



**HAL**  
open science

## Essays on attitudes towards economic inequality

Max Lobeck

► **To cite this version:**

Max Lobeck. Essays on attitudes towards economic inequality. Economics and Finance. Université Panthéon-Sorbonne - Paris I, 2021. English. NNT : 2021PA01E043 . tel-03552460

**HAL Id: tel-03552460**

**<https://theses.hal.science/tel-03552460>**

Submitted on 2 Feb 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Université Paris 1 Panthéon-Sorbonne — Paris School of Economics

## Ph.D. Thesis

Submitted for the Degree of Doctor of Philosophy in Economics

*Prepared and defended on July 15, 2021 by*

MAX LOBECK

## Essays on Attitudes towards Economic Inequality

Thesis advisors: Nicolas Jacquemet and Claudia Senik

### Jury:

<i>President</i>	Marc Fleurbaey	Paris School of Economics, CNRS
<i>Reviewer</i>	Urs Fischbacher	University of Konstanz
<i>Reviewer</i>	Bertil Tungodden	Norwegian School of Economics
<i>Examiner</i>	Beatrice Boulu-Reshef	University of Orléans
<i>Advisor</i>	Nicolas Jacquemet	Paris School of Economics, University Paris 1
<i>Advisor</i>	Claudia Senik	Paris School of Economics, Sorbonne University



Université Paris 1 Panthéon-Sorbonne — Paris School of Economics

## Thèse

pour l'obtention du titre de Docteur en Sciences Économiques

*Présentée et soutenue publiquement le 15 juillet 2021 par :*

MAX LOBECK

## Essais sur les attitudes envers l'inégalité économique

Supervisé par Nicolas Jacquemet and Claudia Senik

### Jury

<i>President</i>	Marc Fleurbaey	Paris School of Economics, CNRS
<i>Reviewer</i>	Urs Fischbacher	Universität de Konstanz
<i>Reviewer</i>	Bertil Tungodden	Norwegian School of Economics
<i>Examiner</i>	Beatrice Boulu-Reshef	Université d'Orléans
<i>Advisor</i>	Nicolas Jacquemet	Paris School of Economics, Université Paris 1
<i>Advisor</i>	Claudia Senik	Paris School of Economics, Sorbonne University



# Dissertation Abstract

The three chapters of this dissertation study individual attitudes towards economic inequality. The common focus of the chapters is to study how these attitudes are shaped by the economic environment and context. It, thereby, improves our general understanding of the prevalence, shape, and consequences of attitudes towards economic inequality.

The first chapter, *Preferences over Income Distribution: Evidence from a Choice Experiment*, is joint work together with Sophie Cêtre, Claudia Senik, and Thierry Verdier. Using a choice experiment in the lab, we assess the relative importance of different motives behind attitudes to income inequality. We elicit subjects' preferences regarding pairs of payoff distributions within small groups, in a firm-like setting. We find that distributions that satisfy the Pareto-dominance criterion attract unanimous suffrage: all subjects prefer larger inequality provided it makes everyone weakly better off. This is true no matter whether payoffs are based on merit or luck. Unanimity only breaks once subjects' positions within the income distribution are fixed and known ex-ante. Even then, 75% of subjects prefer Pareto-dominant distributions, but 25% of subjects choose to decrease inequality by decreasing top incomes *without* redistributing that money and, thus, without making others better off. Furthermore, a majority of subjects embrace a more equal distribution if their own income or overall efficiency is not at stake. When their own income is at stake and the sum of payoffs remains unaffected, 20% of subjects are willing to pay for a lower degree of inequality.

The second chapter, *Principal's Distributive Preferences and the Incentivization of Agents*, is joint work with Sophie Cêtre. There, we ask whether principals' distributive preferences affect the allocation of incentives within firms. We document a robust relationship between French employers' fairness preferences and the incentive contracts they choose for their workers. To establish causality, we run a Principal-Agent lab experiment, framed as a firm setting. In the experiment, subjects are randomized in the principal or worker position. Principals must choose piece rate wage contracts for two workers that differ in terms of ability. Workers have to choose an effort level that is non-contractible. Principals are either paid in proportion to the output produced (Stakeholder treatment) or paid a fixed wage (Spectator treatment). We study how principals make trade-offs between incentive concerns (motivating workers to maximize output) and their own normative distributive preferences. We find that, despite the firm-frame and the moral hazard situation, principals do hold egalitarian concerns, but they are sensitive to both extensive and intensive margin incentives. We characterize the heterogeneity in distribu-

---

tive preferences by positing a utility function that incorporates the principal's other-regarding preferences and we estimate it using a finite mixture model and show that principals can be categorized as performance-maximizing, inequality targeting, and egalitarian.

The third chapter, *Motivating Beliefs in a Just World* studies whether individuals distort their beliefs about the relative importance of effort and luck to motivate themselves to exert effort. To that end, I develop a novel experimental design where past experience of success or failure serves as a noisy signal about the true importance of effort in achieving success. To test whether individuals distort their beliefs to motivate future effort, I vary the moment in time when subjects are informed about an effortful task to be performed later in the experiment. Subjects who receive the information before belief elicitation face an incentive to distort their beliefs to motivate effort in the later task. The results show that such individuals are more likely to believe that their effort is important for success. Motivating belief distortion is particularly pronounced for subjects who receive disincentivizing news about the true state of the world, i.e. that success depends on luck rather than on effort. I additionally test whether motivating belief distortion affects subjects' willingness to distribute money between two *other* individuals as a third-party spectator. I find no evidence that distributive behavior differs across the two treatment groups. These results suggest that individuals' luck-effort beliefs not only depend on past or current events that inform about the true state of the world but are also endogenous to the incentive structure individuals expect to face.

# Résumé de la thèse

Les trois chapitres de cette thèse étudient les attitudes individuelles à l'égard de l'inégalité économique. L'objectif commun de ces chapitres est d'étudier comment ces attitudes sont façonnées par l'environnement et le contexte économiques. Cette thèse permet donc d'améliorer ainsi notre connaissance générale de la prévalence, de la forme et des conséquences des attitudes envers l'inégalité.

Le premier chapitre, *Preferences over Income Distributions: Evidence from a Choice Experiment* est un travail conjoint avec Sophie Cètre, Claudia Senik, et Thierry Verdier. À l'aide d'une expérience de choix en laboratoire, nous évaluons l'importance relative de différents motifs qui sous-tendent les attitudes envers l'inégalité des revenus. Nous sollicitons les préférences des sujets concernant des paires de distributions de gains au sein de petits groupes, dans un cadre semblable à celui d'une entreprise. Nous constatons que les distributions qui satisfont au critère de dominance de Pareto suscitent un suffrage unanime : tous les sujets préfèrent une plus grande inégalité, à condition qu'elle améliore légèrement la situation de chacun. Ceci est vrai, que les gains soient basés sur le mérite ou la chance. L'unanimité ne se brise que lorsque les positions des sujets dans la distribution des revenus sont fixes et connues ex ante. Cependant, même dans ce cas, 75 % des sujets préfèrent les distributions à dominante Pareto. Mais on observe tout de même que 25 % des sujets rejettent ces inégalités en diminuant les revenus au sommet de la distribution *sans* redistribuer cet argent. Par ailleurs, nous montrons qu'une majorité, nous montrons qu'une majorité de sujets adopte une distribution plus égale si leur propre revenu ou l'efficacité globale n'est pas en jeu. Lorsque leur propre revenu est en jeu et que la somme des gains n'est pas affectée, 20 % des sujets sont prêts à payer pour obtenir un degré d'inégalité plus faible.

Le deuxième chapitre, *Principal's Distributive Preferences and the Incentivization of Agents* est un travail conjoint avec Sophie Cètre. Dans ce chapitre, nous nous demandons si les préférences distributives des employeurs affectent l'allocation des incitations au sein des entreprises. Nous documentons une relation robuste entre les préférences distributives des employeurs français et les méthodes d'incitation qu'ils choisissent pour leurs travailleurs. Pour établir la causalité, nous menons une expérience de laboratoire Principal-Agent, dans un cadre semblable à celui d'une entreprise. Dans l'expérience, les sujets sont aléatoirement placés dans la position d'employeur (principal) ou de travailleur (agent). Les principaux doivent choisir des contrats de salaire à la pièce pour deux travailleurs qui diffèrent en termes de capacité. Les travailleurs, quant à eux,



---

doivent choisir un niveau d'effort qui n'est pas contractuel. Les employeurs sont soit payés en proportion de la production (traitement Stakeholder), soit payés à un salaire fixe (traitement Spectator). Nous étudions comment les employeurs acceptent de faire des compromis entre les préoccupations d'incitation (motiver les travailleurs à maximiser la production) et leurs propres préférences normatives distributives. Nous constatons que, en présence d'aléa moral, les employeurs ont des préoccupations égalitaires, mais qu'ils sont sensibles aux incitations, à la fois à la marge extensive et intensive. Nous caractérisons l'hétérogénéité des préférences distributives en introduisant une fonction d'utilité qui incorpore les autres préférences du principal. Nous l'estimons à l'aide d'un modèle "finite mixture" et montrons que les principaux peuvent être classés en trois catégories : maximisation de la performance, faible et fort égalitarisme.

Le troisième chapitre, *Motivating Beliefs in a Just World* étudie si les individus déforment leurs croyances sur l'importance relative de l'effort et de la chance pour se motiver à faire des efforts. À cette fin, je développe un nouveau plan expérimental dans lequel l'expérience passée de succès ou d'échec sert de signal sur l'importance réelle de l'effort dans la réussite. Pour vérifier si les individus déforment leurs croyances pour motiver un effort futur, je fais varier le moment où les sujets sont informés d'une tâche exigeant un effort qui sera effectuée plus tard dans l'expérience. Les sujets qui reçoivent l'information avant la révélation des croyances sont incités à déformer ces croyances pour motiver l'effort à fournir dans la tâche future. Les résultats montrent que ces individus sont plus susceptibles de croire que leur effort est important pour réussir. La distorsion des croyances motivantes est particulièrement prononcée chez les sujets qui reçoivent des informations dissuasives sur l'état réel du monde, c'est-à-dire que le succès dépend de la chance plutôt que de l'effort. Je vérifie en outre si la distorsion des croyances motivantes affecte la volonté des sujets de distribuer de l'argent entre deux *autres* individus en tant que spectateur neutre. Je ne trouve aucune preuve que le comportement distributif diffère entre les deux groupes de traitement. Ces résultats suggèrent que les croyances des individus en matière de chance et d'effort ne dépendent pas seulement des événements passés ou actuels, mais sont également endogènes à la structure d'incitation à laquelle les individus s'attendent à être confrontés.

# Acknowledgments

This dissertation benefited from the help of numerous people and I want to take the opportunity to thank them. Without them, this four-year-long journey would have been much less insightful, and I would certainly not be at the point where I am right now.

Above all, I want to thank my two advisors, Claudia Senik and Nicolas Jacquemet. Without their terrific support, I would not have been able to complete this dissertation. I first met Claudia in the second year of my Master's, where I started working with her on a project on preferences over income distribution, and which would ultimately become Chapter 1. With this project commenced a fruitful mentorship. During the last four years, Claudia has supported and helped me wherever she could. The feedback she gave me and the discussions we had, shaped and improved all three chapters of this thesis. I first met Nicolas as a Master's student who took his excellent course on experimental economics. For the PhD, I was looking for a co-supervisor; given his expertise in experimental economics, Nicolas was the perfect fit. Over the years, we developed a very fruitful advisor-advisee relationship, and I am very thankful for the invaluable feedback and guidance he provided me on my research and professional development. I am also grateful to Nicolas for integrating me into the Economics and Psychology group.

This thesis would have not been the same without the collaborators that have directly contributed to the first and second chapters. This is particularly the case for Sophie Cêtre with whom I co-authored the first two chapters. After having already met previously during the Master, it was through our collaboration on Chapter 1 that we really got to know each other. While working on Chapter 1 for more than a year, we saw a gap in the literature and decided to tackle a new research question which became Chapter 2. Writing these chapters with Sophie was an extremely enriching experience and a good example of a smooth and fruitful collaboration. I also want to thank Thierry Verdier for his time, energy, and insights that shaped Chapter 1 of this thesis.

