 to participate in our study, and we paid \$0.50 to complete the task.⁶ Respondents were randomly allocated to one of the sixteen experimental conditions (on average, 152 respondents per condition). In the Online Supplement (Table OS3), we show that the sample is balanced for observable characteristics such as gender, education, income, and ethnicity, among others, across our 16 payment scheme conditions.

⁶ In total, 4,805 MTurkers clicked on our task. Of these respondents, 216 did not pass the simple CAPTCHA, another 815 did not finish reading the instructions (without seeing the scenarios), 1,335 did not pass the control questions and were excluded from the study, and 8 did not answer the fairness question, leaving us with $n=2,431$ observations (see the Online Supplement for more details about dropout rates). In the Online Supplement we show that there is no differential attrition across conditions.

Table 1. Earnings structure and fairness judgments by the payment scheme

Payment Scheme		Gini	Avg. pay	Top 10% pay	Bottom 10% pay	Fairness judgment non-incentivized ^a Study 1 / Study 2
Piece Rate:						
T1:	Low with Base Pay	0.17	6.40	7.48	5.32	8.51 / 8.46
T2:	Medium No Base Pay	0.45	6.40	9.28	3.52	8.47 / 7.96
T3:	High with Entrance Fee	0.90	6.00	11.40	0.60	7.02 / 6.73
T4:	Exponential	0.91	6.08	25.84	1.28	7.50 / 7.01
Discrete Bonus:						
T5:	Low	0.17	6.84	7.92	5.68	7.83 / 7.93
T6:	High	0.87	6.64	12.32	0.88	7.17 / 6.97
T7:	Low with Time Advantage	0.17	6.84	7.92	5.68	7.19 / 6.86
T8:	High with Time Advantage	0.87	6.64	12.32	0.88	7.17 / 6.40
T9:	Low with Time Handicap	0.17	6.84	7.92	5.68	6.81 / 6.43
T10:	High with Time Handicap	0.87	6.64	12.32	0.88	6.71 / 6.58
Tournament:						
T11:	Low	0.18	6.80	8.00	5.60	8.41 / 7.85
T12:	High	0.82	6.60	12.00	1.20	6.59 / 6.67
T13:	Low with Time Advantage	0.18	6.80	8.00	5.60	6.55 / 6.29
T14:	High with Time Advantage	0.82	6.60	12.00	1.20	5.35 / 5.40
T15:	Low with Time Handicap	0.18	6.80	8.00	5.60	7.17 / 6.82
T16:	High with Time Handicap	0.82	6.60	12.00	1.20	6.55 / 6.37

Notes: Low and high refer to the degree of inequality. Pay (in USD) corresponds to the amount an average (top 10% or bottom 10%) worker would have received as payment under the respective payment scheme.

a: Non-incentivized fairness judgments, measured on a scale from 0 (completely unfair) to 10 (completely fair).

3.1.2 Fairness Assessment

In Study 1, respondents judged the overall fairness of the payment scheme using a simple non-incentivized measure, indicating their judgment on an 11-point scale from completely unfair (0) to completely fair (10):

Let us now consider the overall picture of the payment mechanism described above. Different payment mechanisms to reward people for their work may be considered more or less fair. We are interested in how people think about the above-described mechanism. To what extent do you think this payment mechanism is fair? Please indicate your fairness judgment on a scale from 0 (completely unfair) to 10 (completely fair).

Completely unfair ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Completely fair

3.2 Results

We first observe that fairness perceptions vary substantially across conditions, ranging from 5.35 in the high-inequality tournament with time advantage (T14) to 8.51 in the condition with a low piece rate (T1), see Table 1.⁷

This relative comparison is consistent with our findings for partial fairness judgments in incentivized laboratory games with a student sample (Fehr et al. 2020). In this previous work, we implemented similar versions of these two payment schemes (T1 and T14) and showed that subjects in the perceived unfair condition become less trusting and trustworthy in subsequent interactions than those in the perceived fair condition. The results suggest that (i) partial (in Fehr et al., 2020) and impartial (in the current design) observers have similar perceptions about the relative fairness of payment schemes and (ii) that these fairness perceptions predict downstream behavior. However, in Fehr et al. (2020), we could not disentangle the differential impact of different payment scheme

⁷ We show the full distribution of fairness judgments for all 16 treatments in the Online Supplement.

properties on fairness judgments and downstream behavior.

Next, we study the contribution of the different properties of the payment schemes to fairness judgments. In particular, we analyze the effect of inequality in outcomes as measured by the Gini coefficient, the impact of discrete bonus and tournament incentives compared to piece rate contracts, and additional contract features such as handicaps, time advantages, and exponential schemes. Table 2 shows the results of regressing fairness judgments on inequality, treatment indicators, and indicators for time handicaps, time advantages, and highly skewed piece rates, using all 16 payment schemes. In some specifications, we also include sociodemographic controls and interaction terms for the time advantage and handicap to test if advantages and handicaps are perceived differently in individual discrete bonus schemes compared to tournament schemes.

We find that all design features of contracts that affect fairness lead to lower fairness judgments compared to the low-incentive piece rate contract, rated as most fair (see Table 1). We observe a negative impact of the Gini coefficient on fairness views. The same is true for discrete bonuses and tournament incentives, with tournament incentives judged less favorably than discrete bonus schemes. Both time advantages and time handicaps reduce fairness judgments. Interestingly, the convex incentives in the exponential piece rate scheme do not exert an influence beyond its effect through the Gini. This is despite its similarity to a discrete bonus scheme, with payments becoming significant only after a certain level of performance has been achieved. Its more gradual increase in payment, as opposed to a discrete performance threshold, appears to make the scheme seem fairer.

Table 2. Multivariate analyses of fairness judgments – Study 1

	Fairness judgment			
	(1)	(2)	(3)	(4)
Gini	-1.33*** (0.17)	-1.30*** (0.17)	-1.32*** (0.16)	-1.28*** (0.17)
Discrete Bonus=1	-0.32** (0.15)	-0.24 (0.15)	-0.52*** (0.17)	-0.45*** (0.17)
Tournament=1	-0.74*** (0.14)	-0.70*** (0.15)	-0.55*** (0.16)	-0.51*** (0.17)
Time Advantage=1	-0.94*** (0.16)	-0.99*** (0.17)	-0.28 (0.22)	-0.28 (0.23)
Time Handicap=1	-0.69*** (0.15)	-0.68*** (0.16)	-0.73*** (0.22)	-0.72*** (0.23)
Tournament x Time Advantage			-1.29*** (0.31)	-1.39*** (0.33)
Tournament x Time Handicap			0.07 (0.30)	0.10 (0.31)
Exponential Scheme=1	-0.02 (0.19)	0.01 (0.21)	-0.02 (0.19)	-0.01 (0.20)
Constant	8.72*** (0.10)	8.61*** (0.32)	8.72*** (0.10)	8.59*** (0.32)
Controls	No	Yes	No	Yes
Observations	2,431	2,182	2,431	2,182
R-squared	0.08	0.12	0.09	0.13
<i>F-statistics (Discrete Bonus = Tournament)</i>	10.16***	11.69***	0.02	0.08
<i>F-statistics (Time Bonus = Time Handicap)</i>	2.27	3.27*	3.60*	3.26*

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. OLS regression on fairness judgments with robust standard errors in parentheses. Piece rate is the reference category. Controls include gender, level of education, personal income (categorical), indicators for ethnic groups and for living in rural areas, political orientation (1: Liberal; 6: Conservative), self-perceived social status, and employment status.

The specifications in the last two columns suggest that time handicaps for successful workers are never considered fair. Therefore, it seems that a handicap is at odds with

fundamental conceptions of merit, even though it aims at leveling the playing field.⁸ In contrast, time advantages are judged neutrally in individual-level discrete bonuses but strongly negatively in the case of competitive tournament conditions. The latter effect is remarkable because tournament-like settings involve better resources for successful people (scholarships, grants, special training, networking opportunities) in many practical contexts. The results show that all typical design features of incentive contracts come at a cost in terms of reduced fairness perceptions.⁹

Next, we explore whether these patterns are robust for different population subgroups or whether they differ across demographic groups. Such differences can be driven, for example, by individual differences in personality (Fulmer and Shaw, 2018) or differences in intrinsic and non-monetary motivators (Benabou and Tirole, 2003, Erkal, Gangadharan, and Koh, 2018). Table 3 presents the results of estimating the full model (see Table 2, column 4) separately for high- and low-income individuals, politically conservative and liberal individuals, and males and females.

We observe that the above-identified pattern of effects is remarkably consistent across different groups. Still, there are some noteworthy differences. For example, high-income respondents are less averse to bonus and tournament payment schemes. This could be interpreted as a self-serving bias attributing their achievements to a greater extent to their merits (see, for example, Fehr and Vollmann, 2022). Similarly, negative views of handicaps for successful workers seem to be driven predominantly by high-income and conservative respondents, who may perceive these design features as violating basic meritocratic

⁸ This result contrasts with findings from Schildberg-Hörisch et al. (2020). They show in a lab experiment that subjects do not judge affirmative action rules (quota rules for low performance due to (i) bad luck, (ii) low productivity, and (iii) short working time), as less fair than no affirmative action.

⁹ Online Appendix OS9 shows that the results are robust to binarizing fairness judgments at either a score of 5 or the median score of 7. Online Appendix OS10 shows that they are robust to excluding the Gini from the regressions. The convex piece rate scheme then becomes negative, capturing the effect of inequality.

principles. Furthermore, male subjects view tournament incentives more negatively. This association with gender is a bit surprising, given the literature showing that females, unlike men, often avoid competitive payment schemes (Niederle and Vesterlund, 2007; Ors et al., 2013; Buser et al., 2014).

Table 3. Fairness view analyses by income, political view, and gender

	Income		Political view		Gender	
	(1)	(2)	(3)	(4)	(5)	(6)
	High	Low	Conservative	Liberal	Female	Male
Gini	-1.08*** (0.24)	-1.48*** (0.25)	-1.03*** (0.28)	-1.41*** (0.21)	-1.25*** (0.25)	-1.26*** (0.24)
Discrete Bonus=1	-0.12 (0.23)	-0.76*** (0.25)	-0.12 (0.27)	-0.62*** (0.22)	-0.50* (0.26)	-0.45** (0.23)
Tournament=1	-0.19 (0.14)	-0.83*** (0.24)	-0.66** (0.27)	-0.44** (0.21)	-0.28 (0.23)	-0.68*** (0.24)
Time Advantage=1	-0.28 (0.32)	-0.35 (0.32)	-0.09 (0.35)	-0.40 (0.29)	-0.29 (0.33)	-0.27 (0.31)
Time Handicap=1	-1.14*** (0.31)	-0.29 (0.33)	-1.43*** (0.36)	-0.28 (0.29)	-0.61* (0.31)	-0.75** (0.33)
Tournament x Time Advantage	-1.30*** (0.45)	-1.47*** (0.48)	-1.66*** (0.52)	-1.16*** (0.42)	-1.77*** (0.46)	-1.03** (0.46)
Tournament x Time Handicap	0.39 (0.44)	-0.21 (0.45)	0.70 (0.54)	-0.31 (0.39)	-0.61* (0.31)	0.45 (0.46)
Exponential Scheme=1	0.31 (0.27)	-0.34 (0.31)	0.13 (0.31)	-0.11 (0.27)	-0.21 (0.34)	0.10 (0.25)
Constant	8.53*** (0.49)	9.03*** (0.43)	8.72*** (0.49)	9.38*** (0.39)	8.96*** (0.43)	8.44*** (0.46)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,102	1,080	794	1,388	1,033	1,142
R-squared	0.11	0.15	0.11	0.12	0.14	0.12

Notes: *** p<0.01, ** p<0.05, * p<0.1. OLS regression on fairness judgments with robust standard errors in parentheses. High/Low income, Conservative/Liberal is the median split. Controls include gender, level of education, personal income (categorical, high: ≥ 1.500 \$; not included in regression (1) and (2)), indicators for ethnic groups and for living in a rural area, political orientation (1: Liberal; 6: Conservative, not included in regression (3) and (4)), self-perceived social status, and employment status. Results are robust with or without control variables.

4. Study 2

4.1. Sample and Fairness Assessments

4.1.1 Respondents

We conducted Study 2 on Prolific with the exact same 16 condition as in Study 1. Again, we provided a short description of a task scenario, including the task, task requirements, and the expected payment for completing the task. After reading the study material, respondents answered four comprehension questions. If they failed a question, they first got a notification that their solution was incorrect, and if they failed a second time, they learned the correct solution. In any case, respondents could then proceed to the study. On average, 16 percent of the respondents failed to answer some or all comprehension questions correctly before moving to the fairness judgments. In Table OS5 of the Online Supplement, we show that these respondents respond less strongly (though qualitatively in the same way) to the scenarios than those who correctly answered the comprehension checks on their first attempt. This is what we would expect if respondents do not fully and accurately understand the consequences of the mechanisms. In the main text, we report results that include all respondents, thus providing a conservative lower bound on fairness judgments.

We recruited 2,423 US respondents on Prolific in 2022. That is, on average, 151 respondents per condition. Payment included a fixed amount of £1 and the opportunity to earn another £1.20 through incentivized fairness judgments (see below for details). At the median, respondents earned £1.40 for a task that took about 8 minutes. In the Online Supplement, we show that the sample is balanced concerning observable characteristics such as gender, education, income, and ethnicity, among others, across our 16 payment scheme conditions.

4.1.2 Fairness Assessment

In Study 2, we first replicate the basic fairness judgment task employed in Study 1. We then additionally collect the following measures: First, we extend the overall fairness judgment by including judgments regarding the outcome and procedural fairness of the payment scheme. The two items are measured on the same 11-point scale as the overall measure and are described as follows:

Let us now consider two specific aspects of fairness. First, a payment mechanism may involve fair and balanced procedures, although the resulting distribution of payments to different workers may not be fair. That is what we call *procedural fairness*. Second, a payment mechanism may result in a distribution of payments to other workers that can be considered fair, irrespective of whether the procedures that led to the outcomes were fair. This is what we call *outcome fairness*.

After collecting respondents' non-incentivized fairness judgments, we collect incentivized measures of fairness views using the method of Houser and Xiao (2011). The method involves incentivizing respondents to coordinate on the most salient fairness judgment by paying them for correctly indicating the majority (modal) judgment of all other respondents. It can thus be interpreted as a measure of the perceived social standard, i.e., a fairness norm in the group. The task was described as follows (overall fairness measure).

We are now interested in how you think the majority of participants in this study think about the mechanism's overall fairness. So please indicate to what extent you think that other people consider this payment mechanism as fair. Please indicate your fairness judgment on a scale from 0 (completely unfair) to 10 (completely fair). Important: The other participants in this study make the same judgment as you do. If your judgment matches the rating that is given by the largest number of other participants, you will receive an additional payment of £0.40 (and the same is true for all other participants).

Subsequently, respondents indicated their incentivized choices for the outcome and procedural fairness measures. In the following, we denote the incentivized measures as *HX fairness judgments* (Houser-Xiao).

4.2. Results: Consistency of Different Fairness Measures

The rightmost column of Table 1 shows that the observed pattern of basic overall fairness judgments closely replicates in the new sample. A multivariate regression analysis (see Table A1 in the Appendix) confirms that pay inequality, measured by the Gini coefficient, and tournament incentives hurt overall fairness judgments. As in Study 1, the discrete bonus scheme is judged less negatively than the tournament, but in Study 2, its effect is not significant. Both time advantage and handicap reduce fairness judgments, and we replicate the negative interaction of time advantages with tournament incentives. Thus, Study 1 results on non-incentivized fairness judgments are robust and replicate closely in Study 2.

In Study 2, we can also compare the non-incentivized fairness measures (overall, procedural fairness, and outcome fairness) to the incentivized HX-fairness social norms counterparts. We find that the HX fairness measures are somewhat lower than the non-incentivized judgments in 15 of the 16 treatments for the overall fairness measure, 13 treatments for the outcome fairness measure, and in all 16 treatments for the procedural fairness measure (see Table OS6). However, the basic patterns of relative judgments remain very similar across incentivized and non-incentivized measures.¹⁰ We also find that for all three measures, the variation in non-incentivized fairness judgments is

¹⁰ Vesely (2015) shows that incentivized methods elicit fairness judgments that are indistinguishable from non-incentivized measures. Our results suggest that while relative fairness judgments across conditions are robust with respect to incentivization, levels might be affected by incentives.

significantly larger than the variation in the HX measures. This suggests that participants perceive a clear fairness norm.¹¹

4.3. Results: Analysis of Incentivized Fairness Measures

Having established the consistency of the different fairness measures in Study 1 and Study 2 (non-incentivized and incentivized), we will now focus on the analysis of the incentivized overall HX fairness measure and on the differential effects on the incentivized HX procedural fairness and HX outcome fairness measures.

4.3.1 HX overall fairness

The overall HX fairness measure shows the same pattern as the fairness measure reported in Study 1. In particular, the low-incentive piece rate (T1) elicits the highest HX fairness judgments (8.05), and the high-inequality tournament with time advantage (T14) elicits the lowest judgments (5.49).¹² The multivariate analyses for the overall HX fairness measure in Table 4 closely replicate the findings of Study 1 reported in Table 2. Again, we find a significant negative effect for discrete bonus schemes, smaller than the negative effect of tournaments. Both time handicaps and, more strongly so, time advantages reduce the HX fairness judgment. These direct effects emerge consistently over all specifications. The interaction between tournament incentives and time advantages is again negative, but the standard errors are large.

Conducting subsample analyses, as in Table 3, reveals that low-income (vs. high-income) and more liberal (vs. conservative) respondents are less concerned about time ad-

¹¹ The standard deviation for non-incentivized measures are 2.53 (overall fairness), 2.49 (procedural fairness) and 2.76 (outcome fairness), which are each significantly larger ($p < 0.01$) than the corresponding HX measures of 2.28 (overall fairness), 2.29 (procedural fairness) and 2.55 (outcome fairness).

¹² In the Online Supplement, we show histograms of the full distribution of HX fairness judgments for all 16 treatments.

Table 4. Multivariate analyses of HX fairness judgment – Study 2

	HX Fairness judgment			
	(1)	(2)	(3)	(4)
Gini	-1.03*** (0.14)	-1.03*** (0.14)	-1.03*** (0.14)	-1.03*** (0.14)
Discrete Bonus=1	-0.24* (0.14)	-0.25* (0.14)	-0.28* (0.16)	-0.28* (0.15)
Tournament=1	-0.49*** (0.13)	-0.50*** (0.14)	-0.46*** (0.15)	-0.46*** (0.15)
Time Advantage=1	-0.95*** (0.13)	-0.93*** (0.13)	-0.81*** (0.18)	-0.80*** (0.18)
Time Handicap=1	-0.62*** (0.13)	-0.63*** (0.13)	-0.65*** (0.18)	-0.65*** (0.18)
Tournament x Time Advantage			-0.27 (0.26)	-0.26 (0.26)
Tournament x Time Handicap			0.05 (0.25)	0.04 (0.25)
Exponential Scheme=1	-0.08 (0.22)	-0.09 (0.22)	-0.08 (0.22)	-0.09 (0.22)
Constant	7.91*** (0.11)	8.21*** (0.19)	7.91*** (0.11)	8.20*** (0.19)
Controls	No	Yes	No	Yes
Observations	2,423	2,423	2,423	2,423
R-squared	0.07	0.07	0.07	0.07
<i>F-statistics (Discrete Bonus = Tournament)</i>	5.56**	5.58**	1.04	1.11
<i>F-statistics (Time Bonus = Time Handicap)</i>	5.81**	5.22**	0.82	0.75

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. OLS regression on HX fairness judgment with robust standard errors in parentheses. Piece rate is the reference category. Controls include gender, level of education, personal income (categorical), indicators for ethnic groups and for living in a rural area, political orientation (1: Liberal; 6: Conservative), self-perceived social status, and employment status.

vantages and time handicaps (see Table A2 in the Appendix). Only the latter finding replicates Study 1 results. Females are less concerned about time advantages and more

concerned about time handicaps. The subsample results suggest that the overall patterns are robust and are not driven by certain groups holding specific fairness views.

4.3.2 HX Procedural and Outcome Fairness

We now analyze the role of procedural and outcome fairness in individuals' judgments. We have already seen that outcome inequality in the form of the Gini coefficient has a robust negative effect on fairness judgments. However, several other properties of the payment schemes have different effects on fairness judgments that are not captured by a purely outcome-based perspective.

We first observe that outcome fairness judgments are substantially and significantly lower than procedural fairness judgments in 13 of the 16 treatments and not significantly different in T9, T14, and T15 (see Table OS6). That is, overwhelmingly, the payment schemes are judged more positively from a procedural than from an outcome fairness perspective. Moreover, when regressing the overall HX fairness measure on the HX procedural and HX outcome measures, we find a significantly stronger partial correlation with procedural than with outcome fairness (coefficients of 0.533 versus 0.370, $F=20.59$, $p<0.01$, see Table OS7). This suggests that the more positive procedural judgments also have a more substantial effect on the overall judgment. It supports the view that organizations can benefit from making procedural justice perspectives more salient among their employees (e.g., Trautmann and Wakker, 2010; Schmidt and Trautmann, 2023). This is particularly relevant when one cannot guarantee equal outcomes, for example, because of the need for payment schemes. That is, procedural fairness is easier to attain than outcome fairness, which is also reflected in the higher fairness judgments for procedural fairness across all payment schemes.

Next, we assess how the properties of the payment scheme differentially affect the two fairness properties. To this end, we define the relative distance between procedural and outcome fairness as:

$$\Delta_{\text{pro-out}} = \frac{\text{HX procedural fairness} - \text{HX outcome fairness}}{\text{HX procedural fairness}}$$

Note that $\Delta_{\text{pro-out}}$ increases if outcome fairness is judged relatively lower compared to procedural fairness and decreases if procedural fairness is judged relatively lower compared to outcome fairness for a specific payment scheme.

Table 5. Multivariate analyses of $\Delta_{\text{pro-out}}$

	$\Delta_{\text{pro-out}}$			
	(1)	(2)	(3)	(4)
Gini	0.10** (0.04)	0.10** (0.04)	0.10** (0.04)	0.10** (0.04)
Discrete Bonus=1	-0.01 (0.04)	-0.01 (0.04)	-0.04 (0.05)	-0.04 (0.05)
Tournament=1	-0.03 (0.04)	-0.03 (0.04)	0.01 (0.04)	0.01 (0.04)
Time Advantage=1	-0.09*** (0.04)	-0.10*** (0.04)	-0.01 (0.05)	-0.02 (0.05)
Time Handicap=1	-0.08** (0.04)	-0.08** (0.04)	-0.05 (0.05)	-0.05 (0.05)
Tournament x Time Advantage			-0.16** (0.07)	-0.16** (0.07)
Tournament x Time Handicap			-0.06 (0.07)	-0.07 (0.07)
Exponential scheme=1	-0.02 (0.05)	-0.02 (0.05)	-0.02 (0.05)	-0.02 (0.05)
Constant	0.03 (0.04)	0.02 (0.05)	0.03 (0.04)	0.02 (0.05)
Controls	No	Yes	No	Yes
Observations	2,394	2,394	2,394	2,394
R-squared	0.01	0.01	0.01	0.01
F-statistics (Discrete Bonus = Tournament)	0.54	0.73	1.09	0.99
F-statistics (Time Bonus = Time Handicap)	0.09	0.12	0.57	0.47

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. OLS regression on $\Delta_{\text{pro-out}}$ with robust standard errors in parentheses. Piece rate is the reference category. Controls include gender, level of education, personal income (categorical), indicators for ethnic groups and for living in a rural area, political orientation (1: Liberal; 6: Conservative), self-perceived social status, and employment status.

Table 5 presents the results. This analysis confirms the positive effect of the Gini coefficient on the relative difference between procedural and outcome fairness judgments, which means that higher inequality affects outcome fairness judgments more negatively. We do not observe significant differences between the discrete bonus and tournament scheme indicators. That is, they influence both fairness judgments similarly. In contrast, time advantages and handicaps strongly and negatively affect the difference: they are perceived as more procedurally unfair. For the time advantage, this effect is particularly strong for the tournament condition. The result explains our finding that handicap schemes (i.e., those payment schemes that constrain the resources of more successful workers to level the playing field) are judged negatively despite their desirable effect on equality. Although they positively impact equality, which is good for outcome fairness, they are seen as unfavorable from a procedural fairness point of view.

Although providing additional resources for either the more successful (time advantage) or the less successful (time handicap) is common in many practical settings, they seem to violate basic procedural fairness norms. An advantage or handicap does not just provide higher or lower rewards to different people but intervenes directly in the activities or even the interaction of the workers. This is perceived as undesirable. If such interventions are implemented in practice, our findings suggest that providing a clear justification is essential to prevent adverse fairness effects, as has been discussed in the context of affirmative action for example.

5. Are Third-Parties' Fairness Perceptions Economically Relevant?

Studies 1 and 2 have shown that the properties of a compensation scheme strongly influence how spectators perceive its fairness. We now turn to the question of whether the elicited spectator fairness judgments have behavioral consequences. In particular, we test whether spectators are willing to forgo some of their payment to avoid payment schemes they judge as unfair.

We run an online survey (Study 3) in which student participants rate the fairness of two payment schemes (similar to studies 1 and 2) and can decide to forgo some of their participation fee to implement the fairer scheme in a subsequent laboratory experiment implementing the task as described in the scenarios they assessed. We do this in two different conditions. In the first condition, spectators' decision to implement a fair or unfair scheme has no potential effect on themselves. In the second condition, the spectators may themselves participate in the subsequent lab session for which they decide on the payment scheme. In this way, we can (i) assess whether spectator judgments have behavioral consequences and (ii) whether such consequences differ between impartial and potentially partial judges.

5.1. Sample and Design

5.1.1 Design

Study 3 consists of two experiments. First, we conduct an online experiment similar to Studies 1 and 2. Second, we ran two additional laboratory sessions in which participants actually completed the slider task under one of two possible payment schemes. Respondents in the online experiment of Study 3 have the opportunity to influence the payment scheme of the lab experiment and possibly participate in it.

The online study closely follows the design of Studies 1 and 2 but differs in two aspects: First, respondents now rate two instead of only one payment scheme. The first payment scheme they rate is condition T14 (tournament with high inequality and time advantage). This payment scheme has been rated the most unfair among respondents in the two previous studies (see Table 1). The second scheme is condition T1 (piece rate with low inequality and base pay), which was consistently rated as the fairest payment scheme. Respondents assessed the overall fairness, outcome fairness, and procedural fairness as in Study 2 (all non-incentivized). Second, respondents are faced with another decision after

rating the two payment schemes. They can decide to forgo part of their participation fee of €6 to influence the payment scheme of the described slider task that we implemented in an actual laboratory session. We elicited this decision using an 11-item multiple price list, where respondents can decide to either i) keep the €6 participation fee and implement the unfair tournament scheme or ii) forgo part of their participation fee and implement the fairer piece rate scheme. Each step in the multiple price list corresponds to sacrificing 30 Cents –starting from a participation fee of €6 down to €3– in favor of the piece rate scheme.

Respondents in the online study are randomly assigned to one of two treatments. Respondents in the *impartial* treatment group were informed that participants for the laboratory session would be recruited from the same subject pool but that they themselves would not be able to participate. Respondents in the *partial* treatment, on the other hand, learn that the participants for their laboratory session will be recruited exclusively from online survey participants. Both sets of respondents only know that they will make a decision relevant to one laboratory session; they do not know about the existence of the other treatment and, thus, the second laboratory session. The different treatments are not introduced until the payment scheme decision page. That is, up to this point, instructions and fairness ratings are identical for both treatments, and the treatments thus cannot influence the fairness judgments.

All respondents are informed that one will be randomly selected after the data collection is finished. For the selected respondent, one of the decisions from the multiple price list is then chosen at random and implemented in the laboratory session. Respondents learn (before deciding) that only the randomly selected respondent will have to forgo part of their participation fee (depending on their choices). Otherwise, they receive the €6 participation fee in full. This helps to ensure that respondents make their decision *as-if* they can directly influence the lab sessions' payment scheme without

needing to worry that they might sacrifice part of their endowment without an effect on the laboratory incentives.

5.1.2 Respondents

Respondents for the online experiment of Study 3 were recruited from the subject pool of the AWI lab at Heidelberg University (consisting mostly of students). This ensured that some respondents in the partial condition could be re-invited for the laboratory session. As in the previous studies, respondents first read a short description of the task. They then received the description of the two payment schemes and the expected payments of different workers. After that, they answered four comprehension questions with the same warnings for incorrect answers as in Study 2. In total, 289 respondents participated in the online experiment, 144 in partial and 145 in impartial treatment. On average, 13 percent of the respondents failed to answer some or all comprehension questions correctly for the Tournament condition before moving to the fairness judgments, and 12 percent in the Piece Rate condition. These numbers are somewhat lower than in the Prolific sample in Study 2. As described above, respondents received a participation fee of €6, but for two randomly selected respondents, the actual earnings could be lower, depending on their decisions.

For the laboratory sessions, we recruited a total of 30 participants, 16 in the impartial and 14 in the partial session. Participants in the impartial session were recruited from the AWI subject pool, excluding everyone who participated in the online experiment; participants in the partial session were recruited exclusively from respondents who had previously taken part in the online partial treatment. For both sessions, the randomly selected respondents of the online experiment decided to forgo part of their participation fee to implement the piece rate condition (T1 in Table 1). Including the participation fee, participants in the impartial lab session received an average amount of €10.61, and participants in the partial lab session, on average, earned €11.10.

5.2. Results

Table 6 shows the fairness judgments of the student sample, which replicate the findings of Studies 1 and 2. Respondents find the piece rate much fairer than the tournament incentives (7.81 vs. 4.31, p -value < 0.0001 , two-sided paired sample t -test).¹³ Again, procedural fairness judgments are significantly higher than outcome fairness judgments. Overall, fairness judgments are somewhat lower than in the previous MTurk and Prolific samples.

Table 6. Study 3 – summary of fairness judgments

Payment Scheme	Fairness overall	Procedural fairness	Outcome fairness
Piece Rate: Low with Base Pay (T1)	7.81	8.01	7.27***
Tournament: High with Time Advantage (T14)	4.31	5.30	4.70**

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$ indicate significant differences between procedural and outcome fairness judgment (two-sided paired sample t -test).

Looking at the WTP to implement the fairer piece rate incentive scheme, we find a positive WTP in both the impartial and the partial condition. On average, participants are willing to pay €0.90 in the impartial and €0.77 in the partial conditions. There is no statistical difference in the WTP between the partial and impartial raters.

The histogram in Figure 1 plots the distribution of the WTP, pooling the partial and impartial conditions. We see a tri-modal pattern where about 50 percent of the individuals are basically unwilling to pay a significant amount to implement the fairer payment

¹³ It turned out that the two lab sessions were implemented with the piece-rate scheme. The average fairness judgments was 8.00 (directly elicited before doing the task) and does not differ significantly from the spectator fairness judgments of participants in the online experiment in Study 3 (which are drawn from the same student subject pool, $p=0.827$).

scheme (0 or 0.15 euros), 15 percent are willing to pay 1.05 euros, and about 12 percent are willing to pay 3 euros – the maximum possible WTP allowed in the experiment. Importantly, there is a positive correlation between the WTP and how much fairer the piece rate is compared to the tournament scheme (measured by the difference between the fairness rating of the piece rate minus that of the tournament scheme, Spearman's $\rho = 0.28$, $p\text{-value} < 0.001$): the more unfair an individual considers T14 (tournament) relative to T1 (piece rate) in terms of overall fairness, the more they are willing to forgo to implement T1 in the lab session. The same holds for the difference in procedural (Spearman's $\rho = 0.21$) and outcome fairness (Spearman's $\rho = 0.17$) and WTP ($p\text{-values} < 0.01$).

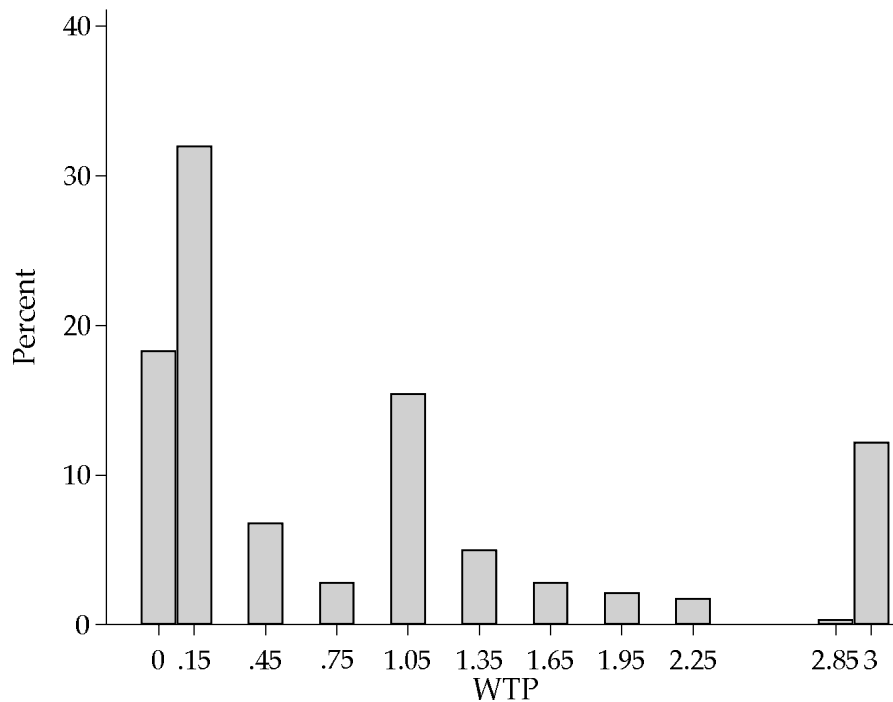


Figure 1. Distribution of the WTP

Study 3 thus shows that the fairness judgments measured in our studies have behavioral consequences and that these are observed both for impartial and potentially partial spectators. That is, people are willing to put their money where their mouth is. In the broader organizational context, we interpret these results as evidence that fairness

judgments are driven by distributional and procedural features of payment schemes and that these judgments may lead to financially relevant behavior.

6. Fairness and the Trade-Off Between Equality and Merit

We have seen that inequality has a detrimental effect on fairness perceptions. Does this negative effect imply that an equal-pay scheme with flat payments that are unrelated to performance differences is most desirable from a fairness point of view? To answer this question, we ran an additional experiment in which respondents assessed the fairness of a piece-rate scheme with low inequality (T1) or a tournament scheme with high inequality (T14) and then a flat payment scheme with a fixed payoff of €6.40. We conducted the study on Prolific with 200 participants. For more details, see Appendix A.3.

The fairness judgments for the tournament and piece rate schemes replicate our previous results. Respondents rate the piece rate as fairer than the tournament (7.98 vs. 5.53). However, respondents rated the flat-payment scheme only slightly better than the tournament, with a score of 5.87.¹⁴ Thus, in the context of the task used in the current paper, a complete absence of performance-based rewards is not perceived as fair, possibly due to violating merit. The negative effect of inequality is conditional on at least some link to performance.

7. Conclusion

Incentive contracts are an essential feature in many employment settings where asymmetric information prevents the payment of fixed, output-independent wages. While incentive contracts are thus often necessary, they also lead to inequality, success, or failure and often to competition between winners and losers (Verhaeghe, 2014). The current

¹⁴ Respondents perceive the flat-payment scheme slightly fairer if they first assessed the tournament scheme than the piece-rate scheme. However, the differences are not significant.

paper systematically shows how these issues affect fairness perceptions of different pay-for-performance schemes.

Our findings suggest that the benefits of providing incentives need to be traded off against unintended side effects due to violation of employees' fairness norms. Such trade-offs are not trivial, with potentially opposing effects of payment scheme features on fairness. For example, contradicting the idea of merit-based payments, steeper incentives are uniformly judged negatively because of their implications for inequality between high and low-skilled workers. On the other hand, a flat scheme unrelated to performance differences is also perceived as unfair. Similarly, we observe that handicaps are perceived as unfair despite leveling the playing field. This may be because flat payments and handicaps contradict some basic notions of procedural fairness and merit that are stronger than the effect of inequality (Dong et al., 2024). Indeed, these views are also captured in popular culture, for example, in Kurt Vonnegut's (1968) famous short story *Harrison Bergeron*, which ridicules the use of handicaps to induce equality at a societal level.