During my thesis, I have received exceptional support from some professors and I want to thank them on this occasion. First, I want to thank Béatrice Boulu-Reshef who was part of my thesis committee. She followed my work from the first day onward and was always there if I had questions and needed her advice. Second, I want to thank Roberto Galbiati and Shachar Kariv for their invaluable support on the academic job market. Third, this thesis profited from research visits at the University of California, Berkeley, and the Wissenschaftszentrum Berlin. These visits would not have been possible without the sponsorship by Shachar Kariv and Dorothea Kübler. The time spent in Berkeley and Berlin significantly helped to improve the chapters of the thesis through interacting with Shachar and Dorothea, as well as the scholars

---

based at these institutions. At this point, I also want to thank the organizers of the FAIR Spring School of Behavioral Economics at Rady School of Management, UC San Diego, for organizing and allowing me to participate in this wonderful event. This event is truly unique and helped me grow as a researcher through significantly widening my academic horizon.

I, furthermore, want to thank Urs Fischbacher, Marc Fleurbaey, and Bertil Tungodden for being part of my jury. I found the discussions we had during the pre-defense very enriching and they certainly improved the chapters in the months preceding the defense.

I also want to thank collaborators that did not contribute directly to the chapters of this thesis but with whom I worked or am still working on other projects in the course of the last four years. These projects have indirectly contributed to the successful completion of this thesis by making me learn new empirical methods, dive into new literature, or simply draft papers. Thanks go to Roberto Galbiati and Emeric Henry with whom I work on a fascinating project on laws and norms, Fabrice Murtin and Louis Putterman with whom I worked on data from the TrustLab project, and Morten Støstad with whom I work on a project on inequality externalities.

The chapters of this thesis also benefited indirectly from the experiences I made teaching at University Paris 1 and PSE. In these four years, I had the pleasure of working with Francis Bloch, Stéphane Gauthier, Christos Ioannou and Jean-Phillipe Tropeano and I want to thank them on this occasion for their support and collaboration.

During the thesis, there have been numerous instances, where I met with researchers at PSE (and other institutions) and asked them for their help or advice. Risking that I forget somebody, I want to thank Liza Charrion, Andrew Clark, Fabrice Etilé, Jeanne Hagenbach, Fabrice LeLec, Suanna Oh, Leonardo Pejsachowicz, Jean-Christophe Vergnaud and Nikhil Vellodi. I also want to thank PSE, Paris 1 and CEPREMAP for institutional support and I want to thank in particular Véronique Guillotin for managing the PhD cohort, as well as Sylvie Lambert and David Margolis for directing the PhD program and Job Market training.

A big thank you also go to the fellow PhD students that I met at PSE. During the last years, I had the chance of meeting and working alongside a lot of brilliant young researchers. Most of the time during my PhD, until March 2020, I spend in the office. Thanks, thus, go to Manon François, Martin Jégard and Morten Støstad for making the “Moffice” a stimulating place at PSE. I want to thank the fellow students from Economics and Psychology—Quentin Cavalan, Nina Rapoport, Sofia Mun, Justine Jouxte, Hélène Korda, Bin Wang, Olivier Savrimoutoo, Rémi Yin, Antoine Hémon—for the exciting hours on Zoom discussing papers, after March 2020, and interesting discussions on Fridays at l’Intermède before March 2020. I also want to thank other PSE students, in particular, Juan Luksic, Euiyoung Jung, Santiago Garriga, Florencia Pinto, Alvaro Zuniga, Oscar Barrera, Sara Signorelli, Yajna Govind, Paolo Santini, Elias Bouacida, Giulio Iacobelli, Marion Leroutier, Paul Brandily, Laura Khoury, Sarah Schneider, and many many more, for the pleasure to interact with them regularly over the last over years.

During my visits, I met a lot of bright researchers that made these research stays a truly enriching experience. I want to thank my fellow scholars in Berkeley, in particular, Kevin Ducbao Tran, Ariane Salem, Benedikt Busch, Sebastian Rast, David Birke and Paulus Wagner, for making these 4 months a truly enriching experience. I would also like to thank Robert Stüber

---

for inspiring discussions and mutual support on the Job Market: first in San Diego, then at the WZB during my visit, and, finally, on the phone wherever we were. On this occasion, I want to thank all the other researchers I met at WZB and Berkeley for making these welcoming and inspiring places to visit.

Ich möchte mich auch herzlich bei meinen Freunden in Berlin und Umgebung bedanken. Besonderen Dank gehen an Arnim Seidlitz für inspirierende Diskussionen, u.a. über Academia und Research, bei Herthaspielen oder wo auch immer wir gerade waren; Kai Rövenich für tief-sinnige Debatten, u.a. über Politik, Ökonomie und Management, beim Erklimmen diverser Gipfel des Balkans, Marokkos, Kataloniens und des Dreiländerecks; Nicolai Braun für inspirierende Diskussionen, u.a. über das politische System und die Gesellschaft, in diversen Berliner Bars, Felix Sodemann und Reentje Streuter für spannende Reflexionen, u.a. über die politische Ökonomie der Türkei, und Florence Pache für ihre tolle Unterstützung in quasi jeder Hinsicht.

An dieser Stelle möchte ich mich auch herzlich bei meiner Familie bedanken. Danke an Fritz, Lisa, Valery, Adrien, Johanna, Alina, Denis und Lauren für die Fondue-Abende, Wanderungen und Skiwochenenden in der Genfersee-Region, sowie den gemütlichen Nachmittagen in der Wandlitzer Datsche, die mir erlaubt haben etwas Abstand von der Dissertation zu bekommen. Vielen Dank gehen auch an Großmutter für spannende Telefonate über die Ökonomie der Ungleichheit. Zuletzt möchte ich mich bei meinen Eltern für den bedingungslosen Rückhalt in den letzten vier Jahren bedanken, sowie den Reisen nach Eritrea und Äthiopien, die meinen persönlichen und akademischen Horizont erweitert haben, als auch der Möglichkeit an der Dissertation im schönen Garten in Berlin-Düppel während der Sommermonate zu arbeiten.

Je voudrais remercier tous mes amis parisiens pour le soutien les derniers quatre ans. Pour moi, la thèse sera toujours liée à mon séjour à Paris où j'ai découvert cette belle ville, la culture française et appris cette jolie langue. Merci à Sophie Liu, Emilien Nory, Gaurav Assat, Willy Go, Ava Guez, Alexia Lachavanne, Simon Valayer, Karim El Mernissi et Julien Bergeot pour les discussions passionnants sur le balcon de la coloc ; merci aussi à tous mes amis du COGE pour ces beaux moments musical ; merci à Monica Partarrieu, Bosco et Salvadore pour les après-midi aux parcs parisiennes ; merci à Lorenz Adami et Jonathan Stöterau pour ramener un peu de Berlin à Paris ; merci à Florian Grosset pour les nombreux déjeuners et des lectures de mes papiers préliminaire. Finalement, je voudrais remercier beaucoup Roser Graell pour être à mon côté pendant ces quatre ans, pour les relectures des passages écrit en français, pour me pousser de que parler français, de me garder sain d'esprit pendant ces quatre ans et surtout pour beaucoup des beaux moments.

Un grand merci à tous,  
Max Lobeck

# Contents

<b>Dissertation Abstract</b>	<b>i</b>
<b>Résumé de la thèse</b>	<b>iii</b>
<b>Acknowledgments</b>	<b>v</b>
<b>Table of Contents</b>	<b>xi</b>
<b>General Introduction</b>	<b>1</b>
1 Motivation . . . . .	1
2 Survey of the existing literature . . . . .	4
2.1 Motives behind attitudes towards inequality . . . . .	4
2.2 What shapes attitudes towards inequality? . . . . .	13
2.3 What are the consequences of attitudes towards inequality? . . . . .	22
3 Empirical methods used in this thesis . . . . .	26
4 Summary of each chapter . . . . .	28
5 References . . . . .	36
References . . . . .	46
<b>1 Preferences over Income Distribution</b>	<b>47</b>
1 Introduction . . . . .	48
2 Design of the experiment . . . . .	50
2.1 Overview . . . . .	50
2.2 The Choices . . . . .	51
2.3 Merit versus Luck . . . . .	53
2.4 Control variables . . . . .	54
2.5 Implementation . . . . .	55
3 Results . . . . .	55
3.1 Choices behind the veil of ignorance . . . . .	55
3.2 Choices with known position . . . . .	57
3.3 Observable heterogeneity . . . . .	60
4 Within-subject variation . . . . .	61
5 Robustness Checks . . . . .	61
5.1 Are subjects consistent within a category of choices? . . . . .	61

5.2	Is choosing behind the veil of ignorance the same as choosing between two lotteries? . . . . .	62
5.3	Does the significant treatment effect pick up on over-confidence? . . . . .	62
6	Conclusion . . . . .	63
7	References . . . . .	64
	References . . . . .	66
<b>Appendices to Chapter 1</b>		<b>67</b>
1.A	Tables and Figures . . . . .	67
1.A.1	Tables . . . . .	67
1.A.2	Figures . . . . .	73
1.B	Hypothetical Decisions . . . . .	81
1.C	Screenshots of the Experiment . . . . .	82
1.D	Post-experimental Questionnaire . . . . .	88
1.E	Experimental Instructions . . . . .	91
1.E.1	Introduction of the experiment . . . . .	91
1.E.2	Part 1 . . . . .	91
1.E.3	Part 2 . . . . .	92
1.E.4	Real effort task . . . . .	95
1.E.5	Part 3 . . . . .	96
1.E.6	Questionnaire ( <i>Luck and Merit groups</i> ) . . . . .	97
<b>2</b>	<b>Principals' Distributive Preferences and the Incentivization of Agents</b>	<b>98</b>
1	Introduction . . . . .	99
2	Trading off distributive and incentivization concerns . . . . .	102
3	Managers' preferences: survey evidence . . . . .	104
4	Experimental design . . . . .	106
4.1	Lab setting . . . . .	107
4.2	Workers . . . . .	107
4.3	Principals . . . . .	109
4.4	Summary statistics . . . . .	113
5	Results . . . . .	114
5.1	Effort choices and effort beliefs . . . . .	114
5.2	Belief-based contract trade-offs . . . . .	114
5.3	Principals' choices . . . . .	117
6	Structural Characterization of Distributive Preferences . . . . .	122
6.1	Making distributive decisions ex-ante . . . . .	122
6.2	Utility function specification . . . . .	123
6.3	Pooled results . . . . .	125
6.4	Characterizing heterogeneity in preferences . . . . .	126
6.5	Results from the finite mixture model . . . . .	129
7	Conclusion . . . . .	131

8	References . . . . .	133
	References . . . . .	136
<b>Appendices to Chapter 2</b>		<b>137</b>
2.A	Tables and Figures . . . . .	137
2.B	Robustness Checks . . . . .	147
2.C	Deriving the density of the likelihood on the individual level . . . . .	148
2.D	Counterfactual analyses . . . . .	149
2.D.1	Including social comparisons by agents . . . . .	149
2.E	Instructions . . . . .	156
2.E.1	Principals . . . . .	156
2.E.2	Workers . . . . .	159
2.F	Comprehension test . . . . .	160
2.F.1	Questions principals . . . . .	161
2.F.2	Workers' questions . . . . .	164
2.F.3	Comprehension test performance . . . . .	166
2.G	Aptitude test . . . . .	167
2.G.1	French Questions . . . . .	167
2.G.2	Logic questions . . . . .	168
2.G.3	General knowledge . . . . .	169
<b>3</b>	<b>Motivating Beliefs in a Just World</b>	<b>170</b>
1	Introduction . . . . .	171
2	Framework and Hypotheses . . . . .	176
3	Experimental Design . . . . .	183
3.1	Work Phase 1 . . . . .	184
3.2	Feedback Phase . . . . .	185
3.3	Information Treatment . . . . .	186
3.4	Belief Elicitation Phase . . . . .	187
3.5	Redistribution Stage . . . . .	188
3.6	Work-Phase 2 . . . . .	189
3.7	Post-experimental questionnaire . . . . .	190
3.8	Summary of the variation generated through the design . . . . .	190
4	Procedure . . . . .	191
5	Results . . . . .	193
5.1	Are signals informative? . . . . .	193
5.2	Beliefs across motive- and no-motive-group . . . . .	194
5.3	Who distorts luck-effort beliefs to motivate effort? . . . . .	199
5.4	Distributive Behavior . . . . .	206
6	Conclusion . . . . .	209
7	References . . . . .	211
	References . . . . .	214