In practice, fairness perceptions will be relevant in determining employee satisfaction, cooperation, and turnover. To prevent unanticipated inefficiencies in employment relations, it is thus important to take these perceptions into account when implementing pay-for-performance schemes. The findings of our study offer some guidance in this respect. First, while pay inequality is a determining factor of fairness perceptions, flat payment schemes may not be feasible, at least in contexts in which effort is unobservable and other non-monetary incentives, such as intrinsic motivation, peer effects, or career concerns, are absent. They are seen as particularly unfair. Second, if incentive contracts cannot be avoided, they should be designed carefully and motivated with reference to procedural fairness. That is, discrete bonuses should be employed only if continuous piece rate rewards are not feasible or practical, and tournaments should be implemented only if individual discrete bonuses may not work, e.g., as in the case of non-contractibility.

Employers should then emphasize the procedural fairness of the incentive scheme. This is particularly true if incentives involve resource allocation effects, as is often the case with promotions. A clear justification for why certain features are warranted may reduce potentially harmful consequences due to the perceived unfairness of the bonus contracts.

We designed our studies to limit the impact of contextual features, stakeholder incentives, experience, and selection on fairness judgments. While this allowed us to focus our analysis on the specific features of the payment schemes, it also poses some limitations. First, we assume a constant mapping of fairness to employee behavior. However, situations may vary in the perceived need for strong incentives, depending on factors such as asymmetric information or misaligned preferences. This perceived need for incentives could shape fairness judgments and influence how fairness affects behavior. Second, suppose workers self-select into industries and positions typically affiliated with different types of incentive contracts (Kosfeld and von Siemens, 2011). In that case, this selection may affect both the average fairness perception by the relevant subgroup and the linkage from fairness perceptions to behavior. Our sample splits according to income, politics, and gender did not reveal a clear picture of how observables might affect such selection. However, future work may fruitfully extend the current design to study the effects of self-selection more carefully and to assess better the role of context and the perceived necessity of incentives on fairness perceptions. Third, previous research has shown that fairness judgments evolve over time with experiences with a payment scheme. Fehr et al. (2020) show that for interested parties, and especially the less successful, tournament incentives are judged very negatively. A broader assessment of how direct experiences possibly affect any of the features considered in the current study may reveal how differences in fairness judgments are amplified or curtailed with repeated exposure and experience of outcomes.

Appendix

A.1. Multivariate Analyses of Study-2 Non-incentivized Overall Fairness Measure

Table A1. Multivariate analyses of fairness judgment – Study 2

	Fairness judgment			
	(1)	(2)	(3)	(4)
Gini	-1.18*** (0.16)	-1.15*** (0.16)	-1.17*** (0.16)	-1.15*** (0.16)
Discrete Bonus=1	-0.21 (0.15)	-0.20 (0.15)	-0.26 (0.17)	-0.26 (0.17)
Tournament=1	-0.52*** (0.15)	-0.49*** (0.15)	-0.47*** (0.17)	-0.43** (0.17)
Time Advantage=1	-1.12*** (0.14)	-1.10*** (0.14)	-0.82*** (0.19)	-0.78*** (0.19)
Time Handicap=1	-0.81*** (0.14)	-0.81*** (0.14)	-0.94*** (0.19)	-0.94*** (0.19)
Tournament x Time Advantage			-0.59** (0.29)	-0.63** (0.29)
Tournament x Time Handicap			0.28 (0.28)	0.25 (0.28)
Exponential Scheme=1	-0.24 (0.24)	-0.25 (0.23)	-0.24 (0.24)	-0.25 (0.23)
Constant	8.31*** (0.12)	8.25*** (0.22)	8.31*** (0.12)	8.24*** (0.22)
Controls	No	Yes	No	Yes
Observations	2,423	2,423	2,423	2,423
R-squared	0.07	0.10	0.08	0.10
<i>F-statistics (Discrete Bonus = Tournament)</i>	7.23***	6.22**	1.20	0.73
<i>F-statistics (Time Bonus = Time Handicap)</i>	4.64**	3.89**	0.38	0.63

Notes: *** p<0.01, ** p<0.05, * p<0.1.

OLS regression on fairness judgments with robust standard errors in parentheses. Piece rate is the reference category. Controls include gender, level of education, personal income (categorical), indicators for ethnic groups and for living in a rural area, political orientation (1: Liberal; 6: Conservative), self-perceived social status, and employment status.

A.2. Subsample Analyses of Study-2 incentivized HX Overall Fairness Measure

Table A2. HX fairness judgments by income, political view, and gender – Study 2

	HX Fairness judgment					
	Income		Political view		Gender	
	(1) High	(2) Low	(3) Conservative	(4) Liberal	(5) Female	(6) Male
Gini	-0.98*** (0.22)	-1.00*** (0.21)	-1.14*** (0.27)	-0.89*** (0.17)	-0.88*** (0.21)	-1.07*** (0.20)
Discrete Bonus=1	-0.18 (0.23)	-0.32 (0.23)	-0.26 (0.28)	-0.34* (0.19)	-0.28 (0.22)	-0.32 (0.22)
Tournament=1	-0.46** (0.22)	-0.46** (0.22)	-0.51** (0.26)	-0.42** (0.19)	-0.50** (0.23)	-0.45** (0.20)
Time Advantage=1	-1.12*** (0.27)	-0.64** (0.27)	-1.11*** (0.34)	-0.62*** (0.22)	-0.57** (0.26)	-0.94*** (0.26)
Time Handicap=1	-0.96*** (0.28)	-0.49** (0.24)	-1.01*** (0.33)	-0.36* (0.21)	-0.94*** (0.26)	-0.36 (0.24)
Tournament x Time Advantage	-0.23 (0.39)	-0.20 (0.38)	-0.03 (0.49)	-0.42 (0.32)	-0.50 (0.38)	-0.10 (0.37)
Tournament x Time Handicap	0.30 (0.39)	-0.08 (0.36)	0.55 (0.46)	-0.31 (0.31)	-0.04 (0.39)	-0.02 (0.34)
Exponential Scheme=1	-0.50 (0.33)	0.34 (0.32)	-0.04 (0.39)	-0.19 (0.27)	-0.25 (0.31)	0.06 (0.31)
Constant	8.28*** (0.35)	8.11*** (0.28)	9.00*** (0.37)	7.96*** (0.23)	8.11*** (0.28)	8.21*** (0.26)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,094	1,078	699	1,635	1,071	1,304
R-squared	0.09	0.07	0.10	0.07	0.08	0.08

Notes: *** p<0.01, ** p<0.05, * p<0.1. OLS regression on fairness judgment with robust standard errors in parentheses. High/Low income, Conservative/Liberal is the median split. Controls include gender, level of education, personal income (categorical, high: ≥ 1.500 \$; not included in regression (1) and (2)), indicators for ethnic groups and for living in a rural area, political orientation (0: Liberal; 1: Conservative, not included in regression (3) and (4)), self-perceived social status, and employment status. Results are robust with or without control variables.

A.3. Scenario without Performance Pay

To probe the limits of the negative effects of inequality, we ran a small-scale experiment with a scenario that entirely eliminates inequality. We conducted the study with 200 participants on Prolific, using a design similar to Study 3.¹⁵ In particular, we first let participants assess either the fairness of the tournament scheme T14 or the piece rate scheme T1 and then let them assess the fairness of a *Flat* payment scheme, described as follows:

The payment mechanism is as follows:

Both workers perform the task individually and independently. The time limit per round to work on the sliders is 120 seconds. Workers receive a flat payment of €1.60 per round of work, independently of the actual number of slides set.

Our data containing more than 600 workers show that workers can correctly place about 20 sliders per round on average within a time limit of 120 seconds. The most skilled workers (top 10%) can place about 29 sliders in 120 seconds on average, while the least skilled workers (bottom 10%) can place about 11 sliders in 120 seconds on average. Irrespective of their actual performance, workers will receive the flat payment of €1.60 per round, amounting to a payment of €6.40 in total. Note also that both workers know their own and the other worker's performance and payoffs in each round and in total.

¹⁵ The experiment took about 10 minutes. Participants received a fixed amount of £2 for their participation.

References

- Almås, I., Cappelen, A., & Tungodden, B. (2020). Cutthroat capitalism versus cuddly socialism: Are Americans more meritocratic and efficiency-seeking than Scandinavians? *Journal of Political Economy*, 128, 1753–1788.
- Ambrose, M.L., Seabright, M. A., & Schminke, M. (2002). Sabotage in the workplace: The role of organizational injustice. *Organizational Behavior and Human Decision Processes*, 89, 947–965.
- Ambuehl, S., Niederle, M., & Roth, A. E. (2015). More money, more problems? Can high pay be coercive and repugnant? *American Economic Review*, 105(5), 357–360.
- Ambuehl, S., & Ockenfels, A. (2017). The ethics of incentivizing the uninformed: A vignette study. *American Economic Review*, 107(5), 91–95.
- Bartling, B., Cappelen, A. W., Ekström, M., Sorensen, E., & Tungodden, B. (2018). Fairness in Winner-Take-All Markets. NHH Dept. of Economics Discussion Paper.
- Bénabou, R. & Jean Tirole, J. (2003). Intrinsic and Extrinsic Motivation, *The Review of Economic Studies*, 70 (3), 489–520.
- Bénabou, R., & Tirole, J. (2016). Bonus culture: Competitive pay, screening, and multitasking. *Journal of Political Economy*, 124(2), 305–370.
- Breza, E., Kaur, S., & Shamdasani, Y. (2018). The morale effects of pay inequality. *Quarterly Journal of Economics*, 133(2), 611–663.
- Bell, B., Pedemonte, S., & Van Reenen, J. (2021). CEO Pay and the Rise of Relative Performance Contracts: A Question of Governance? *Journal of the European Economic Association*, 19, 2513–2542.
- Buser, T., Niederle, M., & Oosterbeek, H. (2014). Gender, Competitiveness, and Career Choices. *The Quarterly Journal of Economics*, 129(3), 1409–1447.
- Card, D., Mas, A., Moretti, E., & Saez, E. (2012). Inequality at Work: The Effect of Peer Salaries on Job Satisfaction, *American Economic Review*, 102(6), 2981–3003.
- Cappelen, A.W., Falch, R., & Tungodden, B. (2020). Fair and Unfair Income Inequality. In: Zimmermann, K. (ed.), *Handbook of Labor, Human Resources and Population Economics*. Springer, Cham.

- Chen, D. L., Schonger, M., & Wickens, C. (2016). oTree—An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9, 88–97.
- Dahl, G., & M. Ransom. (1999). "Does Where You Stand Depend on Where You Sit? Tithing Donations and Self-Serving Beliefs," *American Economic Review* 89, 703–27.
- Dillard, J. F., & Fisher, J. G. (1990). Compensation schemes, skill level, and task performance: An experimental examination. *Decision Sciences*, 21(1), 121–137.
- Dong, Lu, Lingbo Huang, & Jamie W. Lien (2024). "They Never Had a Chance": Unequal Opportunities and Fair Redistributions. *Economic Journal*, forthcoming.
- Eriksson, T., Poulsen, A., & Villeval, M.-C. (2009). Feedback and incentives: Experimental evidence. *Labour Economics*, 16, 679–688.
- Erkal, N., Gangadharan, L., & Koh, B. H. (2018). Monetary and non-monetary incentives in real-effort tournaments. *European Economic Review*, 101, 528-545.
- Fehr, D. (2018). Is increasing inequality harmful? Experimental evidence. *Games and Economic Behavior*, 107, 123–134.
- Fehr, D., & Vollmann, M. (2022). Misperceiving Economic Success: Experimental Evidence on Meritocratic Beliefs and Inequality Acceptance, CESifo Working Paper 9983.
- Fehr, D., Rau, H., Trautmann, S. T., & Xu, Y. (2020). Inequality, fairness and social capital. *European Economic Review*, 129, 103566.
- Finseraas, H., Johnsen, Å. A., Kotsadam, A., & Torsvik, G. (2016). Exposure to female colleagues breaks the glass ceiling—Evidence from a combined vignette and field experiment. *European Economic Review*, 90, 363–374.
- Fulmer, I. S., & Shaw, J. D. (2018). Person-Based Differences in Pay Reactions: A Compensation-Activation Theory and Integrative Conceptual Review. *Journal of Applied Psychology*, 103, 939–958.
- Gill, D., & Prowse, V. (2012). A structural analysis of disappointment aversion in a real effort competition. *American Economic Review*, 102, 469–503.
- Greenberg, J. (1987). Reactions to Procedural Injustice in Payment Distributions: Do the Means Justify the Ends? *Journal of Applied Psychology* 72, 55-61.

- Gupta, N., & Shaw, J. D. (2014). Employee compensation: The neglected area of HRM research. *Human Resource Management Review* 24, 1–4.
- Harbring, C., Irlenbusch, B., Kräkel, M., & Selten, R. (2007). Sabotage in corporate contests—an experimental analysis. *International Journal of the Economics of Business*, 14(3), 367-392.
- Houser, D. & Xiao, E. (2011). Classification of natural language messages using a coordination game. *Experimental Economics* 14, 1–14.
- Jawahar, I. M. & Stone, T. (2011). Fairness Perceptions and Satisfaction with Components of Pay Satisfaction. *Journal of Managerial Psychology*, 26, 297–312.
- Kahneman, D., Knetsch, J., & R. Thaler. (1986). Fairness as a Constraint on Profit Seeking: Entitlements in the Market," *American Economic Review* 76, 728–41.
- Kao, A.C., Jager, A. J., Koenig, B. A., Moller, A. C., Tutty, M.A., Williams, G. C., & Wright, S. M. (2018). Physician Perception of Pay Fairness and its Association with Work Satisfaction, Intend to Leave Practice, and Personal Health. *Journal of General Internal Medicine*, 33, 812–817.
- Kleinlercher, D., & Stöckl, T. (2018). On the provision of incentives in finance experiments. *Experimental Economics*, 21(1), 154-179.
- Konow, J. (2003). Which is the fairest one of all? A positive analysis of justice theories. *Journal of Economic Literature*, 41(4), 1188–1239.
- Kosfeld, M., & Von Siemens, F. A. (2011). Competition, cooperation, and corporate culture. *RAND Journal of Economics*, 42(1), 23–43.
- Kőszegi, B. (2014). Behavioral Contract Theory. *Journal of Economic Literature*, 52, 1075–1118.
- Kübler, D., Schmid, J., & Stüber, R. (2018). Gender Discrimination in Hiring Across Occupations: A Nationally-Representative Vignette Study. *Labour Economics*, 55, 215–229.
- Lazear, E. P. (2018). Compensation and Incentives in the Workplace. *Journal of Economic Perspectives*, 32(3), 195-214.
- Milgrom, P. & Roberts, J. (1992). *Economics, Organization & Management*. Pearson.

- Murayama, K., & Elliot, A. J. (2012). The competition–performance relation: A meta-analytic review and test of the opposing processes model of competition and performance. *Psychological Bulletin*, 138(6), 1035.
- Niederle, M., & Vesterlund, L. (2007). Do Women Shy Away From Competition? Do Men Compete Too Much? *The Quarterly Journal of Economics*, 122(3), 1067–1101.
- Ors, E., Palomino, F., & Peyrache, E. (2013). Performance gender gap: does competition matter? *Journal of Labor Economics*, 31(3), 443–499.
- Oyer, P. (2000). A theory of sales quotas with limited liability and rent sharing. *Journal of Labor Economics*, 18(3), 405–426.
- Prendergast, C. (1999). The Provision of Incentives in Firms. *Journal of Economic Literature*, 37, 7–63.
- Roemer, J. E. (1998). *Equality of opportunity*. Harvard University Press.
- Samek, A., Kapteyn, A., & Gray, A. (2020). Using Vignettes to Improve Understanding of Social Security and Annuities, *NBER Working Paper No. w26176*.
- Schildberg-Hörisch, H., Trieu, C., & Willrodt, J. (2020). Perceived Fairness and Consequences of Affirmative Action Policies, *IZA Discussion Paper 13202*.
- Schmidt, R. & Trautmann, S.T. (2023). Implementing (Un)fair Procedures? Favoritism and Process Fairness when Inequality is Inevitable. *Journal of Law, Economics and Organization*, 39, 199–234.
- Trautmann, S. T. (2023). Procedural fairness and equality of opportunity. *Journal of Economic Surveys*, 37, 1697–1714.
- Trautmann, S. T. & Wakker, P. P. (2010). Process Fairness and Dynamic Consistency. *Economics Letters* 109, 187–189.
- Verhaeghe, P. (2014). *What about me? The struggle for identity in a market-based society*, Scribe, London, UK.
- Vesely, S. (2015). Elicitation of normative and fairness judgments: Do incentives matter? *Judgment and Decision Making* 10, 191–197.
- Vonnegut, K. (1968). *Welcome to the Monkey House*. Dial Press, NY.

Xu, Y., & Marandola, G. (2023). The (Negative) Effects of Inequality on Social Capital.
Journal of Economic Surveys 37, 1562–1588.

ONLINE SUPPLEMENT

Contents

OS.1 INSTRUCTIONS

OS.2 STUDY 1: ATTRITION AND BALANCE STATISTICS

OS.3 STUDY 1: HISTOGRAMS OF FAIRNESS JUDGMENTS

OS.4 STUDY 2: ATTRITION AND BALANCE STATISTICS

OS.5 STUDY 2: HISTOGRAMS OF OVERALL FAIRNESS JUDGMENTS AND HX
FAIRNESS JUDGMENTS

OS6. STUDY 2: SUBSAMPLE ANALYSES BY TASK COMPREHENSION

OS7. STUDY 2: SUMMARY OF FAIRNESS MEASURES

OS8. STUDY 2: COMPOSITION OF OVERALL FAIRNESS MEASURES

OS9. ROBUSTNESS CHECK: BINARIZED MEASURES

OS10. ROBUSTNESS CHECK: Excluding GINI

OS.1 INSTRUCTIONS

Note: We show the correct answers to the comprehension questions in curly brackets after each question.

General Instructions

Dear participant,

Thank you for taking part in the questionnaire. Please first read all information carefully, and then answer the question.

Two workers have to perform a task on a computer. They will be paid according to some payment scheme, which will be described in detail later. The task requires the workers to place a slider in the central position of a slider bar. The slider initially appears at the far-left position in the bar. One has to use the mouse to move the slider to the target position.

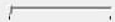
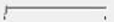
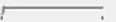
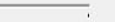
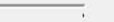
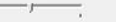
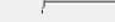
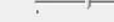
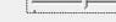
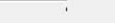

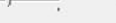
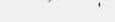
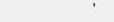
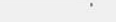
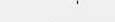

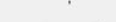
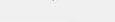
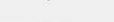
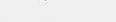

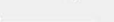
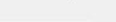
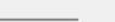
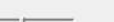
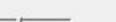
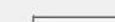
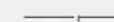
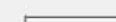
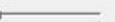

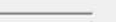



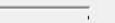
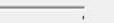

Initial position:



Target position:



The keyboard has been disabled in order to make the task sufficiently challenging. The task consists of 48 sliders to be correctly placed. Placing 48 sliders within a time limit (described later) is called a round. The task is repeated for *four* rounds. The actual screen the workers have seen is like the following.

	0		0		0
	0		0		50
	0		50		50
	0		50		50
	50		0		0
	0		0		50
	0		0		50
	0		63		50
	50		0		0
	0		50		50
	0		50		0
	0		50		0
	0		0		0

Payment scheme: Piece rate

Low inequality with base pay (Low_with base pay)

The payment mechanism is as follows:

Both workers perform the task individually and independently. The time limit per round to work on the sliders is 120 seconds. For each correctly positioned slider within the time limit per round, workers receive \$0.03. Additionally, they also receive \$1 each round for participating in the task. Their payment per round can be summarized as $\$1 + \$0.03 \times (\text{Number of sliders})$.

Our data containing more than 600 workers show that workers can correctly place about 20 sliders per round on average within a time limit of 120 seconds. An average-skilled worker would, therefore, receive \$1.60 ($=\$0.03 \times 20 + \1) per round. The total earnings after four rounds would be \$6.40 ($=\1.60×4) for an average-skilled worker.

The most skilled workers (top 10%) can place about 29 sliders in 120 seconds on average, amounting to a payment of \$1.87 ($=\$0.03 \times 29 + \1) per round, or \$7.48 in total. The least skilled workers (bottom 10%) can place about 11 sliders in 120 seconds on average, amounting to a payment of \$1.33 ($=\$0.03 \times 11 + \1) per round, or \$5.32 in total. Thus, the difference in total earnings between the two workers is \$2.16 after four rounds if one worker is skilled and the other is not. Note also that both workers know their own and the other worker's performance and payoffs in each round and in total.

i) If all aspects of the task and the payment of workers are clear, please answer the following questions about the task:

Assume that worker A placed 13 sliders correctly and worker B placed 27 sliders correctly in round 1.

What are the round 1 earnings of worker A (in \$)? _____ {1.39}

What are the round 1 earnings of worker B (in \$)? _____ {1.81}

How many seconds does worker A have available for the task in round 2? _____ {120}

How many seconds does worker B have available for the task in round 2? _____ {120}

ii) Let us now consider the overall picture of the payment mechanism described above. Different payment mechanisms to reward people for their work may be considered more or less fair. We are interested in how people think about the above described mechanism. To what extent do you think this payment mechanism is fair? Please indicate your fairness judgment on a scale from 0 (completely unfair) to 10 (completely fair).

Completely unfair ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Completely fair

Payment scheme: Piece rate

Medium inequality without base pay (Medium No Base Pay)

The payment mechanism is as follows:

Both workers perform the task individually and independently. The time limit per round to work on the sliders is 120 seconds. For each correctly positioned slider within the time limit per round, workers receive \$0.08. Their payment per round can be summarized as $\$0.08 \times (\text{Number of sliders})$.

Our data containing more than 600 workers show that workers can correctly place about 20 sliders per round on average within a time limit of 120 seconds. An average-skilled worker would, therefore, receive \$1.60 ($=\0.08×20) per round. The total earnings after four rounds would be \$6.40 ($=\1.60×4) for an average-skilled worker.

The most skilled workers (top 10%) can place about 29 sliders in 120 seconds on average, amounting to a payment of \$2.32 ($=\0.08×29) per round, or \$9.28 in total. The least skilled workers (bottom 10%) can place about 11 sliders in 120 seconds on average, amounting to a payment of \$0.88 ($=\0.08×11) per round, or \$3.52 in total. Thus, the difference in total earnings between the two workers is \$5.76 after four rounds if one worker is skilled and the other is not. Note also that both workers know their own and the other worker's performance and payoffs in each round and in total.

i) If all aspects of the task and the payment of workers are clear, please answer the following questions about the task:

Assume that worker A placed 13 sliders correctly and worker B placed 27 sliders correctly in round 1.

What are the round 1 earnings of worker A (in \$)? _____ {1.04}

What are the round 1 earnings of worker B (in \$)? _____ {2.16}

How many seconds does worker A have available for the task in round 2? _____ {120}

How many seconds does worker B have available for the task in round 2? _____ {120}

ii) Let us now consider the overall picture of the payment mechanism described above. Different payment mechanisms to reward people for their work may be considered more or less fair. We are interested in how people think about the above described mechanism. To what extent do you think this payment mechanism is fair? Please indicate your fairness judgment on a scale from 0 (completely unfair) to 10 (completely fair).

Completely unfair ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Completely fair

Payment scheme: Piece rate

High inequality without base pay but with an entrance fee (High with Entrance Fee)

The payment mechanism is as follows:

Both workers perform the task individually and independently. The time limit per round to work on the sliders is 120 seconds. The task requires a mandatory participation fee of \$6 to be paid by each worker (or \$1.50 per round). For each correctly positioned slider within the time limit per round, workers receive \$0.15. Their payment per round can be summarized as $\$0.15 \times (\text{Number of sliders})$.

Our data containing more than 600 workers show that workers can correctly place about 20 sliders per round on average within a time limit of 120 seconds. An average-skilled worker would, therefore, receive \$1.50 ($=\$0.15 \times 20 - \1.50) per round. The total earnings after four rounds would be \$6 ($=\1.50×4) for an average-skilled worker (already accounting for the participation fee).

The most skilled workers (top 10%) can place about 29 sliders in 120 seconds on average, amounting to a payment of \$2.85 ($=\$0.15 \times 29 - \1.50) per round, or \$11.40 in total (already accounting for the participation fee). The least skilled workers (bottom 10%) can place about 11 sliders in 120 seconds on average, amounting to a payment of \$0.15 ($=\$0.15 \times 11 - \1.50) per round, or \$0.60 in total (already accounting for the participation fee). Thus, the difference in total earnings between the two workers is \$10.80 after four rounds if one worker is skilled and the other is not. Note also that both workers know their own and the other worker's performance and payoffs in each round and in total.

i) If all aspects of the task and the payment of workers are clear, please answer the following questions about the task:

Assume that worker A placed 13 sliders correctly and worker B placed 27 sliders correctly

in round 1. Please take into account the entrance fee of \$1.50 per round.

What are the round 1 earnings of worker A (in \$)? _____ {0.45}

What are the round 1 earnings of worker B (in \$)? _____ {2.55}

How many seconds does worker A have available for the task in round 2? _____ {120}

How many seconds does worker B have available for the task in round 2? _____ {120}

ii) Let us now consider the overall picture of the payment mechanism described above. Different payment mechanisms to reward people for their work may be considered more or less fair. We are interested in how people think about the above described mechanism. To what extent do you think this payment mechanism is fair? Please indicate your fairness judgment on a scale from 0 (completely unfair) to 10 (completely fair).

Completely unfair ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Completely fair

Payment scheme: Piece rate

Extreme inequality, no base pay, no entrance fee, exponential reward function (Exponential)

The payment mechanism is as follows:

Both workers perform the task individually and independently. The time limit per round to work on the sliders is 120 seconds. The total earnings for the task in a round depend on the number of sliders workers can correctly place within the time limit, which is summarized in the table below. The upper rows show the number of correctly placed sliders and the lower rows show the total payoffs in dollar.

Sliders	1	2	3	4	5	6	7	8	9	10	11	12
Total	0.01	0.03	0.04	0.06	0.08	0.11	0.14	0.17	0.21	0.26	0.32	0.38
Payoffs												

Sliders	13	14	15	16	17	18	19	20	21	22	23	24
Total	0.46	0.55	0.66	0.78	0.92	1.09	1.29	1.52	1.79	2.11	2.48	2.91
Payoffs												
Sliders	25	26	27	28	29	30	31	32	33	34	35	36
Total	3.42	4.01	4.70	5.52	6.46	7.58	8.87	10.39	12.17	14.25	16.69	19.54
Payoffs												
Total	37	38	39	40	41	42	43	44	45	46	47	48
Sliders												
Payoffs	22.87	26.77	31.34	36.67	42.92	50.23	58.78	68.78	80.49	94.18	110.21	128.95

Our data containing more than 600 workers show that workers can correctly place about 20 sliders per round on average within a time limit of 120 seconds. An average-skilled worker would therefore receive \$1.52 per round for doing the task (see the table). The total earnings after four rounds would be \$6.08 for an average-skilled worker.

The most skilled workers (top 10%) can place about 29 sliders in 120 seconds on average, amounting to a payment of \$6.46 (see table) per round, or \$25.84 in total. The least skilled workers (bottom 10%) can place about 11 sliders in 120 seconds on average, amounting to a payment of \$0.32 (see table) per round, or \$1.28 in total. Thus, the difference in total earnings between the two workers is \$24.56 after four rounds if one worker is skilled and the other is not. Note also that both workers know their own and the other worker's performance and payoffs in each round and in total.

i) If all aspects of the task and the payment of workers are clear, please answer the

following questions about the task:

Assume that worker A placed 13 sliders correctly and worker B placed 27 sliders correctly in round 1.

What are the round 1 earnings of worker A (in \$)? _____ {0.46}

What are the round 1 earnings of worker B (in \$)? _____ {4.70}

How many seconds does worker A have available for the task in round 2? _____ {120}

How many seconds does worker B have available for the task in round 2? _____ {120}

ii) Let us now consider the overall picture of the payment mechanism described above. Different payment mechanisms to reward people for their work may be considered more or less fair. We are interested in how people think about the above described mechanism. To what extent do you think this payment mechanism is fair? Please indicate your fairness judgment on a scale from 0 (completely unfair) to 10 (completely fair).

Completely unfair 0 0 0 0 0 0 0 0 0 0 0 Completely fair

Payment scheme: Discrete bonus target

Low inequality, no time advantage, no time handicap (Low)

The payment mechanism is as follows:

Both workers perform the task individually and independently. The time limit per round to work on the sliders is 120 seconds. For each correctly positioned slider within the time limit per round, workers receive \$0.02. Additionally, in each round, there will also be a bonus payment of \$1.40 if a worker correctly positioned 21 sliders or more. If the total number of sliders placed is below 21, then the bonus payment is \$1.20.

Our data containing more than 600 workers show that workers can correctly place about 20 sliders per round on average within a time limit of 120 seconds. If an average-skilled worker just manages to place 21 sliders in a round, his payoff would be \$1.82 ($=\$0.02 \cdot 21 + \1.40) this round. If, on the other hand, he just fails to get 21 sliders in a round, his payoff would be \$1.60 ($=\$0.02 \cdot 20 + \1.20) for this round. Suppose that he manages to correctly place 21 sliders in two rounds, but only places 20 sliders in the other two rounds. The total earnings after four rounds would then be \$6.84 ($=\$1.82 \cdot 2 + \$1.60 \cdot 2$) for the average-skilled worker.

The most skilled workers (top 10%) can place about 29 sliders in 120 seconds on average, amounting to a payment of \$1.98 ($=\$0.02 \cdot 29 + \1.40) per round, or \$7.92 in total. The least skilled workers (bottom 10%) can place about 11 sliders in 120 seconds on average, amounting to a payment of \$1.42 ($=\$0.02 \cdot 11 + \1.20) per round, or \$5.68 in total. Thus, the difference in total earnings between the two workers is \$2.24 after four rounds if one worker is skilled and the other is not. Note also that both workers know their own and the other worker's performance and payoffs in each round and in total.

i) If all aspects of the task and the payment of workers are clear, please answer the

following questions about the task:

Assume that worker A placed 13 sliders correctly and worker B placed 27 sliders correctly in round 1.

What are the round 1 earnings of worker A (in \$)? _____ {1.46}

What are the round 1 earnings of worker B (in \$)? _____ {1.94}

How many seconds does worker A have available for the task in round 2? _____ {120}

How many seconds does worker B have available for the task in round 2? _____ {120}

ii) Let us now consider the overall picture of the payment mechanism described above. Different payment mechanisms to reward people for their work may be considered more or less fair. We are interested in how people think about the above described mechanism. To what extent do you think this payment mechanism is fair? Please indicate your fairness judgment on a scale from 0 (completely unfair) to 10 (completely fair).

Completely unfair ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Completely fair

Payment scheme: Discrete bonus target

High inequality, no time advantage, no time handicap (High)

The payment mechanism is as follows:

Both workers perform the task individually and independently. The time limit per round to work on the sliders is 120 seconds. For each correctly positioned slider within the time limit per round, workers receive \$0.02. Additionally, in each round, there will also be a bonus payment of \$2.50 if a worker correctly positioned 21 sliders or more. If the total number of sliders placed is below 21, then the bonus payment is \$0.00.

Our data containing more than 600 workers show that workers can correctly place about 20 sliders per round on average within a time limit of 120 seconds. If an average-skilled worker just manages to place 21 sliders in a round, his payoff would be \$2.92 ($=\$0.02 \cdot 21 + \2.50) this round. If, on the other hand, he just fails to get 21 sliders in a round, his payoff would be \$0.40 ($=\$0.02 \cdot 20$) for this round. Suppose that he manages to correctly place 21 sliders in two rounds, but only places 20 sliders in the other two rounds. The total earnings after four rounds would then be \$6.64 ($=\$2.92 \cdot 2 + \$0.40 \cdot 2$) for the average-skilled worker.

The most skilled workers (top 10%) can place about 29 sliders in 120 seconds on average, amounting to a payment of \$3.08 ($=\$0.02 \cdot 29 + \2.50) per round, or \$12.32 in total.

The least skilled workers (bottom 10%) can place about 11 sliders in 120 seconds on average, amounting to a payment of \$0.22 ($=\0.02×11) per round, or \$0.88 in total. Thus, the difference in total earnings between the two workers is \$11.44 after four rounds if one worker is skilled and the other is not. Note also that both workers know their own and the other worker's performance and payoffs in each round and in total.

i) If all aspects of the task and the payment of workers are clear, please answer the following questions about the task:

Assume that worker A placed 13 sliders correctly and worker B placed 27 sliders correctly in round 1.

What are the round 1 earnings of worker A (in \$)? _____ {0.26}

What are the round 1 earnings of worker B (in \$)? _____ {3.04}

How many seconds does worker A have available for the task in round 2? _____ {120}

How many seconds does worker B have available for the task in round 2? _____ {120}

ii) Let us now consider the overall picture of the payment mechanism described above. Different payment mechanisms to reward people for their work may be considered more or less fair. We are interested in how people think about the above described mechanism. To what extent do you think this payment mechanism is fair? Please indicate your fairness judgment on a scale from 0 (completely unfair) to 10 (completely fair).

Completely unfair ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Completely fair

Payment scheme: Discrete bonus target

Low inequality, with time advantage (Low with Time Bonus)

The payment mechanism is as follows:

Both workers perform the task individually and independently. For each correctly positioned slider within the time limit per round, workers receive \$0.02. Additionally, in each round, there will also be a bonus payment of \$1.40 if a worker correctly positioned 21 sliders or more. If the total number of sliders placed is below 21, then the bonus payment is \$1.20.

The initial time limit in the first round is 120 seconds. Depending on performance, the time budget is reduced or increased in the subsequent round. In particular, if the worker correctly places 21 or more sliders, 6 seconds are added to his current time budget. If the worker correctly places less than 21 sliders, 6 seconds are subtracted from his current time budget. See the example for illustration.

Example: A worker places 20 sliders correctly in round 1. His time budget in round 2 will be 114 seconds. In round 2 he then places 21 sliders correctly. His time budget in round 3 will be 120 seconds. In round 3 he then places 22 sliders correctly. His time budget in round 4 will be 126 seconds.

In general, a shorter time budget to work on the task in a round makes it more difficult, and a longer time budget makes it less difficult, to correctly place many sliders and to reach the threshold for the bonus payment, in subsequent rounds.

Our data containing more than 600 workers show that workers can correctly place about 20 sliders per round on average within a time limit of 120 seconds. If an average-skilled worker just manages to place 21 sliders in a round, his payoff would be \$1.82 ($=\$0.02 \cdot 21 + \1.40) this round. If, on the other hand, he just fails to get 21 sliders in a round, his payoff would be \$1.60 ($=\$0.02 \cdot 20 + \1.20) for this round. Suppose that he manages to correctly place 21 sliders in two rounds, but only places 20 sliders in the other two rounds. The total earnings after four rounds would then be \$6.84 ($=\$1.82 \cdot 2 + \$1.60 \cdot 2$) for the average-skilled worker.

The most skilled workers (top 10%) can place about 29 sliders in 120 seconds on average, amounting to a payment of \$1.98 ($=\$0.02 \cdot 29 + \1.40) per round, or \$7.92 in total. The least skilled workers (bottom 10%) can place about 11 sliders in 120 seconds on average, amounting to a payment of \$1.42 ($=\$0.02 \cdot 11 + \1.20) per round, or \$5.68 in total. Thus, the difference in total earnings between the two workers is \$2.24 after four rounds if one worker is skilled and the other is not.

Note that these calculations do not account for increased or reduced time budgets due to managing or not managing to reach the threshold of 21 sliders in earlier rounds. This will lead, on average, to larger numbers of sliders compared to the number for 120 seconds for those with more time after passing the threshold, and on average lower numbers of sliders compared to the number for 120 seconds for those with less time after not passing the threshold. Note also that both workers know their own and the other worker's performance and payoffs in each round and in total.

i) If all aspects of the task and the payment of workers are clear, please answer the following questions about the task:

Assume that worker A placed 13 sliders correctly and worker B placed 27 sliders correctly in round 1.

What are the round 1 earnings of worker A (in \$)? _____ {1.46}

What are the round 1 earnings of worker B (in \$)? _____ {1.94}

How many seconds does worker A have available for the task in round 2? _____ {114}

How many seconds does worker B have available for the task in round 2? _____ {126}

ii) Let us now consider the overall picture of the payment mechanism described above. Different payment mechanisms to reward people for their work may be considered more

or less fair. We are interested in how people think about the above described mechanism.