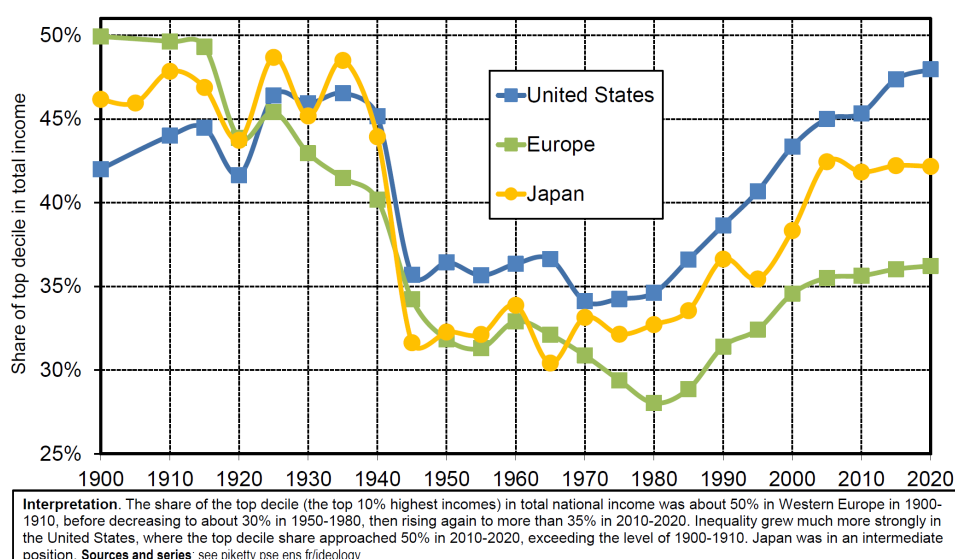
<b>Appendices to Chapter 3</b>	<b>215</b>
3.A Figures . . . . .	215
3.B Tables . . . . .	216
3.C Questionnaire . . . . .	240
3.D Attention Checks . . . . .	240
3.E Pre-analysis plan . . . . .	242
3.F Instructions . . . . .	247
<b>4 Résumé détaillé en français</b>	<b>273</b>
1 Motivation . . . . .	273
2 Revue de la littérature . . . . .	276
2.1 Raisons des attitudes envers l'inégalité . . . . .	276
2.2 Qu'est-ce qui influence les attitudes envers l'inégalité? . . . . .	286
2.3 Quelles sont les conséquences des attitudes à l'égard de l'inégalité? . . . . .	297
3 Méthodes empiriques utilisées dans cette thèse . . . . .	301
4 Résumé des chapitres . . . . .	304
4.1 Motivating Beliefs in a Just World . . . . .	307
5 Références . . . . .	312
Références . . . . .	321



# General Introduction

## 1 Motivation

Figure 1: Inequality 1900 to 2020 in Western Europe, Japan, and the US. Source: Piketty (2020)



Economic inequality has increased dramatically around the world: While the Top 10% income share in the US was at roughly 35% in 1980, it has risen to nearly 50% in 2020. As shown in Figure 1, income inequality has now reached a dimension not seen since before the Second World War. The pictured revival of inequality is not confined to the US but characterizes a global dynamic. Inequality has been steadily rising in Europe, Japan, China, Russia and, in particular, India, where the Top 10% now captures more than 55% of total income (Alvaredo et al., 2018).

The unveiling of this striking dynamic in economic inequality brought inequality back to the center of the academic and political debate.<sup>1</sup> Indeed, President Barack Obama coined economic inequality as “the defining challenge of our time” (Obama, 2013). This debate was—and still is—multifaceted: It asks about the origins and consequences of economic inequalities, as well as posing the question of whether these inequalities are justified in the first place. Individuals’ attitudes towards economic inequality are fundamental for answering each of these questions and the aim of this dissertation is, thus, to advance our understanding of these attitudes.

<sup>1</sup>This was made possible by the publication and dissemination of new data on top-income shares through academic studies such as Atkinson et al. (2011), Alvaredo et al. (2013) and Piketty (2014).

**Why should we study attitudes towards economic inequality?** Economic inequality is fundamentally a societal *choice*. Societies have the means to reduce inequality by redistributing income, for example, through levying taxes that rise progressively in income.<sup>2</sup> Attitudes towards economic inequality are central to understanding why they make this choice. As governments set these policies, they are influenced—at least to some extent—by the preferences and demands of their citizens. Thus, understanding how people form attitudes towards inequalities is indispensable if we want to better understand the dynamics of economic inequality itself. As an example, the conservative revolution of the 1980s popularized the sentiment that high levels of redistribution hampers economic growth. This compounded the meritocratic narrative that the rich earned their high incomes through hard work; thereby justifying pre-tax inequality (Piketty, 2020; Sandel, 2020). These narratives and beliefs shaped economic policy in the last decades and led to a dismantling of the Welfare state in Western economies, even by center-left governments.

While these dynamics are partly driven by the supply of ideas in the 1990s and early 2000s, they are also the product of a popular demand for lower levels of redistribution.<sup>3</sup> One striking feature of this period is that, at least in the US, the rise in income inequality was not accompanied by an increased demand for redistribution (Ashok et al., 2015). This may partly explain why we did not see greater efforts to curb inequality. It, thus, illustrates the importance of understanding *attitudes* towards economic inequality for better understanding the *evolution* of inequality.

Studying attitudes towards economic inequality is, furthermore, essential for understanding the consequences of inequality. Assuming that individuals care about inequality implies that it affects their utility and behavior. Attitudes towards inequality may also affect outcomes and behavior in the labor market: If individuals dislike working for an organization with high levels of inequality, inequality matters for the functioning of incentive contracts, well-being at work, and labor supply.

Putting aside that attitudes towards inequality are important for better understanding the dynamics and consequences of inequality, their nature makes them a fascinating topic to study in itself. This is mainly because attitudes towards inequality are (i) far from homogeneous, (ii) dependent on the importance of competing motives that may or may not conflict with each other, (iii) subject to behavioral biases, and (iv) context dependent.

**Main research questions** While the dissertation comprises three self-contained research papers, they all speak to one overarching research question: *How does the economic environment shape attitudes towards economic inequality?* As I will show below in Section 2, the prior literature has made great strides in identifying why individuals may dislike inequality. However, our understanding of how these motives interact with the economic and institutional environment is still limited. Advancing our knowledge at this frontier is crucial because it is key

---

<sup>2</sup>That policies matter for the level of inequality is supported by the observation that the dynamics of inequality correlate with the top marginal tax rate, a measure of tax progressivity (Piketty, 2014; Saez and Zucman, 2020).

<sup>3</sup>Assessing the causal effect of citizens' preferences for redistribution and implemented levels of redistribution is difficult. This is partly because that attitudes towards inequality are not the only thing that matters in elections. Thus, it may be hard to disentangle the effect of views on inequality with other political goals that may be on the ballot. Nonetheless, surveys have shown that voters care about distributive fairness in elections (see Almås et al., 2020, for the US and Norway).

for better understanding the dynamics of inequality acceptance and how policies shape attitudes towards economic inequality.<sup>4</sup>

Chapter 1 systematically assesses how preferences over income distributions depend on the context in which they are revealed. To that end, it will jointly study how attitudes towards inequality depend on (1) the relative *shape* that inequalities take, (2) *certainty* of holding a given position within an income distribution, and (3) whether this position is determined through differences in *luck* or differences in *effort*. The results help to better understand in which contexts individuals may accept or reject inequality. For example, large inequalities that only favor the very top attain unanimous suffrage if all individuals weakly benefit from these inequalities and all individuals have a chance to benefit from these inequalities. This acceptance of inequality is independent of how one earned this rank. However, unanimity breaks once positions in the distribution are fixed and some may not benefit from inequalities. Then, a sizable minority of these subjects reject such inequalities by decreasing incomes at the top of the distribution *without* redistributing that money and, thus, without making others better off. This illustrates the importance of income prospects in shaping inequality attitudes.

Chapter 2 studies managers' attitudes towards inequalities in the context of the firm. The research question asks whether managers' attitudes towards inequalities affect the provision of incentives in firms and whether they are affected by the incentive scheme a manager faces herself. In the paper, we find that managers' attitudes towards inequality affect the choice of incentives. We further show that distributive concerns are partly—but not fully—crowded out by the incentives that a manager faces herself. This implies that the effect of managers' distributive preferences on incentive schemes persists, even if the manager is incentivized to act against her fairness views. It connects to the overarching research question of this thesis by applying the relevance of inequality attitudes to the context of the firm and by testing how they depend on the incentive environment the decision maker faces herself.

Chapter 3 studies a crucial determinant of attitudes towards inequality: beliefs about the source of inequality. As I will show below, individuals are more willing to accept inequality if they believe that it reflects differences in effort rather than luck. In the chapter, I ask whether individuals distort these beliefs for motivating purposes. Through that, I test whether they are endogenous to the incentive structure that individuals face in their economic environment. The analysis reveals that individuals indeed distort luck-effort beliefs for motivating purposes, leading to an overestimation of the importance of effort relative to luck in getting ahead.

The remainder of the introduction is structured as follows: Section 2 features a survey of the existing literature on inequality attitudes; Section 3 outlines the empirical methods that I use for answering the research questions; and Section 4 features a detailed summary of each chapter. The reader can find a French version of the introduction at the end of the dissertation starting at Page 273.

---

<sup>4</sup>This thesis studies individuals' attitudes towards *economic* inequality. Throughout the thesis, I will often refer to these attitudes towards economic inequality simply by attitudes towards inequality. Attitudes towards other forms of inequality, such as gender or race inequality, are not the subject of this dissertation but are a fascinating research agenda.

## 2 Survey of the existing literature

This section provides a survey of the literature on attitudes towards economic inequality that will help the reader put this thesis's contributions into the wider context of the existing literature. The review starts by asking why and whether individuals hold preferences over income distributions. I will then summarize the existing evidence on how these attitudes towards inequality are shaped by experiences and the economic environment, before concluding the review with a presentation of the *consequences* of inequality attitudes.

### 2.1 Motives behind attitudes towards inequality

Throughout this thesis, I will argue that individuals care about economic inequality. One may thus ask why individuals should form preferences over income distributions in the first place. In the following section, I will outline concepts developed in recent years and point to evidence in favor or against these theoretical concepts.

#### Self-regarding motives for inequality attitudes

The literature that asks why we form attitudes towards inequality mostly focuses on other-regarding preferences. Nonetheless, there are several situations where individuals may prefer some distributions of income over others for completely self-regarding reasons.

One reason individuals want to reduce inequalities is that they expect to benefit from reducing these inequalities through redistribution. This idea is particularly present in political economy models that build on the framework introduced by Romer (1975) and Meltzer and Richard (1981). Empirical studies that use survey evidence, often identify a negative correlation between the demand for redistribution and current or expected household income (see e.g. Alesina and La Ferrara, 2005; Alesina and Giuliano, 2011; Owens and Pedulla, 2014). Laboratory experiments that exogenously assign initial levels of income often find that, on average, individuals assign a positive weight on the self-regarding part of their utility function (e.g. Durante et al., 2014). Further, a non-negligible minority (16.75%) of the US-adult population has found to be completely selfish, meaning that they always choose the option that maximizes their income (Fisman et al., 2017).<sup>5</sup> The link between maximizing one's income and inequality attitudes may sound obvious at first sight. However, the economic environment influences how selfish motives map into inequality attitudes, often showing, at first sight, surprising results. I will return to this literature in Section 2.2, when discussing how uncertainty shapes attitudes towards inequality.

#### Other-regarding preferences over income distributions

While selfish motives for redistribution can be a powerful motivation to explain why individuals reject or embrace inequalities, there are numerous instances where this is insufficient to ratio-

---

<sup>5</sup>The experiment used by Fisman et al. (2017) uses a modified dictator game that varies the cost of giving, as well as the efficiency-equality trade-off within subjects. Repeated decision allow the authors to identify dictators' selfishness and their willingness to trade-off equality and efficiency (see below).

nalize behavior. One way of conceptualizing this divergence is by assuming that individuals intrinsically care about the prevailing distribution of income.