To what extent do you think this payment mechanism is fair? Please indicate your fairness judgment on a scale from 0 (completely unfair) to 10 (completely fair).

Completely unfair ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Completely fair

Payment scheme: Discrete bonus target

High inequality, with time advantage (High with Time Bonus)

The payment mechanism is as follows:

Both workers perform the task individually and independently. For each correctly positioned slider within the time limit per round, workers receive \$0.02. Additionally, in each round, there will also be a bonus payment of \$2.50 if a worker correctly positioned 21 sliders or more. If the total number of sliders placed is below 21, then the bonus payment is \$0.00.

The initial time limit in the first round is 120 seconds. Depending on performance, the time budget is reduced or increased in the subsequent round. In particular, if the worker correctly places 21 or more sliders, 6 seconds are added to his current time budget. If the worker correctly places only less than 21 sliders, 6 seconds are subtracted from his current time budget. See the example for illustration.

Example: A worker places 20 sliders correctly in round 1. His time budget in round 2 will be 114 seconds. In round 2 he then places 21 sliders correctly. His time budget in round 3 will be 120 seconds. In round 3 he then places 22 sliders correctly. His time budget in round 4 will be 126 seconds.

In general, a shorter time budget to work on the task in a round makes it more difficult, and a longer time budget makes it less difficult, to correctly place many sliders and to reach the threshold for the bonus payment, in subsequent rounds.

Our data containing more than 600 workers show that workers can correctly place about 20 sliders per round on average within a time limit of 120 seconds. If an average-skilled worker just manages to place 21 sliders in a round, his payoff would be \$2.92 ($=\$0.02 \cdot 21 + \2.50) this round. If, on the other hand, he just fails to get 21 sliders in a round, his payoff would be \$0.40 ($=\$0.02 \cdot 20$) for this round. Suppose that he manages to correctly

place 21 sliders in two rounds, but only places 20 sliders in the other two rounds. The total earnings after four rounds would then be \$6.64 ($=\$2.92*2+\$0.40*2$) for the average-skilled worker.

The most skilled workers (top 10%) can place about 29 sliders in 120 seconds on average, amounting to a payment of \$3.08 ($=\$0.02*29+\2.50) per round, or \$12.32 in total. The least skilled workers (bottom 10%) can place about 11 sliders in 120 seconds on average, amounting to a payment of \$0.22 ($=\$0.02*11$) per round, or \$0.88 in total. Thus, the difference in total earnings between the two workers is \$11.44 after four rounds if one worker is skilled and the other is not.

Note that these calculations do not account for increased or reduced time budgets due to managing or not managing to reach the threshold of 21 sliders in earlier rounds. This will lead, on average, to larger numbers of sliders compared to the number for 120 seconds for those with more time after passing the threshold, and on average lower numbers of sliders compared to the number for 120 seconds for those with less time after not passing the threshold. Note also that both workers know their own and the other worker's performance and payoffs in each round and in total.

i) If all aspects of the task and the payment of workers are clear, please answer the following questions about the task:

Assume that worker A placed 13 sliders correctly and worker B placed 27 sliders correctly in round 1.

What are the round 1 earnings of worker A (in \$)? _____ {0.26}

What are the round 1 earnings of worker B (in \$)? _____ {3.04}

How many seconds does worker A have available for the task in round 2? _____ {114}

How many seconds does worker B have available for the task in round 2? _____ {126}

ii) Let us now consider the overall picture of the payment mechanism described above. Different payment mechanisms to reward people for their work may be considered more or less fair. We are interested in how people think about the above described mechanism. To what extent do you think this payment mechanism is fair? Please indicate your fairness judgment on a scale from 0 (completely unfair) to 10 (completely fair).

Completely unfair ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Completely fair

Payment scheme: Discrete bonus target

Low inequality, with time handicap (Low with Time Handicap)

The payment mechanism is as follows:

Both workers perform the task individually and independently. For each correctly positioned slider within the time limit per round, workers receive \$0.02. Additionally, in each round, there will also be a bonus payment of \$1.40 if a worker correctly positioned 21 sliders or more. If the total number of sliders placed is below 21, then the bonus payment is \$1.20.

The initial time limit in the first round is 120 seconds. Depending on performance, the time budget is reduced or increased in the subsequent round. In particular, if the worker correctly places 21 or more sliders, 6 seconds are subtracted from his current time budget, a time handicap. If the worker correctly places less than 21 sliders, 6 seconds are added to his current time budget, a time benefit. See the example for illustration.

Example: A worker places 20 sliders correctly in round 1. His time budget in round 2 will be 126 seconds. In round 2 he then places 21 sliders correctly. His time budget in round 3 will be 120 seconds. In round 3 he then places 22 sliders correctly. His time budget in round 4 will be 114 seconds.

In general, a shorter time budget to work on the task in a round makes it more difficult, and a longer time budget makes it less difficult, to correctly place many sliders and to reach the threshold for the bonus payment, in subsequent rounds.

Our data containing more than 600 workers show that workers can correctly place about 20 sliders per round on average within a time limit of 120 seconds. If an average-skilled worker just manages to place 21 sliders in a round, his payoff would be \$1.82 ($=\$0.02 \cdot 21 + \1.40) this round. If, on the other hand, he just fails to get 21 sliders in a round, his payoff would be \$1.60 ($=\$0.02 \cdot 20 + \1.20) for this round. Suppose that he manages to correctly place 21 sliders in two rounds, but only places 20 sliders in the other two rounds. The total earnings after four rounds would then be \$6.84 ($=\$1.82 \cdot 2 + \$1.60 \cdot 2$) for the average-skilled worker.

The most skilled workers (top 10%) can place about 29 sliders in 120 seconds on average, amounting to a payment of \$1.98 ($=\$0.02 \cdot 29 + \1.40) per round, or \$7.92 in total. The least skilled workers (bottom 10%) can place about 11 sliders in 120 seconds on average, amounting to a payment of \$1.42 ($=\$0.02 \cdot 11 + \1.20) per round, or \$5.68 in total. Thus, the difference in total earnings between the two workers is \$2.24 after four rounds if one worker is skilled and the other is not.

Note that these calculations do not account for increased or reduced time budgets due to managing or not managing to reach the threshold of 21 sliders in earlier rounds. This will lead, on average, to lower numbers of sliders compared to the number for 120 seconds for those with less time after passing the threshold, and on average larger numbers of sliders compared to the number for 120 seconds for those with more time after not passing the threshold. Note also that both workers know their own and the other worker's performance and payoffs in each round and in total.

i) If all aspects of the task and the payment of workers are clear, please answer the following questions about the task:

Assume that worker A placed 13 sliders correctly and worker B placed 27 sliders correctly in round 1.

What are the round 1 earnings of worker A (in \$)? _____ {1.46}

What are the round 1 earnings of worker B (in \$)? _____ {1.94}

How many seconds does worker A have available for the task in round 2? _____ {126}

How many seconds does worker B have available for the task in round 2? _____ {114}

ii) Let us now consider the overall picture of the payment mechanism described above. Different payment mechanisms to reward people for their work may be considered more or less fair. We are interested in how people think about the above described mechanism. To what extent do you think this payment mechanism is fair? Please indicate your fairness judgment on a scale from 0 (completely unfair) to 10 (completely fair).

Completely unfair ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Completely fair

Payment scheme: Discrete bonus target

High inequality, with time handicap (High with Time Handicap)

The payment mechanism is as follows:

Both workers perform the task individually and independently. For each correctly positioned slider within the time limit per round, workers receive \$0.02. Additionally, in each round, there will also be a bonus payment of \$2.50 if a worker correctly positioned 21 sliders or more. If the total number of sliders placed is below 21, then the bonus payment is \$0.00.

The initial time limit in the first round is 120 seconds. Depending on performance, the

time budget is reduced or increased in the subsequent round. In particular, if the worker correctly places 21 or more sliders, 6 seconds are subtracted from his current time budget, a time handicap. If the worker correctly places less than 21 sliders, 6 seconds are added to his current time budget, a time benefit. See the example for illustration.

Example: A worker places 20 sliders correctly in round 1. His time budget in round 2 will be 126 seconds. In round 2 he then places 21 sliders correctly. His time budget in round 3 will be 120 seconds. In round 3 he then places 22 sliders correctly. His time budget in round 4 will be 114 seconds.

In general, a shorter time budget to work on the task in a round makes it more difficult, and a longer time budget makes it less difficult, to correctly place many sliders and to reach the threshold for the bonus payment, in subsequent rounds.

Our data containing more than 600 workers show that workers can correctly place about 20 sliders per round on average within a time limit of 120 seconds. If an average-skilled worker just manages to place 21 sliders in a round, his payoff would be \$2.92 ($=\$0.02 \times 21 + \2.50) this round. If, on the other hand, he just fails to get 21 sliders in a round, his payoff would be \$0.40 ($=\0.02×20) for this round. Suppose that he manages to correctly place 21 sliders in two rounds, but only places 20 sliders in the other two rounds. The total earnings after four rounds would then be \$6.64 ($=\$2.92 \times 2 + \0.40×2) for the average-skilled worker.

The most skilled workers (top 10%) can place about 29 sliders in 120 seconds on average, amounting to a payment of \$3.08 ($=\$0.02 \times 29 + \2.50) per round, or \$12.32 in total. The least skilled workers (bottom 10%) can place about 11 sliders in 120 seconds on average, amounting to a payment of \$0.22 ($=\0.02×11) per round, or \$0.88 in total. Thus, the difference in total earnings between the two workers is \$11.44 after four rounds if one worker is skilled and the other is not.

Note that these calculations do not account for increased or reduced time budgets due to managing or not managing to reach the threshold of 21 sliders in earlier rounds. This will lead, on average, to lower numbers of sliders compared to the number for 120 seconds for those with less time after passing the threshold, and on average larger numbers of sliders compared to the number for 120 seconds for those with more time after not passing the threshold. Note also that both workers know their own and the other worker's performance and payoffs in each round and in total.

i) If all aspects of the task and the payment of workers are clear, please answer the following questions about the task:

Assume that worker A placed 13 sliders correctly and worker B placed 27 sliders correctly in round 1.

What are the round 1 earnings of worker A (in \$)? _____ {0.26}

What are the round 1 earnings of worker B (in \$)? _____ {3.04}

How many seconds does worker A have available for the task in round 2? _____ {126}

How many seconds does worker B have available for the task in round 2? _____ {114}

ii) Let us now consider the overall picture of the payment mechanism described above. Different payment mechanisms to reward people for their work may be considered more or less fair. We are interested in how people think about the above described mechanism. To what extent do you think this payment mechanism is fair? Please indicate your fairness judgment on a scale from 0 (completely unfair) to 10 (completely fair).

Completely unfair ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Completely fair

Payment scheme: Tournament

Low inequality, no time advantage, no time handicap (Low)

The payment mechanism is as follows:

The two workers form a competitive group. That is, the goal of the workers in each round is to put a larger number of sliders to the target position within the given time limit than the other worker in the group. The time limit per round to work on the sliders is 120 seconds. The worker who placed more sliders in a group is the winner of this round. It does not matter how large the difference actually is, it only matters who of the two workers positioned more sliders correctly. In each of the four rounds, the winner of the round receives \$2. The loser receives \$1.40. In case of a tie where both players have the same number of correctly positioned sliders, the winner for this round will be randomly determined.

Our data containing more than 600 workers show that workers can correctly place about 20 sliders per round on average within a time limit of 120 seconds. How much a worker can earn per round depends on his own performance, the performance of his opponent, and in the event of a tie, a flip of a coin. If both workers in a group win two of the four rounds, then each would receive \$6.80 ($=\$2*2+\$1.40*2$).

The most skilled workers (top 10%) can place about 29 sliders per round on average, while the least skilled workers (bottom 10%) can place about 11 sliders per round on average. If a worker wins four rounds in a row, then he can earn \$8 ($=\$2*4$), while his opponent earns \$5.60 ($=\$1.40*4$), a difference of \$2.40 after four rounds. Note also that both workers know their own and the other worker's performance and payoffs in each round and in total.

i) If all aspects of the task and the payment of workers are clear, please answer the

following questions about the task:

Assume that worker A placed 13 sliders correctly and worker B placed 27 sliders correctly in round 1.

What are the round 1 earnings of worker A (in \$)? _____ {1.40}

What are the round 1 earnings of worker B (in \$)? _____ {2.00}

How many seconds does worker A have available for the task in round 2? _____ {120}

How many seconds does worker B have available for the task in round 2? _____ {120}

ii) Let us now consider the overall picture of the payment mechanism described above. Different payment mechanisms to reward people for their work may be considered more or less fair. We are interested in how people think about the above described mechanism. To what extent do you think this payment mechanism is fair? Please indicate your fairness judgment on a scale from 0 (completely unfair) to 10 (completely fair).

Completely unfair ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Completely fair

Payment scheme: Tournament

High inequality, no time advantage, no time handicap (High)

The payment mechanism is as follows:

The two workers form a competitive group. That is, the goal of the workers in each round is to put a larger number of sliders to the target position within the given time limit than the other worker in the group. The time limit per round to work on the sliders is 120 seconds. The worker who placed more sliders in a group is the winner of this round. It does not matter how large the difference actually is, it only matters who of the two workers positioned more sliders correctly. In each of the four rounds, the winner of the round receives \$3. The loser receives \$0.30. In case of a tie where both players have the

same number of correctly positioned sliders, the winner for this round will be randomly determined.

Our data containing more than 600 workers show that workers can correctly place about 20 sliders per round on average within a time limit of 120 seconds. How much a worker can earn per round depends on his own performance, the performance of his opponent, and in the event of a tie, a flip of a coin. If both workers in a group win two of the four rounds, then each would receive \$6.60 ($=\$3*2+\$0.30*2$).

The most skilled workers (top 10%) can place about 29 sliders per round on average, while the least skilled workers (bottom 10%) can place about 11 sliders per round on average. If a worker wins four rounds in a row, then he can earn \$12 ($=\$3*4$), while his opponent earns \$1.20 ($=\$0.30*4$), a difference of \$10.80 after four rounds. Note also that both workers know their own and the other worker's performance and payoffs in each round and in total.

i) If all aspects of the task and the payment of workers are clear, please answer the following questions about the task:

Assume that worker A placed 13 sliders correctly and worker B placed 27 sliders correctly in round 1.

What are the round 1 earnings of worker A (in \$)? _____ {0.30}

What are the round 1 earnings of worker B (in \$)? _____ {3.00}

How many seconds does worker A have available for the task in round 2? _____ {120}

How many seconds does worker B have available for the task in round 2? _____ {120}

ii) Let us now consider the overall picture of the payment mechanism described above. Different payment mechanisms to reward people for their work may be considered more or less fair. We are interested in how people think about the above described mechanism.

To what extent do you think this payment mechanism is fair? Please indicate your fairness judgment on a scale from 0 (completely unfair) to 10 (completely fair).

Completely unfair ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Completely fair

Payment scheme: Tournament

Low inequality, with time advantage (Low with Time Bonus)

The payment mechanism is as follows:

The two workers form a competitive group. That is, the goal of the workers in each round is to put a larger number of sliders to the target position within the given time limit than the other worker in the group. The worker who placed more sliders in a group is the winner of this round. It does not matter how large the difference actually is, it only matters who of the two workers positioned more sliders correctly. In each of the four rounds, the winner of the round receives \$2. The loser receives \$1.40. In case of a tie where both players have the same number of correctly positioned sliders, the winner for this round will be randomly determined.

The initial time limit in the first round is 120 seconds. Depending on performance, the time budget is reduced or increased in the subsequent round. In particular, for the winner, 6 seconds are added to his current time budget. For the loser, 6 seconds are subtracted from his current time budget. See the example for illustration.

Example: Worker A places more sliders correctly in round 1 than worker B. His time budget in round 2 will be 126 seconds, worker B's time budget will be 114 seconds. In round 2 worker A again places more sliders correctly than worker B. His time budget in round 3 will be 132 seconds, worker B's time budget will be 108 seconds. In round 3, worker A now places less sliders correctly than worker B. His time budget in round 4 will be 126 seconds, worker B's time budget will be 114.

In general, a shorter time budget to work on the task in a round makes it more difficult, and a longer time budget makes it less difficult, to correctly place more sliders than the opponent.

Our data containing more than 600 workers show that workers can correctly place about 20 sliders per round on average within a time limit of 120 seconds. How much a worker can earn per round depends on his own performance, the performance of his opponent, and in the event of a tie, a flip of a coin. If both workers in a group win two of the four rounds, then each would receive \$6.80 ($=\$2*2+\$1.40*2$).

The most skilled workers (top 10%) can place about 29 sliders per round on average, while the least skilled workers (bottom 10%) can place about 11 sliders per round on average. If a worker wins four rounds in a row, then he can earn \$8 ($=\$2*4$), while his opponent earns \$5.60 ($=\$1.40*4$), a difference of \$2.40 after four rounds. Note also that both workers know their own and the other worker's performance and payoffs in each round and in total.

Note that these calculations do not account for increased or reduced time budgets due to winning or losing earlier rounds. This will lead, on average, to larger numbers of sliders compared to the number for 120 seconds for those with more time, and on average lower numbers of sliders compared to the number for 120 seconds for those with less time. Note also that both workers know their own and the other worker's performance and payoffs in each round and in total.

i) If all aspects of the task and the payment of workers are clear, please answer the following questions about the task:

Assume that worker A placed 13 sliders correctly and worker B placed 27 sliders correctly in round 1.

What are the round 1 earnings of worker A (in \$)? _____ {1.40}

What are the round 1 earnings of worker B (in \$)? _____ {2.00}

How many seconds does worker A have available for the task in round 2? _____ {114}

How many seconds does worker B have available for the task in round 2? _____ {126}

ii) Let us now consider the overall picture of the payment mechanism described above. Different payment mechanisms to reward people for their work may be considered more or less fair. We are interested in how people think about the above described mechanism. To what extent do you think this payment mechanism is fair? Please indicate your fairness judgment on a scale from 0 (completely unfair) to 10 (completely fair).

Completely unfair ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Completely fair

Payment scheme: Tournament

High inequality, with time advantage (High with Time Bonus)

The payment mechanism is as follows:

The two workers form a competitive group. That is, the goal of the workers in each round is to put a larger number of sliders to the target position within the given time limit than the other worker in the group. The worker who placed more sliders in a group is the winner of this round. It does not matter how large the difference actually is, it only matters who of the two workers positioned more sliders correctly. In each of the four rounds, the winner of the round receives \$3. The loser receives \$0.30. In case of a tie where both players have the same number of correctly positioned sliders, the winner for this round will be randomly determined.

The initial time limit in the first round is 120 seconds. Depending on performance, the time budget is reduced or increased in the subsequent round. In particular, for the winner, 6 seconds are added to his current time budget. For the loser, 6 seconds are subtracted from his current time budget. See the example for illustration.

Example: Worker A places more sliders correctly in round 1 than worker B. His time budget in round 2 will be 126 seconds, worker B's time budget will be 114 seconds. In round 2 worker A again places more sliders correctly than worker B. His time budget in round 3 will be 132 seconds, worker B's time budget will be 108 seconds. In round 3, worker A now places less sliders correctly than worker B. His time budget in round 4 will be 126 seconds, worker B's time budget will be 114.

In general, a shorter time budget to work on the task in a round makes it more difficult, and a longer time budget makes it less difficult, to correctly place more sliders than the opponent.

Our data containing more than 600 workers show that workers can correctly place about 20 sliders per round on average within a time limit of 120 seconds. How much a worker can earn per round depends on his own performance, the performance of his opponent, and in the event of a tie, a flip of a coin. If both workers in a group win two of the four rounds, then each would receive \$6.60 ($=\$3*2+\$0.30*2$).

The most skilled workers (top 10%) can place about 29 sliders per round on average, while the least skilled workers (bottom 10%) can place about 11 sliders per round on average. If a worker wins four rounds in a row, then he can earn \$12 ($=\$3*4$), while his opponent earns \$1.20 ($=\$0.30*4$), a difference of \$10.80 after four rounds. Note also that both workers know their own and the other worker's performance and payoffs in each round and in total.

Note that these calculations do not account for increased or reduced time budgets due to winning or losing earlier rounds. This will lead, on average, to larger numbers of sliders compared to the number for 120 seconds for those with more time, and on average lower numbers of sliders compared to the number for 120 seconds for those with less time. Note also that both workers know their own and the other worker's performance and payoffs in each round and in total.

i) If all aspects of the task and the payment of workers are clear, please answer the following questions about the task:

Assume that worker A placed 13 sliders correctly and worker B placed 27 sliders correctly in round 1.

What are the round 1 earnings of worker A (in \$)? _____ {0.30}

What are the round 1 earnings of worker B (in \$)? _____ {3.00}

How many seconds does worker A have available for the task in round 2? _____ {114}

How many seconds does worker B have available for the task in round 2? _____ {126}

ii) Let us now consider the overall picture of the payment mechanism described above. Different payment mechanisms to reward people for their work may be considered more or less fair. We are interested in how people think about the above described mechanism. To what extent do you think this payment mechanism is fair? Please indicate your fairness judgment on a scale from 0 (completely unfair) to 10 (completely fair).

Completely unfair ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Completely fair

Payment scheme: Tournament

Low inequality, with time handicap (Low with Time Handicap)

The payment mechanism is as follows:

The two workers form a competitive group. That is, the goal of the workers in each round is to put a larger number of sliders to the target position within the given time limit than the other worker in the group. The worker who placed more sliders in a group is the winner of this round. It does not matter how large the difference actually is, it only matters who of the two workers positioned more sliders correctly. In each of the four rounds, the winner of the round receives \$2. The loser receives \$1.40. In case of a tie where both players have the same number of correctly positioned sliders, the winner for this round will be randomly determined.

The initial time limit in the first round is 120 seconds. Depending on performance, the time budget is reduced or increased in the subsequent round. In particular, for the winner, 6 seconds are subtracted from his current time budget, a time handicap. For the loser, 6 seconds are added to his current time budget, a time benefit. See the example for illustration.

Example: Worker A places more sliders correctly in round 1 than worker B. His time budget in round 2 will be 114 seconds, worker B's time budget will be 126 seconds. In round 2 worker A again places more sliders correctly than worker B. His time budget in round 3 will be 108 seconds, worker B's time budget will be 132 seconds. In round 3, worker A now places less sliders correctly than worker B. His time budget in round 4 will be 114 seconds, worker B's time budget will be 126.

In general, a shorter time budget to work on the task in a round makes it more difficult, and a longer time budget makes it less difficult, to correctly place more sliders than the opponent.

Our data containing more than 600 workers show that workers can correctly place about 20 sliders per round on average within a time limit of 120 seconds. How much a worker can earn per round depends on his own performance, the performance of his opponent, and in the event of a tie, a flip of a coin. If both workers in a group win two of the four rounds, then each would receive \$6.80 ($=\$2*2+\$1.40*2$).

The most skilled workers (top 10%) can place about 29 sliders per round on average, while the least skilled workers (bottom 10%) can place about 11 sliders per round on average. If a worker wins four rounds in a row, then he can earn \$8 ($=\$2*4$), while his opponent earns \$5.60 ($=\$1.40*4$), a difference of \$2.40 after four rounds. Note also that both workers know their own and the other worker's performance and payoffs in each round and in total.

Note that these calculations do not account for increased or reduced time budgets due to winning or losing earlier rounds. This will lead, on average, to larger numbers of sliders compared to the number for 120 seconds for those with more time, and on average lower numbers of sliders compared to the number for 120 seconds for those with less time. Note also that both workers know their own and the other worker's performance and payoffs in each round and in total.

i) If all aspects of the task and the payment of workers are clear, please answer the following questions about the task:

Assume that worker A placed 13 sliders correctly and worker B placed 27 sliders correctly in round 1.

What are the round 1 earnings of worker A (in \$)? _____ {1.40}

What are the round 1 earnings of worker B (in \$)? _____ {2.00}

How many seconds does worker A have available for the task in round 2? _____ {126}

How many seconds does worker B have available for the task in round 2? _____ {114}

ii) Let us now consider the overall picture of the payment mechanism described above. Different payment mechanisms to reward people for their work may be considered more or less fair. We are interested in how people think about the above described mechanism. To what extent do you think this payment mechanism is fair? Please indicate your fairness judgment on a scale from 0 (completely unfair) to 10 (completely fair).

Completely unfair ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Completely fair

Payment scheme: Tournament

High inequality, with time handicap (High with Time Handicap)

The payment mechanism is as follows:

The two workers form a competitive group. That is, the goal of the workers in each round is to put a larger number of sliders to the target position within the given time limit than the other worker in the group. The worker who placed more sliders in a group is the winner of this round. It does not matter how large the difference actually is, it only matters who of the two workers positioned more sliders correctly. In each of the four rounds, the winner of the round receives \$3. The loser receives \$0.30. In case of a tie where both players have the same number of correctly positioned sliders, the winner for this round will be randomly determined.

The initial time limit in the first round is 120 seconds. Depending on performance, the time budget is reduced or increased in the subsequent round. In particular, for the winner, 6 seconds are subtracted from his current time budget, a time handicap. For the loser, 6 seconds are added to his current time budget, a time benefit. See the example for illustration.

Example: Worker A places more sliders correctly in round 1 than worker B. His time budget in round 2 will be 114 seconds, worker B's time budget will be 126 seconds. In round 2 worker A again places more sliders correctly than worker B. His time budget in round 3 will be 108 seconds, worker B's time budget will be 132 seconds. In round 3, worker A now places less sliders correctly than worker B. His time budget in round 4 will be 114 seconds, worker B's time budget will be 126.

In general, a shorter time budget to work on the task in a round makes it more difficult, and a longer time budget makes it less difficult, to correctly place more sliders than the opponent.

Our data containing more than 600 workers show that workers can correctly place about 20 sliders per round on average within a time limit of 120 seconds. How much a worker can earn per round depends on his own performance, the performance of his opponent, and in the event of a tie, a flip of a coin. If both workers in a group win two of the four rounds, then each would receive \$6.60 ($=\$3*2+\$0.30*2$).

The most skilled workers (top 10%) can place about 29 sliders per round on average, while the least skilled workers (bottom 10%) can place about 11 sliders per round on average. If a worker wins four rounds in a row, then he can earn \$12 ($=\$3*4$), while his opponent earns \$1.20 ($=\$0.30*4$), a difference of \$10.80 after four rounds. Note also that both workers know their own and the other worker's performance and payoffs in each round and in total.

Note that these calculations do not account for increased or reduced time budgets due to winning or losing earlier rounds. This will lead, on average, to larger numbers of sliders compared to the number for 120 seconds for those with more time, and on average lower numbers of sliders compared to the number for 120 seconds for those with less time. Note also that both workers know their own and the other worker's performance and payoffs in each round and in total.

i) If all aspects of the task and the payment of workers are clear, please answer the following questions about the task:

Assume that worker A placed 13 sliders correctly and worker B placed 27 sliders correctly in round 1.

What are the round 1 earnings of worker A (in \$)? _____ {0.30}

What are the round 1 earnings of worker B (in \$)? _____ {3.00}

How many seconds does worker A have available for the task in round 2? _____ {126}

How many seconds does worker B have available for the task in round 2? _____ {114}

ii) Let us now consider the overall picture of the payment mechanism described above. Different payment mechanisms to reward people for their work may be considered more or less fair. We are interested in how people think about the above described mechanism. To what extent do you think this payment mechanism is fair? Please indicate your fairness judgment on a scale from 0 (completely unfair) to 10 (completely fair).

Completely unfair ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Completely fair

Sociodemographic Questionnaire

1. Please indicate your gender:

- A. Female
- B. Male
- C. Prefer not to answer or others

2. What is your year of birth (YYYY): _____.

3. What is your highest degree of completed education?

- A. Not completed high school.
- B. High school.
- C. 2-year college degree
- D. 4-year college degree
- E. Master's degree
- F. Doctoral degree
- G. Professional degree (JD, MD)

4. What is your ethnicity?

- A. White/European-American,
- B. Black/African-American.
- C. Asian/Asian-American/Pacific Islander.
- D. Hispanic/Latino.
- E. Other:_____

5. On a continuum from liberal to conservative, how would you describe your political beliefs?

- A. Strongly liberal.
- B. Moderately liberal.
- C. Slightly liberal.
- D. Slightly conservative.
- E. Moderately conservative.
- F. Strongly conservative.

6. Generally speaking, do you usually think of yourself as a Republican, a Democrat, an Independent, or something else?

- A. Republican.
- B. Democrat.
- C. Independent.
- D. Other: ____

7. Do you live in the United States? Yes/ No


8. In which state do you live? _____

9. What describes best the area where you currently live in?

- A. Urban area.
- B. Rural area.

10. This ladder shows where people in the U.S. stand in society. At the top of the ladder (marked "10") are those people who are doing best, those who have the most money, have the best education and are most respected. At the bottom of the ladder (marked "1") are those who are worst off, have the least money, have the worst education, and are the least

respected. Where would you place yourself on this ladder relative to the people at the top and bottom? Please mark the box next to the rung you think you are standing on!



A vertical ladder with 10 rungs. The rungs are numbered 5 to 10 from bottom to top. Each rung has a small box next to it for marking.

Rung	Box
10	<input type="checkbox"/>
9	<input type="checkbox"/>
8	<input type="checkbox"/>
7	<input type="checkbox"/>
6	<input type="checkbox"/>
5	<input type="checkbox"/>

4
3
2
1

11. What is your current employment status?

- A. Currently employed.
- B. Currently unemployed (but employed before)
- C. Never employed

12. What is your approximate monthly disposable income in USD?

- A. Up to \$1,500
- B. \$1,500 to \$4,500
- C. \$4,500 or more
- D. Don't know or prefer not to answer.

13. In the area where you live, your family financial situation is...

- A. far below the average.
- B. below the average.
- C. at the average.

- D. higher than the average.
- E. far higher than the average
- F. Don't know or prefer not to answer.

14. Consider all income sources in your family, including all sources from your family members, how easy it is to break-even?

- A. Very Difficult.
- B. Somewhat difficult.
- C. Neither Difficult nor easy.
- D. Somewhat Easy.
- E. Very Easy.
- F. Don't know or prefer not to answer.

OS.2 STUDY 1: ATTRITION AND BALANCE STATISTICS

To rule out selective attrition after reading the scenarios as an explanation for our results, we analyze the pattern of attrition across the 16 conditions, and test whether the sample is still balanced across conditions with respect to the characteristics of the respondents. Table OS1 gives an overview of dropouts by treatment. As shown in the table, there is some variation in dropout rates. Table OS2 shows pair-wise comparisons of dropout rates between conditions, using the Fisher's exact test. Overall, there are few systematic differences, except for conditions T3 and T8.

We further examine whether the observed differences affected the sample composition in the conditions. We regress each treatment dummy separately on each of the 7 demographic variables we collected: $Treatment_i = \beta_0 + \beta_{i,j}Covariate_{i,j} + \epsilon_{i,j} \forall i = 1, \dots, 16; j = Demographic\ variable$. The p-values of the estimated coefficients (OLS, robust standard errors) are reported in Table OS3. There are in total merely 5 instances where the p-value is smaller than 0.05 (not corrected for multiple testing), out of 112 regressions. If we correct for multiple testing using Bonferroni's correction, then none of the estimated coefficients is significantly different from zero. In the last row, we also report F-tests when including all covariates into the regression.

There is no instance in which the p-value is less than the critical value of 0.05. We thus conclude that there was no selective attrition affecting the sample composition. The sample is balanced with respect to the observed characteristics across all treatments.

Table OS1. Dropout overview

Treatment	Payment Scheme	Stayed	Dropped out	Total	Dropout rate
Piece-rate:					
1	Low with Base Pay	245	97	342	28%
2	Medium No Base Pay	227	85	312	27%
3	High with Entrance Fee	171	125	296	42%
4	Exponential	236	99	335	30%
Discrete bonus:					
5	Low	147	68	215	32%
6	High	118	73	191	38%
7	Low with Time Advantage	132	80	212	38%
8	High with Time Advantage	102	89	191	47%
9	Low with Time Handicap	138	91	229	40%
10	High with Time Handicap	119	75	195	38%
Tournament:					
11	Low	135	72	208	35%
12	High	140	65	205	32%
13	Low with Time Advantage	132	82	214	38%
14	High with Time Advantage	121	86	207	42%
15	Low with Time Handicap	122	83	205	40%
16	High with Time Handicap	146	73	219	33%
Total		2,431	1,343	3,776	

Table OS2. Study 1 – Pair-wise comparisons of the dropout rate (corrected for multiple testing)

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15
T1															
T2	1.0														
T3	0.04	0.01													
T4	1.0	1.0	0.14												
T5	1.0	1.0	1.0	1.0											
T6	1.0	1.0	1.0	1.0	1.0										
T7	1.0	1.0	1.0	1.0	1.0	1.0									
T8	<0.01	<0.01	1.0	0.01	0.27	1.0	1.0								
T9	0.59	0.35	1.0	1.0	1.0	1.0	1.0	1.0							
T10	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0						
T11	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0					
T12	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.33	1.0	1.0	1.0				
T13	1.0	0.93	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0			
T14	0.24	0.10	1.0	0.61	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0		
T15	0.57	0.25	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	
T16	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.77	1.0	1.0	1.0	1.0	1.0	1.0	1.0

Notes: the entries report the Bonferroni corrected p-value of the Fisher's exact test comparing the dropout rate in treatment T_i to that in treatment T_j where $i \neq j$. When the p-value is greater than 1.0 after correction, we record it as 1.0. P-values below 0.05 are shown in bold

Table OS3. Study 1 – Sample Balance Check

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15	T16
Male	0.42	0.22	0.39	0.04	0.69	0.57	0.35	0.77	0.94	0.42	0.19	0.03	0.12	0.52	0.53	0.99
Education	0.96	0.16	0.65	0.92	0.84	0.32	0.28	0.55	0.39	0.55	0.09	0.34	0.57	0.90	0.03	0.68
Income	0.35	0.66	0.93	0.42	0.65	0.71	0.10	0.08	0.75	0.28	0.35	0.37	0.34	0.02	0.06	0.44
Social status	0.79	0.38	0.41	0.26	0.89	0.49	0.58	0.94	0.67	0.55	0.22	0.57	0.90	0.09	0.37	0.66
Ethnicity	0.87	0.28	0.96	0.62	0.75	0.56	0.66	0.39	0.70	0.00	0.11	0.29	0.88	0.31	0.11	0.16
Rural area	0.76	0.70	0.75	0.33	0.90	0.59	0.06	0.41	0.74	0.42	0.12	0.40	0.77	0.73	0.28	0.06
Employment status	0.35	0.49	0.45	0.85	0.83	0.76	0.79	0.07	0.92	0.43	0.85	0.98	0.47	0.47	0.84	0.86
All	0.94	0.48	0.95	0.25	0.99	0.63	0.10	0.21	0.80	0.06	0.11	0.30	0.79	0.41	0.21	0.47

Notes: the table reports p-values of the estimated coefficient from OLS regressions with robust standard errors. The last row reports p-values of the F-test when including all demographic variables. All p-values reported here are not corrected for multiple-testing. P-values below 0.05 are shown in bold

OS.3 STUDY 1: HISTOGRAMS OF FAIRNESS JUDGMENTS

Figure OS1: Histogram of fairness judgments of piece rate schemes (T1-T4)