**Conceptualizing fairness** To put other-regarding attitudes towards inequality in a common framework, I assume, in the spirit of Cappelen et al. (2013), that if the prevailing distribution of income deviates from the distribution of income that an individual views as fair, she suffers from a moral cost. This can be characterized through the following value function of a given individual  $i$ :

$$V_i(y_i, \mathbf{y}) = y_i - \alpha_i M(|D(\mathbf{y}) - D_i^F(\mathbf{y})|) \quad (1)$$

where  $y_i$  is the individual's own income,  $\mathbf{y}$  is a vector of incomes that may or may not include  $y_i$ ,  $D(\mathbf{y})$  is the prevailing distribution of incomes included in  $\mathbf{y}$ ,  $D^F(\mathbf{y})$  is the distribution of income that  $i$  considers to be the *fair* distribution within the set of feasible income distributions, denoted as  $\Omega$ , and  $M(\cdot)$  is the moral cost that occurs if  $D(\mathbf{y}) \neq D_i^F(\mathbf{y})$  and that is increasing and weakly convex in the difference of the two. Finally,  $\alpha_i$  characterizes the relative importance of maximizing one's own income and minimizing  $M(\cdot)$ .

Before continuing, I want to clarify how to think about this very general value function. First,  $D^F(\mathbf{y})$  depends on the alternatives and resources available (i.e.  $\Omega$ ). Second,  $D^F(\mathbf{y})$  may be context-dependent. For example, people may think differently about fairness at the workplace compared to when they think about redistributing income on the societal level. This implies that  $\mathbf{y}$  may include the payoffs of all individuals in the society or a subset of individuals, e.g. those at the workplace. Third,  $D^F(\mathbf{y})$  may look different if your income is at stake while deciding among income distributions. Fourth,  $D^F(\mathbf{y})$  may differ across individuals, a topic that I will touch on below when discussing the heterogeneity in fairness ideals.

**What enters  $D^F(\mathbf{y})$ ?** The chapters of my thesis build on theoretical approaches that inform about the shape of  $D^F(\mathbf{y})$ . These include normative theories of distributive justice developed by moral philosophers (see Konow, 2003, for a thorough review) as well as the approaches characterized by social preference models developed by behavioral economists (see Cooper and Kagel, 2016, for a thorough review).<sup>6</sup>

One reason for why individuals may prefer some income distributions over others is that they are interested in maximizing social welfare. *Utilitarianism* is the most prominent theory in this spirit. It was introduced by Bentham (1789) and later developed by Sen (1979)'s Welfarist approach. It argues that income should be distributed such that it maximizes the sum of

---

<sup>6</sup>In this review of the literature, I will not discuss approaches based on reciprocity, which are very influential in the social preference literature. I made this decision because the chapters of this thesis do not speak to this literature. Nonetheless, I do believe that reciprocity is a powerful motivator in accepting inequalities.

utilities. To put it differently, utilitarianism stipulates that the distribution of income should aim to maximize the sum of individual well-being or happiness.

$$D^F(\mathbf{y}) = \max_{D(\mathbf{y}) \in \Omega} \left\{ \sum_{k=1}^N U_i(y_k) \right\} \quad (2)$$

On the societal level, utilitarian arguments can lead to a demand to reduce high inequalities via redistribution of income if one assumes that marginal utility is declining with income (Dalton, 1920). Given the subjective nature of utility and the lack of a quantifiable metric for cardinal utility, it is relatively hard to identify distributive preferences that follow the utilitarian principle. This has two reasons: First, those who have to make those decisions, from an impartial or stakeholder position, have to form beliefs about the preferences of those who are affected by the distribution decision. Second, even if individuals could submit their cardinal preferences, they have a strong incentive to deceive the body that decides on distributive policies.<sup>7</sup>

In the social preference literature, the *efficiency motive* received more attention than a truly utilitarian motive. While utilitarianism stipulates that one should maximize the sum of utilities, the efficiency motive is interested in maximizing the sum of incomes.<sup>8</sup>

$$D^F(\mathbf{y}) = \max_{D(\mathbf{y}) \in \Omega} \left\{ \sum_{k=1}^N y_k \right\}. \quad (3)$$

Efficiency concerns as a motive that informs preferences over income distributions are interesting to study because it has direct consequences for policy. Ultimately, progressive taxation is one of the most effective policies to reduce inequality. Since taxes have a distorting effect on labor supply, some output is lost, and the total pie shrinks. This tension creates a trade-off between the efficiency and the egalitarian motive, which has been at the center of numerous studies on attitudes towards inequality (see e.g. Okun, 1975).

While the above motives argue that one needs to implement the distribution of income that is the least “wasteful”, the contrasting concept of justice is *egalitarianism*. Similar to concerns about efficiency, the egalitarian fairness principle considers outcomes rather than processes: It rejects distributions of income that are unequal, no matter how they arose (Nielsen, 1985). In the literature, two theoretical approaches lead to egalitarian motives for redistribution. The first is a disinterested or normative approach that rejects inequality without considering one’s position in the income distribution. Here, the objective function characterizes a desire to minimize income differences:

$$D^F(\mathbf{y}) = \min_{D(\mathbf{y}) \in \Omega} \left\{ \sum_{k=1}^N |y_k - \tilde{y}| \right\}, \quad (4)$$

where  $\tilde{y}$  is the median income.

---

<sup>7</sup>There are, however, some older studies that use hypothetical scenarios and find some support for fairness views that are based on comparing utilities rather than incomes (Yaari and Bar-Hillel, 1984; Konow, 1996, 2003).

<sup>8</sup>One can argue that these two concepts go hand-in-hand for low quantities of income, as often used in laboratory experiments, where the utility can be assumed to be linear in income. My take is that the distinction between utilitarianism and efficiency motives is not always clearly communicated in the social preference literature.

The other approach that has been particularly prominent in the social-preference literature employs a *comparative approach* and assumes that individuals compare their income or consumption to that of their peers (Veblen, 1899; Duesenberry, 1949). The seminal paper by Fehr and Schmidt (1999) develops a framework where individuals suffer from a loss in utility by being better or worse off than other individuals. Their proposed utility function differentiates between concerns about being behind other individuals (envy) and being ahead of other individuals (aheadness aversion (Bruhin et al., 2018)) in the income distribution:

$$U_i(\mathbf{y}) = y_i - \alpha_i \frac{1}{N-1} \sum_{i \neq j} \max\{y_j - y_i, 0\} - \beta_i \frac{1}{N-1} \sum_{i \neq j} \max\{y_i - y_j, 0\}. \quad (5)$$

One key assumption that the authors make is that individuals are more envious than aheadness-averse, i.e.,  $\alpha_i \geq \beta_i$ . How to incorporate difference aversion in the spirit of Fehr and Schmidt into the framework sketched out in (1)? One way would be to assume that  $D^F(\mathbf{y})$  follows (4). However, the weight the agent puts on the moral term in (1),  $\alpha$ , differs whether she is above or below the median of the income distribution.<sup>9</sup>

While egalitarianism deems all inequalities unfair, the *difference principle*, formulated by Rawls (1971), argues that one should reject inequalities unless they help the worse off. Preferences over income distributions then follow a maximin rule:

$$D^F(\mathbf{y}) = \max_{D(\mathbf{y}) \in \Omega} \left\{ \min\{y_1, \dots, y_N\} \right\}. \quad (6)$$

Identifying Rawlsian preferences is difficult. The theory argues that individuals will decide on such an income distribution when choosing from behind the veil of ignorance and attempts to recreate the conditions characterized by Rawls showed weak demand for Rawlsian justice (Michelbach et al., 2003; Frohlich et al., 1987).

Nonetheless, the idea that inequality is acceptable if they benefit the worse-off member of the group is echoed by the model developed by Charness and Rabin (2002). There, individuals care about the income of the worse-off alongside the total payoff. This would translate into the following preferred income distribution with  $\gamma$  characterizing the weight put on the income of the worse-off versus efficiency:

$$D^F(\mathbf{y}) = \max_{D(\mathbf{y}) \in \Omega} \left\{ \gamma \min\{y_1, \dots, y_N\} + (1 - \gamma) \sum_{k=1}^N y_k \right\}. \quad (7)$$

Similar to Charness and Rabin (2002), several theoretical approaches aim to combine the above-listed concepts in a single framework. This is particularly useful for the empirical characterization of attitudes towards inequality because it allows for a parsimonious identification of inequality and efficiency concerns. For example, Bruhin et al. (2018) propose and estimate a model that allows for envy, aheadness aversion, spite, selfishness, and payoff maximizing utilities. The model by Andreoni and Miller (2002) diverges from the approaches surveyed above by

---

<sup>9</sup>Another famous, though less influential, inequality aversion model is Bolton and Ockenfels (2000), where an individual cares about the extent that her income diverges from the average income in the group or society.

modeling other-regarding preferences using a CES-utility function. Thereby, it nests completely selfish, Rawlsian maximin, and utilitarian preferences as a special case. Intriguingly, it allows measuring the degree to which people are difference-averse or concerned about maximizing the total payoff. Fisman et al. (2007) further advance this approach to distinguish between how individuals want to distribute income between themselves and others as well as within a group of others, making a clear distinction between altruism on the one side and distributive preferences on the other side.

The above-cited literature argues that individuals have an intrinsic preference for equality or efficiency. However, one can also assume that individuals always reject an equalization of income because they view the prevailing income distribution as fair:

$$D^F(\mathbf{y}) = D(\mathbf{y}). \tag{8}$$

In the literature, this is typically labeled as the libertarian fairness view (see e.g. Cappelen et al., 2007) and builds on the tradition of libertarian thinkers such as Friedrich A. Hayek that oppose any redistribution of inequality from market transaction.<sup>10</sup> This fairness view is outcome based because the prevailing income distribution is viewed as fair, no matter how it arose.

One thing that the above-listed concepts have in common is that they are purely *outcome-based* or *consequentialist*. In other words, people are not concerned *why* an income distribution attains the shape it attains but rejects or accepts it based on its form, compared to the feasible alternatives. This changes if one assumes that the *process* that generates the income distribution influences inequality attitudes which were conceptualized by theories of desert and equity (Konow, 2003) that I refer to as the meritocratic or liberal egalitarian fairness principle throughout this thesis.

The starting point is to acknowledge that income differences may arise for different reasons such as luck, birth, choice, and effort. The key criterion is that one should equalize differences in income that are outside an individual's control. In contrast, those for which an individual is responsible (Fleurbaey, 2008) or can be held accountable (Konow, 1996) due to their individual choices should not be equalized (see Fleurbaey, 2008, for an introduction to the philosophical foundation of liberal egalitarian theories of justice). This notion can be applied to risk-taking (Dworkin, 2002) and to the extent that effort should be rewarded (Cappelen and Tungodden, 2009). It can be captured (with much simplification) as follows in  $D^F(\mathbf{y})$ :

$$D^F(\mathbf{y}) = \mathbf{x} + \min_{D(\mathbf{y}) \in \Omega} \left\{ \sum_{k=1}^N |w_k - \tilde{w}| \right\}, \tag{9}$$

where we assume that individual income  $y_i$  is made up of  $x_i$  that stands for the part of total income for which an individual can be held accountable and  $w_i$  which is the part of income for which an agent cannot be held accountable. In (9),  $\mathbf{x}$  is, thus, a vector characterizing incomes for which individuals can be held accountable. In contrast,  $w_k$  is the part of income for which

---

<sup>10</sup>Interestingly, Hayek's argument in favor of preserving inequality was *not* based on fairness or meritocracy (see Sandel, 2020, for a discussion).



individual  $k$  cannot be held accountable,  $\tilde{w}$  is the median non-accountable income.

The question that arises from this is: For what should we be held accountable? What enters in  $w_i$  and what enters in  $x_i$ ? This is in the eye of the beholder and depends on individuals' definition of what is within an individual's control and whether this matters for distributive fairness. Some may, for example, argue that the willingness to exert effort is an individual trait, given by birth, and thus outside individual control. Others may argue that one should not equalize differences in income that arise due to differences in talent. And others again argue that one should.<sup>11</sup> While these distinctions depend on what people view as relevant dimensions of accountability, which can be characterized as a preference or conviction, there is also the question of how individuals perceive reality. Individuals may hold similar fairness ideals but hold different beliefs on whether individuals are truly accountable for their outcome because their beliefs about the economic environment diverge (Alesina et al., 2020). For example, one key policy recommendation from meritocratic thinking is to call for equality of opportunity (Arneson, 1989; Roemer, 1998; Sandel, 2020). The extent to which this is achieved is a belief that varies from individual to individual (Alesina et al., 2018), and this belief is the focus of the third chapter.

**What are the relevant motives?** The multiplicity of the motives reviewed above opens up the question of the relevant other-regarding motives behind inequality attitudes. This is, foremost, an empirical question. The empirical literature in experimental economics has a great tradition of identifying and disentangling the different other-regarding motives individuals hold. One part of the literature uses classical games such as the ultimatum game (Güth et al., 1982) or the dictator game (Kahneman et al., 1986b; Forsythe et al., 1994) to measure inequality aversion. A lot of the models cited above were inspired by the behavioral regularities found from these games, which showed that 40 to 50% of low offers are typically rejected in ultimatum games (Ledyard, 1995; Cooper and Kagel, 2016) and subjects in dictator games give on average 28% of their initial payoff (Engel, 2011). Bellemare et al. (2008) for example, fit an augmented version of the Fehr and Schmidt model to behavior in an ultimatum and dictator game and find considerable evidence of envy in the Dutch population but less for aheadness aversion. A problem one faces when identifying inequality attitudes using these games is that they cannot control for reciprocal motives and/or distinguish between the different distributive motives. To control for reciprocal motives and disentangle the different distributive motives from each other, a large experimental literature (including the chapters of this thesis) uses augmented dictator games, where the dictator can choose between a set of income distributions. By varying the set of income distributions ( $\Omega$ ), the experimenters can then disentangle the different motives listed above: selfishness, envy, aheadness aversion, efficiency concerns, and maximin.

Earlier work in this literature found evidence that individuals are less concerned about inequality than they are about the payoff of the least well-off and efficiency (Charness and Rabin, 2002; Kritikos and Bolle, 2001; Engelmann and Strobel, 2004, 2007). This started a debate of whether inequality concerns are not as important as previously thought, or whether the evidence

---

<sup>11</sup>The latter is probably the most frequent way entitlement is defined in the public discourse in the sense that an economic system should give everybody the chance to "rise as far as their *talents* allow them" (Sandel, 2020).

is an artifact of eliciting preferences of economics undergraduate students (see Fehr et al., 2006). The argument is born out of the fact that distributive preferences may be heterogeneous across different sub-population. Hence, findings from one population (students) may not apply to the whole population. To get an answer to this question, one has to characterize the heterogeneity in other-regarding preferences using samples that are representative of the overall population.

Furthermore, mapping heterogeneity in distributive preferences is important to assess the consequences of other-regarding preferences because even minority preferences can sometimes have a strong influence on outcomes. For example, even selfish proposers in an ultimatum game propose a positive amount once a sufficient *minority* of receivers are inequality averse (Fehr and Schmidt, 1999). Similarly, politicians may target distributive policies so that it matches the preferences of an important group of voters (e.g. senior citizens). Standard laboratory experiments (including Chapter 2 of this thesis) have found substantial heterogeneity in distributive preferences, even though the population they study is relatively homogeneous (Andreoni and Miller, 2002; Fisman et al., 2007; Cappelen et al., 2007; Bruhin et al., 2018). This indicates that heterogeneity is likely to matter, even if the population of interest is relatively small and homogeneous (e.g. at the workplace). It also shows that findings based on population averages may be susceptible to hiding relevant preference-types.

In recent years, the use of representative samples allowed us to characterize heterogeneity through observable characteristics. Fisman et al. (2015b), for example, compare other-regarding preferences of a representative sample of Americans with those of students from elite universities. Their design, which is based on Andreoni and Miller (2002), allows identifying where individuals stand on the efficiency-equality trade-off, as well as a degree of selfishness. They show that a higher share of elite students is selfish and they are more likely to be efficient than equality-focused compared to the representative sample of Americans. This is particularly interesting if one considers that elite students are likely to hold influential positions later in life.<sup>12</sup> Other studies that use representative samples have found that a large share of the overall population is more sensitive to inequality concerns compared to efficiency (Almås et al., 2020; Müller and Renes, 2021; Epper et al., 2020). These studies also reveal that a non-negligible part of the population holds libertarian fairness views (Almås et al., 2020).

Overall, this literature shows that both egalitarian and efficiency concerns are relevant motives behind attitudes towards inequality. Efficiency concerns seem to be particularly prominent in convenience samples and less so in representative samples. Nonetheless, there are still differences in elicited preferences among studies that use similar subject pools. Future research should address to what extent this is due to the use of different elicitation methods.

The evidence cited above comes from experiments (or treatments) where the source of inequality is luck. Meritocrats, however, care about the source of inequality. Thus, a large part of the literature studies whether inequality acceptance is sensitive to the source of inequality. Surveys have studied the relevance of beliefs about luck and effort on redistributive preferences. Typically, they find that beliefs about the role of luck and effort in getting ahead are strong

---

<sup>12</sup>It, furthermore, gives empirical bite to Fehr et al. (2006)'s argument that students are more efficiency focused than the overall population.

correlates with inequality acceptance and preferences for redistribution more generally (Alesina and La Ferrara, 2005; Fong, 2001; Alesina et al., 2018). Early experimental approaches studied bargaining or dictator experiments where the role of the first mover is earned by being better at a quiz. These experiments typically found that the final bargain favors the subject that was better at the quiz compared to when the roles were determined randomly (Hoffman et al., 1994; Ruffle, 1998) or that dictators behave more selfishly after winning their role through a contest (Cherry et al., 2002).

More recent approaches use dictator games where subjects generate an initial distribution of income through their decisions. Konow (2000) pioneered this approach by asking pairs of subjects to prepare letters. Individual output was then multiplied by a price, leading to an initial distribution of output value. Dictators (either one subject in a pair or a third party spectator) then have the opportunity to redistribute the *joint* output value while being informed how much each receiver contributed to the joint output value. Across treatments, the experimenter varies whether differences in effort or (exogenous) differences in price account for differences in individual contributions to the joint output value. Konow shows that spectators typically prefer distributions that mirror differences in effort but not those that mirror differences in luck, which is in accordance with the accountability principle outlined above. Frohlich et al. (2004) report findings from a dictator game experiment with a preceding production stage and show that a large share of dictators take into account production differences. Cappelen et al. (2007) study distributive behavior where subjects are asked to make an investment decision. The return to investment is either high or low. They are then matched with somebody else and asked to distribute the joint output from the investment phase. Crucially the subjects know how much the other individual invested and what her rate of return was. Thus, subjects could distinguish between factors that contributed to the total output within and outside the receiver's or their own control. The authors show that subjects are, on average, sensitive to the source of inequality. They are more likely to choose an unequal distribution if the difference in the output result from differences in investment decisions rather than differences in the return on investment. Almås et al. (2020) use an experiment where an impartial spectator is matched with two workers. The spectator knows whether initial differences in income are due to differences in effort or simply luck. They run this experiment on nationally representative samples in Norway and the US and show that the source of initial inequality matters for redistributive decisions on average in both countries. This experiment was replicated around the world, and they identify a strong luck-merit effect in nearly all countries (Almås et al., 2021). Lefgren et al. (2016) study the luck-merit effect in the context of voting for redistribution. They find a large and significant merit effect on voting behavior and show that subjects are willing to vote for taxes that go against their self-interest to reward effort.<sup>13</sup>

While the above studies all reveal strong evidence that the source of inequality matters, this

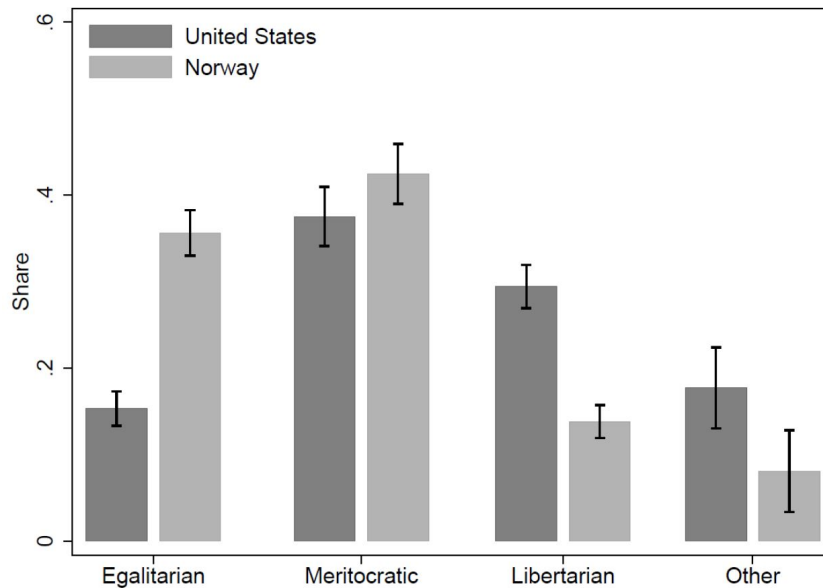
---

<sup>13</sup>Note that differences income may not only result from differences in effort or luck but they may also reflect differences in choice more generally. Several experiments study whether differences in risky choices affect inequality acceptance. Cappelen et al. (2013) find that subjects redistribute on average less, if *ex-post* inequality reflects differences in risky choices rather. Similar results are found in Akbaş et al. (2019). A related paper by Mollerstrom et al. (2015) studies how individuals distinguish between brute- and option-luck (Dworkin, 2002). Their main finding is that individuals can be characterized as “choice compensators” that consider whether an individual

is not automatically the case. For example, recent studies compare redistributive decisions made behind the veil of ignorance and with those made once an individual knows her position. These studies, including Chapter 1 of this thesis, find a luck-merit effect when individuals choose from behind the veil of ignorance but this difference disappears once individuals learn their position in the distribution (Bjerk, 2016; Durante et al., 2014).

As in the empirical literature that studies distributive preferences when the source of inequality is luck, one should ask oneself whether *all* individuals are sensitive to the source of inequality. Recent experiments have asked this question and found considerable heterogeneity in fairness views, categorizing individuals into different fairness types. Cappelen et al. (2007), for example, identify three distinct and prominent fairness types (1) libertarians (accepting of all initial inequality), (2) meritocrats (accepting of inequalities due to differences in effort but not in luck), and (3) egalitarians in a relatively homogeneous group of business students. The study by Almås et al. (2020) identifies the distribution of these fairness views using samples that are representative of the US-American and Norwegian population. As shown in Figure 2, the authors find that each fairness view is represented in both countries. While the study also

Figure 2: Distribution of fairness types in Norway and the United States found in the study by Almås et al. (2020). The figure is taken directly from Almås et al. (2020).



discovers important within-country heterogeneity,<sup>14</sup> it is the cross-country heterogeneity that is, arguably, the most striking. As shown in Figure 2, “meritocratic” is the fairness view that attracts the highest share of individuals in both countries. However, Americans are significantly more likely to be libertarians than Norwegians, and they are significantly less likely to be egalitarian. This difference in the composition of types helps to explain cross-country differences in the demand for redistribution. This is at first sight surprising, given that the previous literature

exposed herself to option luck, even if this did not influence her final income. This is an interesting finding because it indicates that intentions seem more relevant to the distributive decision than outcomes.

<sup>14</sup>In both countries, Women are more likely to reject inequality than men, and conservatives tend to be more accepting of inequality. Socio-economic status is a predictor of inequality acceptance in the US, but not in Norway.

has often highlighted that this difference may rather stem from cross-country heterogeneity in the weight that citizens put on equality vs. efficiency, or beliefs about the source of inequality (both are controlled for by design in this experiment).

## 2.2 What shapes attitudes towards inequality?

The first part of the literature review presented different motives for inequality attitudes and evidence thereof. The evidence shows that people care about the distribution of income and they care about the distribution of income for different reasons: Some are motivated to maximize their payoff, others want to reward people for doing better at a task, while others want to minimize inequality or help the least well-off. The evidence shows that the value function shown in (1) is a relevant characterization of attitudes towards inequality and it lays the necessary foundation for the overarching research question of this thesis.