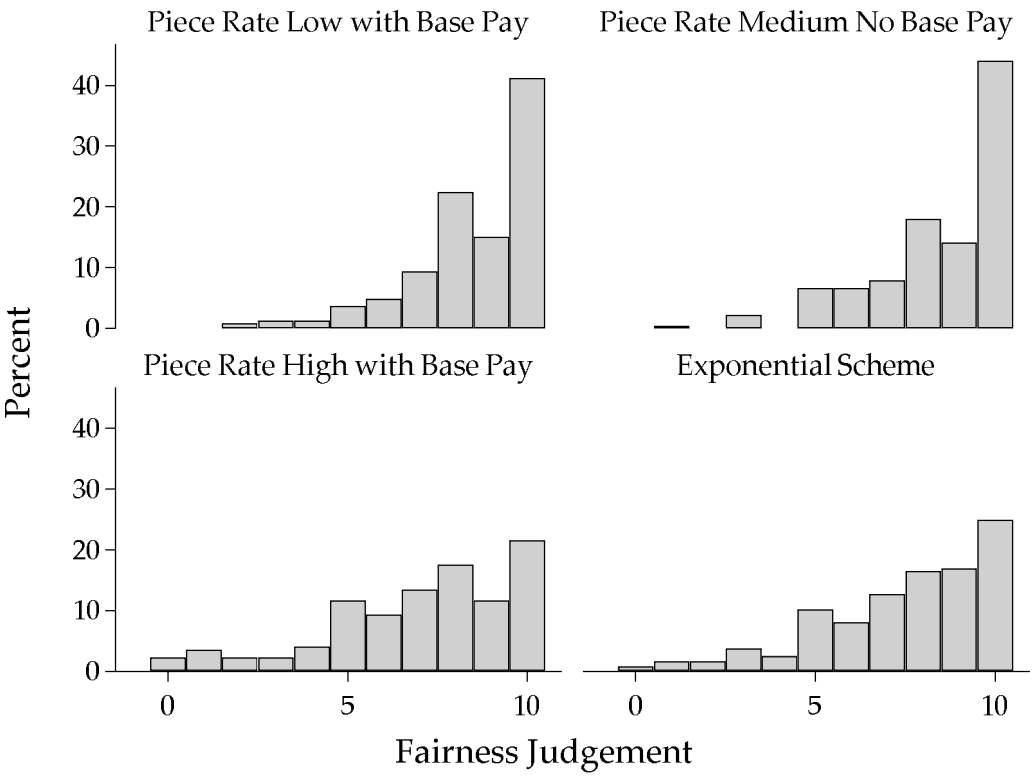


Figure OS2: Histogram of fairness judgments of discrete bonus schemes (T5-T10)

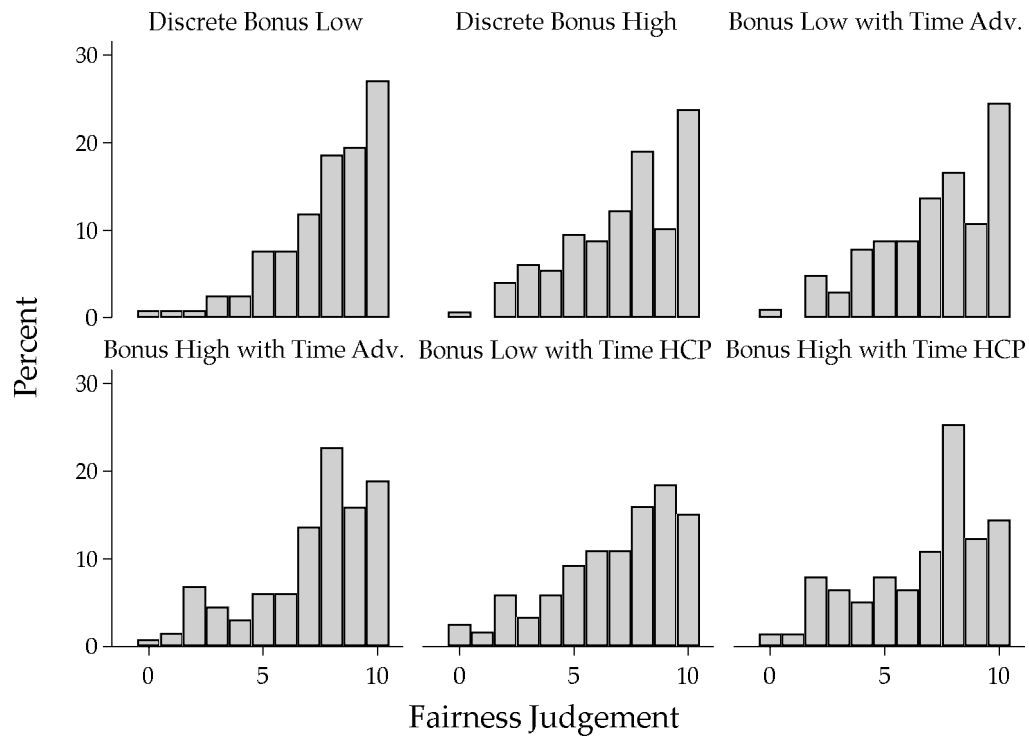
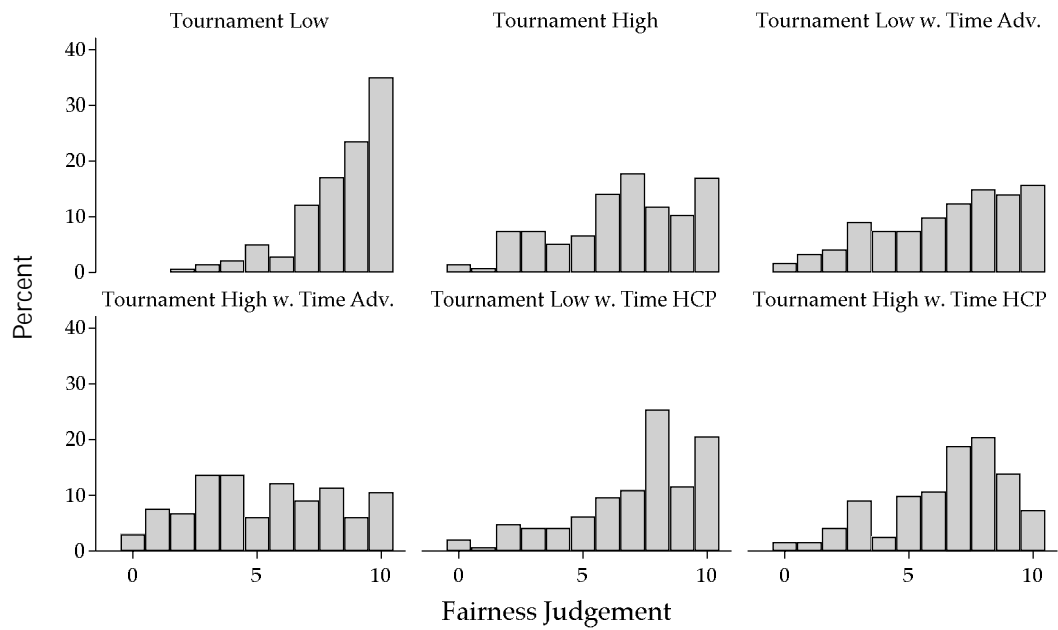


Figure OS3: Histogram of fairness judgments of tournament schemes (T11-T16)



OS.4 STUDY 2: ATTRITION AND BALANCE STATISTICS

Table OS4. Study 2. Sample Balance Check

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15	T16
Gender	0.44	0.41	0.58	0.62	0.83	0.65	0.62	0.73	0.48	0.43	0.77	0.16	0.11	0.95	0.01	0.86
Education	0.08	0.65	0.35	0.99	0.98	0.72	0.55	0.73	0.66	0.94	0.53	0.42	0.95	0.60	0.61	0.37
Income	0.07	0.99	0.28	0.05	0.02	0.18	0.18	0.79	0.57	0.61	0.67	0.30	0.14	0.88	0.74	0.65
Social status	0.44	0.41	0.46	0.56	0.66	0.53	0.84	0.74	0.64	0.52	0.09	0.67	0.65	0.03	0.02	0.12
Ethnicity	0.72	0.72	0.21	0.27	0.88	0.35	0.42	0.91	0.80	0.74	0.89	0.18	0.44	0.96	0.42	0.30
Rural area	0.57	0.04	0.68	0.03	0.62	0.43	0.90	0.15	0.17	0.33	0.28	0.53	0.10	0.18	0.55	0.86
Employment status	0.01	0.76	0.96	0.96	0.67	0.19	0.75	0.13	0.91	0.00	0.24	0.28	0.94	0.86	0.17	0.13
All	0.15	0.51	0.67	0.21	0.45	0.46	0.73	0.64	0.86	0.01	0.74	0.53	0.25	0.31	0.01	0.54

Notes: the table reports p-values of the estimated coefficient from OLS regressions with robust standard errors. The last row reports p-values of the F-test when including all demographic variables. All p-values reported here are not corrected for multiple-testing. P-values below 0.05 are shown in bold

OS.5 STUDY 2: HISTOGRAMS OF OVERALL FAIRNESS JUDGMENTS AND HX
FAIRNESS JUDGMENTS

Figure OS4: Histogram of fairness judgments of piece rate schemes (T1-T4)

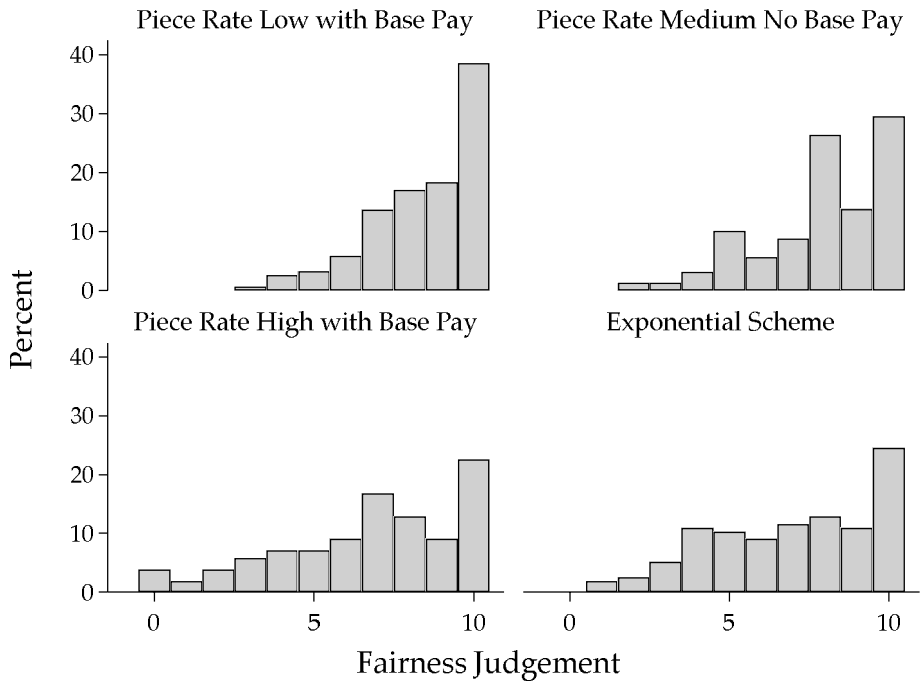


Figure OS5: Histogram of fairness judgments of discrete bonus schemes (T5-T10)

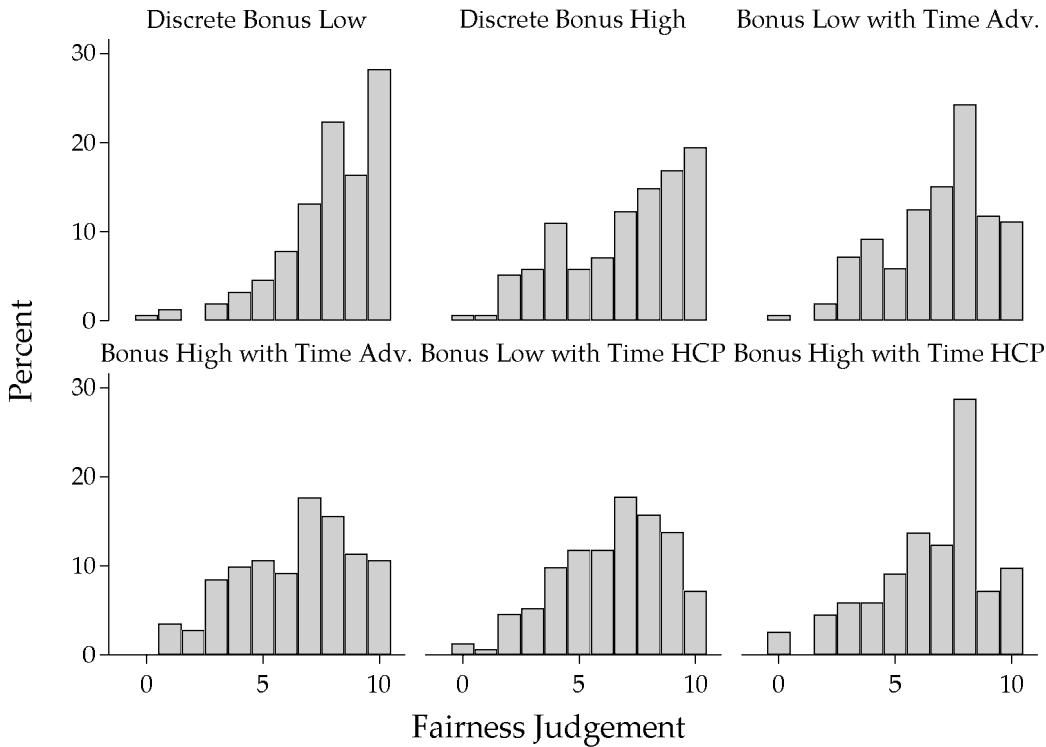


Figure OS6: Histogram of fairness judgments of tournament schemes (T11-T16)

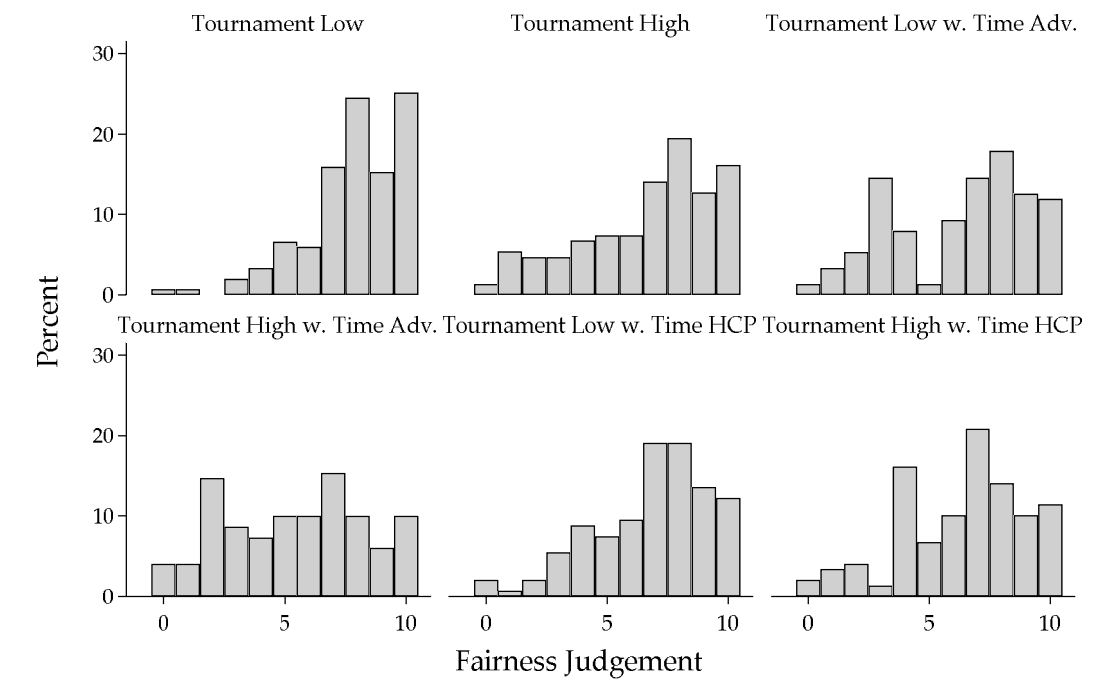


Figure OS7: Histogram of HX fairness judgments of piece rate schemes (T1-T4)

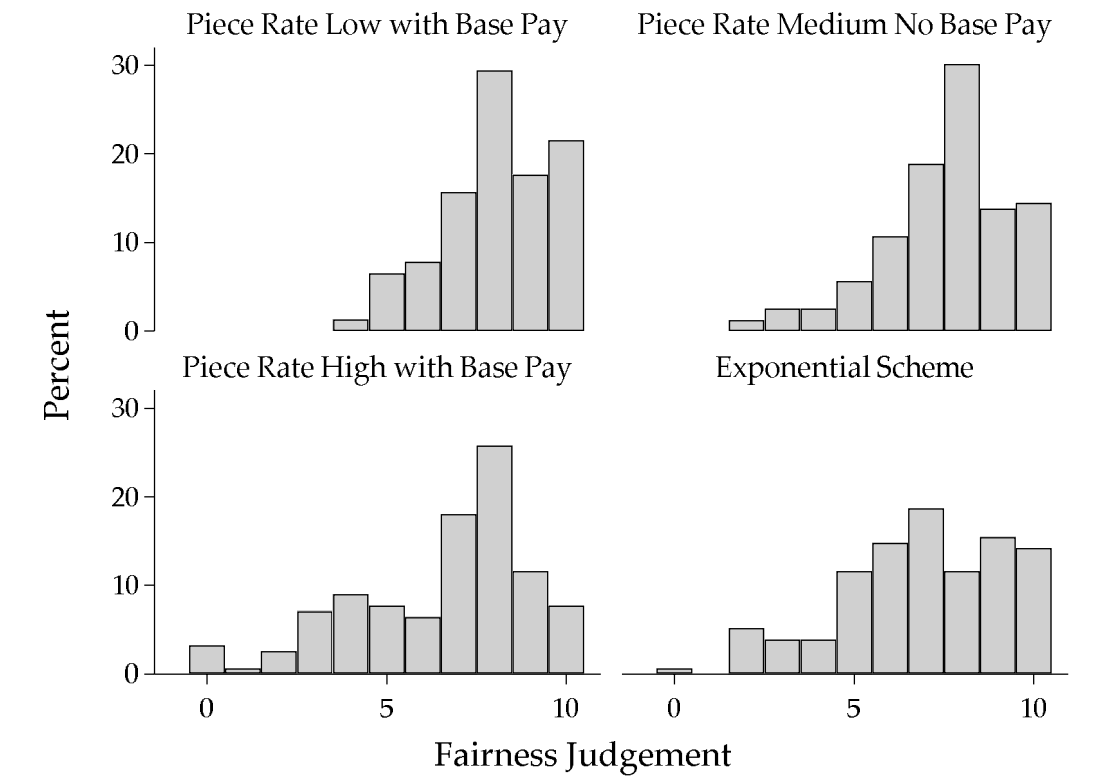


Figure OS8: Histogram of HX fairness judgments of discrete bonus schemes (T5-T10)