In this part, I will present evidence of how these motives interact with the economic environment that we face and the experiences that we make. I am, thus, interested in how the elements in (1) are context-dependent and shaped by the economic environment and experiences. The economic environment and experiences affect inequality attitudes by influencing income prospects  $y_i$ , by shaping other-regarding fairness principles  $D^F(\mathbf{y})$ , and by affecting what enters  $D^F(\mathbf{y})$  (for example, beliefs about the source of inequality), holding the fairness principle constant.

I will tackle this topic from three different angles that I see as being particularly relevant for the general motivation of the chapters of my thesis. First, I will show whether fairness ideals  $D^F(\mathbf{y})$  are malleable and influenced by our surrounding and experiences. Second, I will discuss how economic uncertainty shapes attitudes towards inequality through income prospects. Third, I will present how the literature has approached the question of how the economic environment and experiences influence the way individuals learn about the relative importance of luck and effort.<sup>15</sup>

### Malleability of fairness principles

One way that the economic environment and experiences affect attitudes towards inequality is by shaping the relevant fairness principle itself, i.e.  $D^F(\mathbf{y})$ . Thus, this literature asks how fairness principles change, while holding beliefs about other relevant parameters such as variation in own income and beliefs about the relative importance of luck and effort constant.

**Self-serving fairness views** Above, I showed that individuals care about inequalities not only because they seek to profit from a potential reduction in inequality but also for normative reasons because they find some distributions of income fairer than others. One question that has been targeted by the previous literature asks whether there is an interaction between self- and other-regarding attitudes towards inequality. Self-serving fairness views were introduced

---

<sup>15</sup>Needless to say that there are other channels that shape attitudes towards inequality that I do not discuss in the review because they are not directly related to the chapters of my thesis. This includes, in particular, how social identity shapes preferences for redistribution and other-regarding behavior (Luttmer, 2001; Kranton et al., 2020; Shayo, 2009).

to economics by Babcock et al. (1995) in the context of bargaining. It can be rationalized by assuming that individuals seek to morally justify their selfish demand for higher or lower inequalities by exploiting a moral wiggle room (Dana et al., 2007). This occurs when two desired motives (acting fair and maximizing income) conflict, resulting in cognitive dissonance (Festinger, 1957). This tension can be relieved through such self-serving behavior. A self-serving bias in fairness views or inequality attitudes can come from two sources: First, there may be a self-serving bias in fairness principles, meaning that individuals adopt the fairness principle that yields the highest income for themselves. Second, there may be a self-serving bias in beliefs. There, individuals may exploit uncertainty about the efficiency cost from redistribution or the relative importance of luck and effort to justify their opposition or support for inequality.<sup>16</sup>

Self-serving fairness principles assume that individuals change their normative fairness principle if their own income is at stake. It should be clarified that this is not captured by  $\alpha$  in (1), which characterizes the trade-off between maximizing own income and paying a moral cost if the implemented income distribution is not the normatively preferred income distribution. The idea behind self-serving fairness principles is that  $D^F(\mathbf{y})$  itself may change. For example, an individual may hold libertarian fairness principles if she is at the top of the initial income distribution but adopt egalitarian fairness principles if she is at the bottom of the income distribution. Theoretically and empirically, this has been studied by Konow (2000) that I introduced above. Konow finds evidence for self-serving fairness views as he shows that benevolent dictators that have previously engaged into a distribution decision that affected their own income are more accepting of inequalities due to luck.<sup>17</sup> However, the literature that followed did not find strong evidence of such a self-serving bias. Cappelen et al. (2007) exploit the richness of their design to back out  $\alpha$  and then compare whether subjects stick to their fairness principle in situations where a self-serving bias would predict a different distribution decision compared to the one predicted by the identified fairness principle (holding  $\alpha$  constant). They do not find evidence that this is indeed the case. This result was replicated in the study by Cappelen et al. (2013) on fairness and risk-taking. There, the authors compare distribution behavior by spectators and stakeholders. While stakeholders' distribution decisions are biased to make themselves better off (captured through  $\alpha$ ), the distribution of fairness principles remains stable. The evidence from the recent empirical literature that uses more involved designs, generating richer data, thus, shows that fairness principles are remarkably stable and relatively unlikely to be subject to this type of self-serving bias.<sup>18</sup>

An open empirical question is whether self-serving fairness views are self-serving to one's overall life situation as pointed out by Cappelen et al. (2020b). There is correlational evidence that is in favor of this hypothesis. Almås et al. (2017) show that children with a high socioeconomic background (henceforth, SES) are more accepting of inequality compared to those

---

<sup>16</sup>I will discuss the latter in more detail towards the end of this section.

<sup>17</sup>A very recent experiment by Amasino et al. (2021) replicates Konow (2000). They also study the supply-side mechanism driving this result and find evidence for the hypothesis that the self-serving bias is driven by subjects paying less attention to information about the true degree of merit.

<sup>18</sup>This does not mean that self-serving biases in moral judgments are never relevant. As I will show below, luck-effort beliefs are much more malleable to such self-serving biases and recent work by Barron et al. (2020) provide experimental evidence that is consistent with the notion that individuals act in line with a moral motive (truth-telling or distributive fairness) that yields a higher income.

with low SES and Fisman et al. (2020) show that individuals who experienced an increase in their household income became more selfish.<sup>19</sup> Identifying a causal relationship is, however, difficult because one would need exogenous variation in background, aspirations, and one's current situation. Another open empirical question is how stable such fairness views are across different contexts and whether self-serving behavior is relevant on that dimension. One could, for example, assume that individuals hold different fairness views in the context of a workplace or firm than they do at the societal level.

**Are views on fairness shaped by experiences and institutions?** Fairness principles may also be shaped through the institutional environment and the experiences we make. Again, I will focus here on how  $D^F(\mathbf{y})$  is shaped through institutions and will focus on luck-effort beliefs below.

One strand of the literature studies how education affects fairness views. Jakiela (2015), for example, studies social preferences in rural villages in rural Kenya and at a US University. Her experiment uses a dictator game and varies whether initial income is earned or random. She finds no effect of earned income in Kenya but a strong effect in the US. Interestingly, she finds heterogeneity within the Kenyan sample: Those that lived in communities with access to the main road and a higher level of education were more likely to reward effort. This provides suggestive evidence that exposure to markets and exposure to (Western-style) education may lead to a higher prevalence of meritocratic fairness ideals. Cappelen et al. (2020a) provide causal evidence showing that education shapes fairness views, exploiting random variation from an RCT on early childhood education. They show that attending preschool makes children more egalitarian, while those who were schooled at home, put a higher importance on efficiency. It does, however, not affect their degree of selfishness. Kosse et al. (2020) show that pro-sociality can be shaped by early-childhood interventions. They conduct an RCT, randomizing children into a mentoring program. They show that the program has a large effect on children's pro-sociality—which is captured by  $\alpha$  in (1)—and even closes an initial gap in pro-sociality between children from families with low and high SES. This effect is mediated through having a pro-social role model and social interactions. These studies show that inequality attitudes, in the broader sense, are shaped by education and with whom we interact as a child.

One other question is how personal experiences shape fairness principles. Bauer et al. (2014) study how experiencing conflict shapes egalitarian behavior in Georgia and Sierra Leone. They show that enduring war as a child or young adult makes individuals more egalitarian towards in-group members but not towards out-group members. Fisman et al. (2015a) study how distributive preferences changed in the Great Recession in a setting where the source of income is luck. They find that the recession made subjects more selfish and more concerned about efficiency. Cappelen et al. (2021) focus on how the COVID-19 pandemic shapes fairness views using a priming experiment. They find that the pandemic made Americans more accepting of inequalities due to luck and also more willing to prioritize societal over their own problems. Barr et al. (2016) study whether an unemployment spell affects fairness views. She compares

---

<sup>19</sup>It should be highlighted that the general message of this paper is that other-regarding preferences are remarkable stable within individual and across time.

distributive behavior in the laboratory using a longitudinal design. Exploiting within-subject variation in unemployment spells, she finds that individuals become more egalitarian (less likely to reward differences in effort) after having experienced an unemployment spell. This finding is robust to controlling for meritocratic beliefs and suggests that fairness views changed.

Controlled evidence from the laboratory that demonstrates how experiences shape attitudes towards inequality is scarce. Cassar and Klein (2019) show that experiencing success makes us more accepting of inequality compared to experiencing failure. In their experiment, subjects get a prize and they are then matched with two *other* individuals where one subject won the prize and the other didn't win the prize. They then make a distribution decision. The authors find that those who won a prize are more likely to reward the subject that also won the prize, even if the initial allocation was completely random. They show that this result can be attributed to subjects feeling closer to the fellow loser or winner. More research on the causal effect of experiences on fairness views is, however, still needed.

What does this literature tell us about the way that the economic environment and experiences shape  $D^F(\mathbf{y})$ ? On the one hand, the literature cited above illustrates that fairness principles are remarkably stable to experimental manipulations. On the other hand, fairness principles are to some extent malleable to experiencing macro-economic crises, unemployment, educational programs and they correlate with family background. Putting both together, one can conclude that the source of the heterogeneity may be found in the exposure to cultures (Luttmer and Singhal, 2011), ideas, and fairness views during the childhood or one's impressionable years (Krosnick and Alwin, 1989) and through big life events that may change one's view of the world. One intriguing question for future research is to what extent redistributive institutions affect fairness views, for example by affecting norms. This could help to better understand why fairness views are so heterogeneous across countries.

### **Attitudes towards inequality and uncertainty about income rank**

Individuals face more or less uncertainty about their future position in the income distribution. This variation may shape inequality attitudes through the self-regarding channel by shaping income prospects. Studying how uncertainty affects self-regarding attitudes towards inequality has shown to drive demand for more, as well as less redistribution. The direction of the effect is likely to be driven by expectations about where one ends up in the distribution.

Uncertainty about one's future position may imply that one ends up at the bottom of the income distribution. To insure themselves against possible negative income shocks, individuals may reject inequality and demand redistribution. This idea, originally developed by Harsanyi (1955), arises due to the curvature of the utility function and creates a demand for redistribution as it functions as insurance if one does not make it to the upper half of the income distribution. This becomes particularly relevant if one does not know yet where one stands in the income distribution. It is also an argument for Rawls (1971)'s difference principle that is chosen *before* the position in the income distribution is revealed. Empirically, there is a relationship between risk-aversion and preferences for redistribution. Gärtner et al. (2017) study a representative



sample of Sweden, establishing a significant and robust relationship on the population level. Several other studies tried to quantify to what extent risk-aversion can explain an aversion towards inequality by comparing decisions made in a certain environment with those in an uncertain environment or behind the veil of ignorance. The general finding of this literature is that the insurance motive matters, but it is not sufficient to fully explain any opposition to inequality (e.g. Johansson-Stenman et al., 2002; Carlsson et al., 2005; Schildberg-Hörisch, 2010). The first chapter of this paper adds to this body of evidence. There, we show that insurance motives matter, but they are not capable of fully explaining all opposition to inequality, thus replicating the evidence cited above.

While individuals may reject inequalities because they want to insure themselves against negative income shocks, the literature has also proposed that individuals reject inequality for an, arguably, contrasting motive: an aspiration to benefit from high incomes at the top of the distribution. This mechanism was first introduced by Hirschman and Rothschild (1973) as the so-called tunnel-effect. Bénabou and Ok (2001) later developed the related prospect of upwards mobility (POUM) hypothesis. It builds on the idea that high inequality may be seen as beneficial down the road because they signal one's own income in the future. Thus, beliefs about upwards mobility are key for inequality acceptance in this setting. Experimental evidence of this channel is provided by Checchi and Filippin (2004) and Agranov and Palfrey (2020). In the field, evidence that is consistent with the POUM hypothesis has been found to be particularly relevant in transitioning economies (Senik, 2004; Grosfeld and Senik, 2010) and in firms, where managerial salaries may provide information about ones own future prospects (Clark et al., 2009; Cullen and Perez-Truglia, 2018; Godechot and Senik, 2015). Alesina and La Ferrara (2005) analyze survey data and show that individuals with higher *expected* income growth are more likely to oppose redistribution. Since POUM depends on individual beliefs about upward mobility, it is closely connected to beliefs about one's relative ability. Buser et al. (2020) show that confidence in relative ability is an important predictor of redistributive preferences and the gender gap in *over*-confidence can furthermore explain part of the gender gap in the demand for redistribution. In this thesis, two of my chapters will discuss these ideas. Chapter 1 studies how income prospects can erase any opposition towards inequality and Chapter 3 studies how meritocratic beliefs that are related to mobility, are malleable to the economic incentives that individuals face. Next, I will discuss meritocratic beliefs in greater lengths.

### **Forming beliefs about the importance of luck and effort**

One important determinant of whether individuals think a prevailing distribution of income is fair or unfair is the belief about the source of inequality. Especially for meritocratic individuals, the evaluation of an income distribution depends on the source of inequality. Nonetheless, the source of inequality also matters for self-regarding motives for redistribution because they shape the income prospects discussed above.<sup>20</sup>

---

<sup>20</sup>Individuals who believe that they are more talented and hardworking than others should have higher income prospects if they believe that these things are important for determining one's income rank compared to luck (see Buser et al., 2020, for an analysis of overconfidence in one's ability and inequality acceptance).

The true importance of effort relative to luck is unobserved and uncertain. Uncertainty about the source of inequality implies that we hold beliefs about the true state of the world, and these beliefs determine how  $\mathbf{y}$  can be split into  $\mathbf{w}$  and  $\mathbf{x}$  (see (9)). This leads to the general question of how we form and what influences these beliefs.

**Economic mobility and luck-effort beliefs** One empirically and practically very relevant question is to ask what type of information influences luck-effort beliefs. Arguably, the most common answer to this question is economic mobility. Mobility can be seen as a measure for equality of opportunity because it characterizes the extent to which one's parents' SES determines one's future (Alesina et al., 2001). As this is a factor outside individual control, low levels of economic mobility may signal that socio-economic background is more important to get ahead than hard work. For example, if the likelihood that an average individual that hails from a household from any quintile in the income distribution has a 20% chance of making it to the top quintile, then the society becomes completely mobile as everybody has an equal chance of making it to the top. However, if individuals that hail from the bottom quintile only have a 1% chance to make it to the top quintile, while those from the top quintile have a 90% chance of staying there, mobility is low because the correlation between socio-economic origin and outcomes becomes much higher.

Beliefs about mobility are relevant for inequality attitudes in two ways: First, through the self-regarding POUM channel outlined above and second, through the other-regarding channel discussed in this part of the survey. The other-regarding channel argues that the level of mobility mirrors the degree of inequality in opportunity, which is a key criterion such that differences in outcomes reflect differences in individual choices rather than heritage. Through that, it has an interesting implication for policy-makers. If citizens are more sensitive about equality of opportunity rather than equality of outcomes, they would also demand policies that promote economic mobility.<sup>21</sup>

The theoretical foundation on the effect of mobility on luck-effort beliefs has been first laid out by Piketty (1995). Piketty argues that individuals (or dynasties) learn about luck and effort through mobility experiences. Since learning is costly, different individuals may end up converging to different beliefs about luck and effort, even if they started at the same point within the same economic system. This dynamic arises because two individuals may start off exerting the same amount of effort, but one of the two experienced a positive shock while the other experienced a negative shock. The individual who experienced the negative shock may now revise her beliefs downwards and exert less effort, thereby decreasing the chance of experiencing upwards mobility, while the opposite is true for the other individual that was initially lucky. This diverging dynamic leads to multiple equilibria where some individuals believe that effort is important and demand low levels of redistribution, while the opposite is true for others that experienced low levels of mobility.

In recent years, the empirical literature on economic mobility and luck effort beliefs has considerably developed by making causal assessments between mobility beliefs and luck-effort

---

<sup>21</sup>Indeed, the political discourse in recent years has shifted in this direction and has been attributed to a meritocratic paradigm shift in recent decades (Sandel, 2020).

beliefs, as well as preferences for redistribution. Gathering causal evidence on the relationship is important because the link between luck-effort beliefs and mobility is less obvious than it may seem at first sight. First of all, individuals may have very strong priors on luck-effort beliefs due to their cultural background or because they see an affective value in believing that effort is important for getting ahead (Lerner, 1980) or even the opposite. Second, even if there is total mobility, this does not mean automatically that effort matters for getting ahead. One would, for example, have total mobility in a world where (a) luck is equally distributed across society and luck dominates effort in the income-generating process and (b) everybody works equally hard and luck is equally distributed; then, being lucky becomes the tiebreaker.<sup>22</sup>

First correlational evidence of the relationship between mobility experiences and luck-effort beliefs was documented by Alesina et al. (2001), by showing that occupational mobility experience decreases the demand for redistribution. Giuliano and Spilimbergo (2014), furthermore, show that experiencing a recession shapes individuals' luck-effort beliefs and, thus, preferences for redistribution. This can be viewed as evidence in favor of Piketty's proposed mechanism. One of the more explicit tests of Piketty (1995) is from Gärtner et al. (2019). They use a survey experiment where they confront Swedes with their own mobility experiences. They first show that Swedes generally underestimate the degree of mobility within their lifetime. Hence, mobility beliefs are not necessarily correct. Furthermore, they identify significant correlations between mobility perceptions and luck-effort beliefs, as well as with preferences for redistribution. The second result exploits the experimental component of their survey: In accordance with Piketty (1995), individuals are more likely to believe in the importance of effort to get ahead, upon learning that mobility was higher than they initially thought. Thus, the paper constructs a causal link between learning about one's own mobility and the luck-effort beliefs that individuals hold. The exogenous change in beliefs translated into a change in redistributive preferences among right-wing individuals who previously underestimate their upward mobility.

Alesina et al. (2018) focus on inter-generational mobility, rather than on own mobility experiences.<sup>23</sup> Using a survey experiment in France, Italy, Sweden, the UK, and the US, they first document that Europeans underestimate levels of inter-generational mobility, while US-Americans overestimate mobility. This is striking, as they argue that actual mobility is not that different across the continents. They then exogenously vary respondents' beliefs about intergenerational mobility by showing treated subjects pessimistic information about mobility. This information leads to a decrease in the belief that effort is important to get ahead and an increase in support for redistributive policies that promote equality of opportunity.<sup>24</sup>

---

<sup>22</sup>The latter point becomes particularly relevant in winner-takes-all markets (Frank, 2017).

<sup>23</sup>Piketty (1995)'s focus is on mobility *experiences* on the personal or family level, rather than learning about mobility on a societal level. However, one can apply an inter-generational or aggregate reading of the model. In that case, the predictions would be similar in the sense that learning that mobility is high, is a signal that effort is more likely to be rewarded.

<sup>24</sup>The latter effect is only present for left-wing respondents. They argue that this heterogeneity in the treatment effect may be due to a more general mistrust of the government by right-wing individuals as noted in the previous literature (e.g. Kuziemko et al., 2015). This shows that this must not necessarily translate into a willingness to reduce inequality through policy, as this willingness may depend on other factors that are not directly related to luck-effort beliefs.

**Motivated belief distortion of luck-effort beliefs** One of the contributions by Alesina et al. (2018) and Gärtner et al. (2019) is to show that individuals tend to misperceive economic mobility and that different individuals may have different perceptions of reality. This raises the question of the mechanism that leads to this heterogeneity and misperception in beliefs. While such a misperception may occur under the model developed by Piketty (1995), there are also behavioral theories of motivated belief distortion that can rationalize these empirical findings.

Recent years have produced a stream of papers that study a self-serving attribution bias in luck-effort beliefs. Self-serving behavior, as discussed above, can be applied to luck-effort belief formation in an uncertain environment: To maintain a positive self-image, individuals may deliberately want to attribute failure to things outside of their control (luck) and success to things within their control (effort) (Frank, 2017). This mechanism leads to polarization in luck-effort beliefs, where those who are successful in life end up overemphasizing the importance of the effort to get ahead, ignoring the good fortunes that may have brought them there in the first place; those who struggle in life overemphasizing the importance of luck relative to effort, possibly ignoring that they could have done more in the past. This type of self-serving bias can then also lead to polarization in distributive preferences, as it shapes what individuals regard as fair and unfair.

Recent contributions from the experimental literature provided results that are in line with the aforementioned mechanism. Deffains et al. (2016) show that (exogenously) doing relatively badly in a task is attributed to factors outside of one's control, while the opposite is true for those who did relatively well. This translates into higher and lower demand for redistribution respectively and it polarizes attitudes towards redistribution. Cassar and Klein (2019) find a similar effect using a dictator game to show that this attribution bias affects inequality acceptance. A very recent contribution by Fehr and Vollmann (2020) documents a similar effect, while additionally showing that this effect is not heterogeneous in political affiliation, indicating that this bias is relatively ingrained in the human psyche. Di Tella et al. (2007) presents evidence from the field that is in line with a self-serving attribution bias. They show that land-squatters are more likely to believe that effort is important to get ahead if they were quasi-randomly granted land titles. Finally, one may ask whether this self-serving bias is driven by a desire to maintain a high level of self-confidence or whether individuals want to exploit a moral wiggle room that permits them to morally justify paying lower taxes. Valero (2020) tests the latter hypothesis using a laboratory experiment and find no evidence in favor of it.

While the literature on the attribution bias typically argues that individuals distort beliefs about luck and effort because they want to maintain a positive self-image, one can also assume that individuals distort luck-effort beliefs because they derive a value from believing that effort or luck is important for getting ahead. The most influential theory in economics that incorporates this idea was developed by Bénabou and Tirole (2006). The authors argue that individuals form motivated luck-effort beliefs because they may want to believe that effort is important to get ahead. This demand to maintain the belief that effort is important can either be affective or instrumental.

The affective channel was first proposed by the psychological literature and argues that

people have a desire to believe that effort is rewarded and that everyone gets their just desert. The argument goes that believing in the importance of effort to get ahead (just world beliefs) is anxiety-reducing because individuals have a desire to believe that they can control their life (Lerner, 1980). This is an affective motive for belief distortion, where individuals receive *direct* utility from believing that their effort is rewarded. An individual may now deliberately choose to ignore information that indicates that effort is *not* rewarded to maintain this belief in a just world. This is utility-maximizing as long as the gain from maintaining the belief is higher than the cost from belief distortion.

The instrumental channel assumes that the value from beliefs in a just world is motivational. Bénabou and Tirole (2006) argue that present-biased agents distort beliefs about luck and effort to overcome a misallocation of effort due to their lack of motivation. This type of belief distortion occurs if the gain from belief distortion (overcoming lack of willpower by believing that effort is more important than it actually is) is lower than its cost (potentially exerting too much effort). This trade-off depends on the incentives that individuals face. This implies that the demand for motivating belief distortion is endogenous to the incentive structure that agents expect to face—a topic to which I will get back below. The empirical literature on this type of belief distortion is scarce because the empirical literature in motivated beliefs has mainly focused on ego-relevant beliefs. Chapter 3 marks one of the first contributions that tests whether luck-effort beliefs are distorted for motivating purposes.

**Are luck-effort beliefs endogenous to redistributive policies?** One of the main open empirical questions in this literature is to ask whether these beliefs are endogenous to economic institutions. One question that has received attention in the prior literature is whether luck-effort beliefs are shaped by the redistributive policies we face. Theoretical contributions in political economy show that luck-effort beliefs may indeed be shaped by the levels of redistribution we face and this helps to understand the cross-country variation in luck-effort beliefs and levels of redistribution.

Alesina and Angeletos (2005) capture this idea by arguing that levels of redistribution affect individuals' ability to disentangle success through effort and success through luck. Their argument states that under high levels of redistribution, individuals work and invest less and it, thus, becomes harder to disentangle success through hard work from success through luck. This leads to pessimistic beliefs about the importance of effort and persistent demand for redistribution. The opposite is true if initial levels of redistribution are low. Now, the income distribution correlates much more with effort, sustaining a low demand for redistribution. The authors conclude their article by pointing to the potential roots in the trans-Atlantic differences in luck-effort beliefs and fairness. Similar to Piketty (1995), they identify historical experiences as the principal driver: In Europe, the modern welfare state evolved from a feudal society, where circumstances outside of one's control, such as one's heritage plays an important role in one's fate. The US, argue on the other hand, the authors, evolved from a relatively egalitarian society (within the white, male, and non-enslaved population), where white settlers were less restrained by their heritage in rising through the ranks of society.