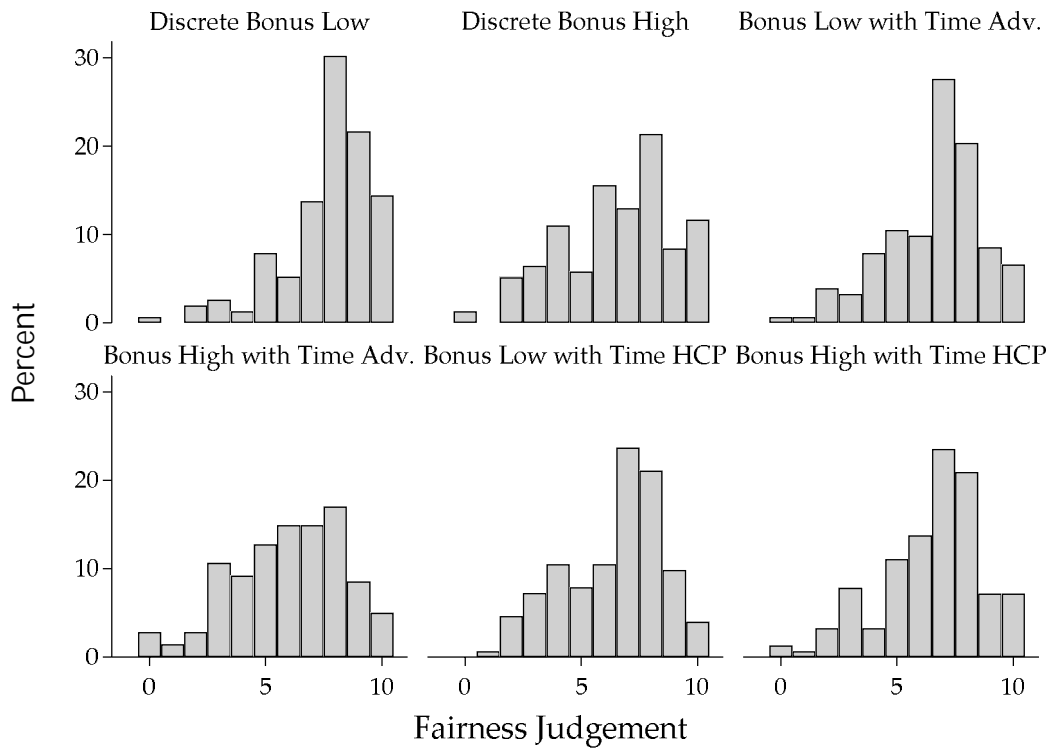
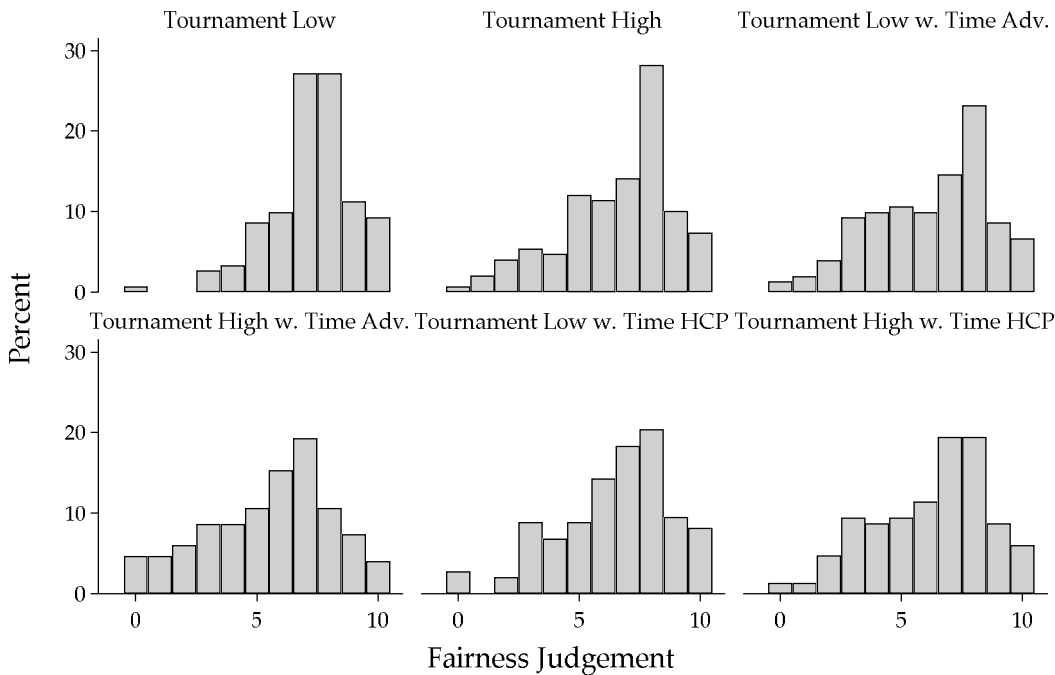


Figure OS9: Histogram of HX fairness judgments of tournament schemes (T11-T16)



OS6. STUDY 2: SUBSAMPLE ANALYSES BY TASK COMPREHENSION

Table OS5. Study 2 – Subsample analyses by task comprehension

	HX Fairness judgment			
	(1)	(2)	(3)	(4)
Gini	-1.14*** (0.16)	-0.44 (0.36)	-1.14*** (0.16)	-0.45 (0.36)
Discrete Bonus=1	-0.22 (0.15)	-0.14 (0.41)	-0.28* (0.17)	0.03 (0.45)
Tournament=1	-0.48*** (0.14)	-0.43 (0.43)	-0.42*** (0.16)	-0.70 (0.51)
Time Advantage=1	-1.00*** (0.14)	-0.66* (0.34)	-0.79*** (0.20)	-0.90** (0.44)
Time Handicap=1	-0.66*** (0.14)	-0.52 (0.34)	-0.66*** (0.19)	-0.77* (0.45)
Exponential Scheme=1	-0.07 (0.23)	0.12 (0.80)	-0.06 (0.23)	0.12 (0.80)
Tournament x Time Advantage			-0.40 (0.28)	0.59 (0.70)
Tournament x Time Handicap			-0.00 (0.27)	0.62 (0.71)
Constant	8.40*** (0.20)	7.04*** (0.60)	8.39*** (0.20)	7.04*** (0.61)
Task Comprehension	Yes	No	Yes	No
Controls	Yes	Yes	Yes	Yes
Observations	2,025	398	2,025	398
R-squared	0.09	0.06	0.09	0.06
F-statistics (Discrete Bonus = Tournament)	5.36**	1.09	0.59	1.84
F-statistics (Time Bonus = Time Handicap)	5.17**	0.20	0.44	0.11

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. OLS regression on HX fairness judgment with robust standard errors in parentheses. Sample split in subjects who passed the comprehension check (Task Comprehension “Yes”) and subjects who answered some or all questions wrong in all attempts (Task Comprehension “No”). Piece rate is the reference category. Controls include gender, level of education, personal income (categorical), indicators for ethnic groups, dummy for living in a rural area, political orientation (0: Liberal; 5: Conservative), self-perceived social status, and employment status.

OS7. STUDY 2: SUMMARY OF FAIRNESS MEASURES

Table OS6. Study 2 – Summary of different fairness measures

	Payment Scheme	Fairness overall	Procedural fairness	Outcome fairness	HX fairness overall	HX procedural fairness	HX outcome fairness
Piece Rate:							
T1:	Low with Base Pay	8.46	8.56	7.01***	8.05	8.05	6.69***
T2:	Medium No Base Pay	7.96	8.23	7.18***	7.55	7.72	6.69***
T3:	High with Entrance Fee	6.73	7.35	6.39***	6.55	7.28	6.03***
T4:	Exponential	7.01	7.47	6.06***	6.90	7.34	6.10***
Discrete Bonus:							
T5:	Low	7.93	8.11	7.55***	7.67	7.74	7.08***
T6:	High	6.97	7.48	5.82***	6.52	6.88	5.69***
T7:	Low with Time Advantage	6.86	6.99	6.47**	6.61	6.78	6.30**
T8:	High with Time Advantage	6.40	6.94	5.79***	5.96	6.61	5.48***
T9:	Low with Time Handicap	6.43	6.58	6.47	6.39	6.57	6.33
T10:	High with Time Handicap	6.58	7.09	5.90***	6.50	6.80	5.61***
Tournament:							
T11:	Low	7.85	7.77	6.82***	7.25	7.38	6.58***
T12:	High	6.67	7.59	5.93***	6.62	7.26	5.87***
T13:	Low with Time Advantage	6.29	6.66	6.42	6.21	6.40	5.97*
T14:	High with Time Advantage	5.40	6.01	5.57	5.49	5.97	5.59
T15:	Low with Time Handicap	6.82	7.00	6.86	6.47	6.49	6.44
T16:	High with Time Handicap	6.37	6.68	5.66***	6.20	6.58	5.66***

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$ indicate significant differences between procedural and outcome fairness judgment (two-sided t-test).

OS8. STUDY 2: COMPOSITION OF OVERALL FAIRNESS MEASURES

Table OS.7 Study 2 – Composition of overall fairness judgments

	Fairness overall	HX fairness overall
Procedural fairness measure	0.51*** (0.02)	
Outcome fairness measure	0.41*** (0.02)	
HX procedural fairness measure		0.53*** (0.02)
HX outcome fairness measure		0.37*** (0.02)
Constant	0.59*** (0.13)	0.70*** (0.12)
Observations	2,423	2,423
R-squared	0.58	0.59
<i>F-statistics (Procedural fairness = Outcome fairness)</i>	9.92***	20.59***

Notes: *** p<0.01, ** p<0.05, * p<0.1. OLS regression on overall and HX fairness judgment with robust standard errors in parentheses.

OS9. ROBUSTNESS CHECK: BINARIZED MEASURES

Tables OS 8 to OS10 present Study-1 and Study-2 baseline regressions vis-à-vis two robustness checks based on either binarizing fairness judgments at the midpoint score of 5, or at the median score of 7. Specification (1) of each Table shows the results discussed in the main text for easier reference.

Table OS8. Robustness Check Binarized Measures Study 1

	Fairness judgments (specification 1: scale 1-10, 2: binary for values >5, 3: binary for values >7)		
	(1)	(2)	(3)
Gini	-1.33*** (0.17)	-0.58*** (0.09)	-0.61*** (0.09)
Discrete bonus incentives	-0.32** (0.15)	-0.19** (0.10)	-0.20** (0.08)
Tournament	-0.74*** (0.14)	-0.33*** (0.09)	-0.40*** (0.08)
Time advantage	-0.94*** (0.16)	-0.42*** (0.09)	-0.33*** (0.08)
Time handicap	-0.69*** (0.15)	-0.26*** (0.09)	-0.21*** (0.08)
Exponential	-0.02 (0.19)	-0.08 (0.12)	-0.06 (0.10)
Constant	8.72*** (0.10)	1.42*** (0.08)	0.83*** (0.07)
Controls	No	No	No
Observations	2,431	2,431	2,431
R-squared	0.08	0.05	0.04
<i>F/Chi2-statistics (Discrete Bonus=Tournament)</i>	10.16***	3.67*	9.79***
<i>F/ Chi2-statistics (Time bonus = Time handicap)</i>	2.27	3.58*	2.24

Notes: *** p<0.01, ** p<0.05, * p<0.1. Specification (1): OLS regression on fairness judgments with robust standard errors in parentheses. Specifications (2)+(3): Probit regressions on fairness judgements with robust standard errors in parentheses. Piece rate is the reference category.

Table OS9. Robustness Check Binarized HX Measures Study 2

	HX Fairness judgments (specification 1: scale 1-10, 2: binary for values >5, 3: binary for values >7)		
	(1)	(2)	(3)
Gini	-1.03*** (0.14)	-0.50*** (0.09)	-0.43*** (0.08)
Discrete bonus incentives	-0.24* (0.14)	-0.15 (0.10)	-0.14 (0.08)
Tournament	-0.49*** (0.13)	-0.25*** (0.09)	-0.22*** (0.08)
Time advantage	-0.95*** (0.13)	-0.44*** (0.08)	-0.49*** (0.07)
Time handicap	-0.62*** (0.13)	-0.27*** (0.08)	-0.38*** (0.07)
Exponential scheme	-0.08 (0.22)	-0.12 (0.13)	-0.24* (0.12)
Constant	7.91*** (0.11)	1.24*** (0.08)	0.41*** (0.07)
Controls	No	No	No
Observations	2,423	2,423	2,423
R-squared	0.07	0.04	0.04
<i>F/Chi2-statistics (Discrete Bonus=Tournament)</i>	5.56**	2.63	2.12
<i>F/Chi2-statistics (Time bonus = Time handicap)</i>	5.81**	4.96**	2.30

Notes: *** p<0.01, ** p<0.05, * p<0.1. Specification (1): OLS regression on HX fairness judgments with robust standard errors in parentheses. Specifications (2)+(3): Probit regressions on fairness judgements with robust standard errors in parentheses. Piece rate is the reference category.

Table OS10. Robustness Check Binarized Measures Study 2

Non-incentivized Fairness judgments (specification 1: scale 1-10, 2: binary for values >5, 3: binary for values >7)			
	(1)	(2)	(3)
Gini	-1.18*** (0.16)	-0.57*** (0.09)	-0.50*** (0.08)
Discrete bonus incentives	-0.21 (0.15)	-0.08 (0.10)	-0.07 (0.08)
Tournament	-0.52*** (0.15)	-0.20** (0.09)	-0.18** (0.08)
Time advantage	-1.12*** (0.14)	-0.45*** (0.08)	-0.51*** (0.07)
Time handicap	-0.81*** (0.14)	-0.31*** (0.08)	-0.44*** (0.07)
Exponential scheme	-0.24 (0.24)	-0.24* (0.13)	-0.17 (0.12)
Constant	8.31*** (0.12)	1.25*** (0.08)	0.58*** (0.07)
Controls	No	No	No
Observations	2,423	2,423	2,423
R-squared	0.07	0.04	0.05
<i>F/Chi2-statistics (Discrete bonus = Tournament)</i>	7.23***	3.58*	3.54*
<i>F/Chi2-statistics (Time bonus = Time handicap)</i>	4.64**	3.63*	0.83

Notes: *** p<0.01, ** p<0.05, * p<0.1. Specification (1): OLS regression on non-incentivized fairness judgments with robust standard errors in parentheses. Specifications (2)+(3): Probit regressions on fairness judgements with robust standard errors in parentheses. Piece rate is the reference category.

OS10. ROBUSTNESS CHECK: Excluding GINI

Table OS11. Robustness Check Multivariate analyses of fairness judgments w/o Gini

– Study 1

	Fairness judgment			
	(1)	(2)	(3)	(4)
Discrete Bonus=1	-0.44*** (0.15)	-0.36** (0.15)	-0.64*** (0.17)	-0.57*** (0.17)
Tournament=1	-0.77*** (0.15)	-0.74*** (0.15)	-0.58*** (0.17)	-0.53*** (0.17)
Time Advantage=1	-0.96*** (0.16)	-1.03*** (0.17)	-0.28 (0.22)	-0.29 (0.23)
Time Handicap=1	-0.67*** (0.15)	-0.65*** (0.16)	-0.71*** (0.22)	-0.70*** (0.23)
Tournament x Time Advantage			-1.31*** (0.32)	-1.44*** (0.33)
Tournament x Time Handicap			0.08 (0.30)	0.11 (0.32)
Exponential Scheme=1	-0.61*** (0.18)	-0.58*** (0.19)	-0.61*** (0.17)	-0.58*** (0.19)
Constant	8.10*** (0.08)	8.01*** (0.32)	8.72*** (0.10)	8.59*** (0.32)
Controls	No	Yes	No	Yes
Observations	2,431	2,182	2,431	2,182
R-squared	0.06	0.09	0.07	0.10
<i>F-statistics (Discrete Bonus = Tournament)</i>	6.46**	7.82***	0.07	0.03
<i>F-statistics (Time Bonus = Time Handicap)</i>	3.07*	4.64**	3.30*	2.90*

Notes: *** p<0.01, ** p<0.05, * p<0.1. OLS regression on fairness judgments with robust standard errors in parentheses. Piece rate is the reference category. Controls include gender, level of education, personal income (categorical), indicators for ethnic groups and for living in rural area, political orientation (1: Liberal; 6: Conservative), self-perceived social status, and employment status.

Table OS12. Robustness Check Multivariate analyses of HX fairness judgments w/o

Gini – Study 2

	HX Fairness judgment			
	(1)	(2)	(3)	(4)
Discrete Bonus=1	-0.25*	-0.25*	-0.29*	-0.29*
	(0.14)	(0.14)	(0.16)	(0.15)
Tournament=1	-0.49***	-0.49***	-0.44***	-0.45***
	(0.14)	(0.14)	(0.15)	(0.15)
Time Advantage=1	-0.94***	-0.93***	-0.79***	-0.79***
	(0.13)	(0.13)	(0.18)	(0.19)
Time Handicap=1	-0.63***	-0.63***	-0.65***	-0.65***
	(0.13)	(0.13)	(0.18)	(0.18)
Tournament x			-0.29	-0.28
Time Advantage			(0.26)	(0.26)
Tournament x			0.04	0.04
Time Handicap			(0.25)	(0.25)
Exponential Scheme=1	-0.49**	-0.50**	-0.49**	-0.50**
	(0.21)	(0.21)	(0.21)	(0.21)
Constant	7.38***	7.67***	7.38***	7.67***
	(0.09)	(0.19)	(0.09)	(0.19)
Controls	No	Yes	No	Yes
Observations	2,423	2,423	2,423	2,423
R-squared	0.05	0.05	0.05	0.05
<i>F-statistics (Discrete Bonus = Tournament)</i>	4.81**	4.80**	0.75	0.79
<i>F-statistics (Time Bonus = Time Handicap)</i>	5.49**	4.92**	0.68	0.63

Notes: *** p<0.01, ** p<0.05, * p<0.1. OLS regression on HX fairness judgment with robust standard errors in parentheses. Piece rate is the reference category. Controls include gender, level of education, personal income (categorical), indicators for ethnic groups and for living in a rural area, political orientation (1: Liberal; 6: Conservative), self-perceived social status, and employment status.