Bénabou and Tirole (2006) use a very different type of mechanism that shows how levels of redistribution may affect luck effort beliefs. As I noted above, the authors argue that individuals want to believe that effort is important for getting ahead if the benefits from belief distortion outweigh the cost from belief distortion. In other words, there must be a demand for motivating belief distortion. This is the case if the marginal gain from effort is sufficiently high, which is more likely to be the case for low levels of redistribution. Thus, the gain from belief distortion increases, as levels of redistribution decrease. This could then lead to a situation, where society has low levels of redistribution and thus engages in motivating belief distortion which leads them to over-estimate the importance of effort in getting ahead. For high levels of redistribution, the incentives to engage in motivating belief distortion are lower and, thus, individuals will be more pessimistic about the importance of effort to get ahead.<sup>25</sup>

The empirical literature on this topic is relatively scarce and should be addressed by future research. This is partly borne out of an identification problem that arises if we assume on the one hand that luck-effort beliefs are a function of redistribution and post-tax inequality and on the other hand predictive for the demand for inequality reduction. Nonetheless, the economics literature has made some progress on this front in recent years. Gärtner et al. (2019) find suggestive evidence for an inter-generational reading of Bénabou and Tirole (2006) by showing that parents teach their kids over-optimistic beliefs about luck and effort. This is particularly the case if parents expect low levels of redistribution in the future, which is a key prediction of the theoretical model. Chapter 3 of this thesis complements the above-mentioned paper by testing another key prediction of the Bénabou and Tirole (2006) model—that luck-effort beliefs are endogenous to the incentives that individuals expect to face. Roth and Wohlfart (2018) show that experiencing high levels of inequality during one’s impressionable years leads individuals to be more convinced that effort is important to get ahead. This could be rationalized through the framework developed by Bénabou and Tirole (2006) as the level of inequality affects the demand for motivating belief distortion by shaping incentives.

What do we learn from the literature summarized above? As shown, luck-effort beliefs are not fixed but a variable over which individuals form beliefs. These beliefs are shaped by how we perceive economic mobility and they are distorted for self-serving purposes. The empirical evidence, especially from the laboratory, suggests that they are—at least to some extent—more malleable to the economic environment than the fairness ideal  $D^F(\mathbf{y})$  itself. This makes them particularly interesting to study and is a fruitful avenue for future research.

### **2.3 What are the consequences of attitudes towards inequality?**

The literature summarized above was primarily targeted at characterizing individuals’ attitudes towards inequality and asking how they are shaped by the economic environment. Chapters 1 and 3 are dedicated to contributing to this question. Chapter 2 also enhances our understanding of the consequences of inequality attitudes by asking whether normative distributive preferences

---

<sup>25</sup>This model could, thus, explain why Americans are too optimistic about social mobility (Alesina et al., 2018) but it is unclear how it can be used to explain why Europeans underestimate social mobility in a society.

affect the provision of incentives in organizations. In this section, I will briefly summarize the literature on how fairness concerns and distributive preferences affect the demand for redistribution and labor market outcomes and interactions. I focus on these applications because the chapters of this thesis are motivated by them. Nonetheless, I want to highlight that fairness views and distributive preferences have consequences above and beyond the political economy of redistribution and labor markets. In particular, fairness considerations and inequality attitudes have shown to be crucial for understanding why firms do not always take advantage of their market power in case of excess demand (see, in particular, Kahneman et al., 1986a), empirical regularities seen in bargaining behavior, as well as interpersonal trust and cooperation (e.g. Ashraf et al., 2006; De Bruyn and Bolton, 2008; Hedegaard et al., 2021).

**Demand for redistribution** One (relatively obvious) consequence of inequality attitudes is the demand for redistribution in an economy, as well as the support for parties that support more or less redistribution (see Alesina and Giuliano, 2011, for an extensive review of the literature). It seems sensible to assume that if individuals reject inequality, then they should demand a reduction of these inequalities through policy. In fact, a lot of the studies surveyed above are explicitly motivated by this question (e.g. the theoretical model by Alesina and Angeletos, 2005) and use the demand for redistribution as an outcome variable characterizing opposition against pre-tax levels of inequality (e.g. Durante et al., 2014; Fong, 2001). Tyran and Sausgruber (2006) study how fairness views affect voting for redistribution, showing that taking into account distributive preferences is crucial for understanding voting outcomes in the laboratory. The link between inequality attitudes and preferences for redistribution can be explained through the self-, as well as other-regarding channel.<sup>26</sup>

While empirical studies often find strong evidence that selfish motives are important predictors for the demand for redistribution (see e.g. Cruces et al., 2013; Alesina and Giuliano, 2011), the evidence is less evident for other-regarding motives. On the one hand, the literature reviewed above identified a robust correlation between fairness views and preferences for redistribution. Almås et al. (2020), for example, show that a majority of sampled Norwegians and Americans say that fairness is an important determinant for whom they vote. They also find a strong correlation between fairness views and the demand for redistribution and this correlation was replicated in a not yet published study by Almås et al. (2021) that surveys countries around the world. Epper et al. (2020), furthermore, show how other-regarding preferences map into redistributive policy preferences.

Related studies have linked distributive preferences with political preferences. Equality-minded Americans, for example, were more likely to be affiliated with the Democratic Party and to vote for Obama in 2012 (Fisman et al., 2017). This correlation could be explained through the different redistributive policy approaches. However, recent survey experiments that exploit exogenous variation in relevant parameters such as inequality perceptions (Kuziemko

---

<sup>26</sup>In the self-regarding channel, it is precisely the belief to gain or lose from redistribution that drives their attitudes towards existing (or expected) inequalities. In the other-regarding channel, it is fairness concerns that drive redistributive preferences. For example, inequality-averse individuals are expected to demand more redistribution than a meritocrat who thinks the effort is an important predictor of pre-tax income.

et al., 2015) or mobility beliefs (Alesina et al., 2018) fail to identify that a shift in attitudes towards existing inequality translates into a shift in policy preferences. This could be either due to a more general distrust of the government’s ability to implement these policies or because the exogenous variation in inequality attitudes is not strong enough to also shift policy preferences.<sup>27</sup>

**Inequality attitudes at the workplace** In the empirical literature on attitudes towards inequality, especially in the social preference literature, there is a strong tradition in focusing on inequality at the workplace. Starting with Akerlof and Yellen (1988), Frank (1984) and Kahneman et al. (1986b) economists have asked how fairness concerns may explain the rigidities we observe in labor markets. Furthermore, social preference models often find a direct application in settings where agents interact in small groups; this is particularly relevant at the workplace. Attitudes towards wage inequality at the workplace have far-reaching implications. As I will show below, wage inequality can have adverse effects on labor supply both on the intensive as well as the extensive margin. This can have significant financial consequences, next to the psychological costs that workers may occur through a loss in job satisfaction. Nonetheless, it may also mean that employers or principals seek to avoid large wage differentials among their employees for profit-maximizing purposes. This would imply an overall reduction in pre-tax inequality levels due to wage compression.

Evidence from laboratory experiments can be classified into two categories. One set of papers that typically uses the gift-exchange game (Fehr et al., 1993) to study how fairness concerns affect labor supply. While this game was designed to capture reciprocal motives, rather than a pure concern for inequality, one can argue that workers are concerned about inequalities in rent sharing. In other words, an agent may refuse to exert effort if the contract benefits the principal over-proportionally. The experimental literature, surveyed by Charness and Kuhn (2011), has found that agents care about the fairness of wages, and they refuse to exert effort if they believe that the principal abuses his or her situation by offering unfair wages. The reciprocal motive is in this situation probably stronger than the pure inequality motive, since the principal’s agency is key for workers behavior. Another set of papers using lab experiments studies how horizontal inequality *among* workers affects labor supply of agents. Theoretically, inequality averse agents suppress labor supply if inequalities become too high (Englmaier and Wambach, 2010; Bartling and Von Siemens, 2010) and there is some empirical evidence that this is indeed the case, especially for disadvantageous inequality (e.g. Clark et al., 2010; Bracha et al., 2015; Gächter and Thöni, 2010; Greiner et al., 2011; Ku and Salmon, 2012) but also some evidence that inequality among workers does not affect wages (e.g. Hennig-Schmidt et al., 2010; Bartling and von Siemens, 2011; Charness and Kuhn, 2007; Gächter et al., 2012). The exact reason for why there is this disparity in the experimental literature is still unclear but one can assume that principal’s agency and the justification (for example luck or effort) of wage inequality are key

---

<sup>27</sup>Furthermore, the willingness to reduce societal inequalities through redistribution may depend on beliefs about the efficiency loss through these policies (Okun, 1975). However, Section 2.1 showed that recent studies found weak support for efficiency concerns in representative samples. This seems to translate to settings that are more explicit in assessing the demand for redistributive policies (see e.g. Stantcheva, 2020).



for explaining the adverse effects of inequality on labor supply in the lab (Bracha et al., 2015; Gross et al., 2015).

A relatively large literature in labor economics has simultaneously emerged to study the effects of wage inequality using field experiments. These studies typically find stronger effects of wage inequality on behavior and job satisfaction.<sup>28</sup> Card et al. (2012) uses a survey experiment, where they randomly inform some employees from the University of California about a website where they can look up the income of their co-workers. The authors find that treated workers with below-median income reported lower job and pay satisfaction and a higher propensity to quit the job. This is consistent with a model where workers care about disadvantageous inequality. A recent field experiment by Cullen and Perez-Truglia (2018) randomly informs workers of a large firm in South-East Asia about their co-workers' wages, varying the hierarchy. They find that horizontal wage inequality has adverse effects on intensive-margin labor supply if they learn that their peers earn more, while it leads individuals to work harder if they learn that their manager earns more than they initially thought. This dichotomy between positive effects of inequality on behavior for vertical inequality and negative effects on behavior for horizontal inequality echoes similar results from the well-being literature (Clark et al., 2009; Godechot and Senik, 2015). Cohn et al. (2014) conduct a field experiment where workers are matched in teams of two. If only one of the workers' wages is cut, the reduction in effort is higher compared to when both workers' wages are cut. This evidence is consistent with a model where subjects are averse to disadvantageous inequality. Breza et al. (2018) mark the most comprehensive field experiment that studies the effect of wage differentials. They show that workers' responses to wage inequality depend on whether inequality can be justified through (salient) productivity differences using a field experiment in India. Through their rich experimental setup they can exploit variation in within-team wage inequality, productivity transparency, and team composition. Exploiting this variation, they identify a strong negative effect of inequality on both output and attendance, particularly of workers that suffer from disadvantageous inequality. The result on the extensive margin labor supply (attendance) is particularly striking because the workers are indeed forgoing a significant amount of income. By exploiting exogenous variation in the ability to observe productivity, they furthermore show that these adverse effects are mitigated if wage differentials can be justified through productivity differences. Finally, in their endline survey, they show that pay inequality had adverse effects on social cohesion within the group.

The evidence from these field experiments clearly shows that inequality at the workplace can have adverse effects on labor supply and well-being. One key feature that determines whether these negative effects occur seems to be the justification of wage differentials. This implies that we may observe more within-firm inequality if managers can effectively communicate the reason behind differences in income. Furthermore, these adverse effects are particularly strong for those who earn lower wages.

As the reader may have noticed, most of this literature is about the inequality attitudes of agents or workers. Distributive preferences of managers have barely been the focus of attention

---

<sup>28</sup>This may not be very surprising, given that concerns about fairness may be much stronger at the actual workplace, to which one has an emotional attachment

by this literature. Studying manager’s distributive preferences is Chapter 2’s contribution.

The preceding review of the literature on attitudes towards inequality showed that individuals hold preferences over income distribution. These attitudes have wide-ranging implications for preferences over redistribution and labor market outcomes. As I showed in the first part of the survey, inequality attitudes are driven by different self- and other-regarding motives. Most of the proposed motives have been identified in controlled environments using laboratory or online experiments. While this speaks in favor of the relevance for each motive, it also unveils a striking plurality of inequality attitudes—both within as well as across different experiments. This opens the question of what drives this plurality. As I showed in the second part of the literature review, inequality attitudes are shaped by the economic environment and the experiences we make—either by changing the income prospects we face, by affecting fairness views, or by changing beliefs about the source of inequality. The chapters of this thesis advance this literature (i) by showing how the importance of difference motives for inequality attitudes interact with the choice environment, (ii) by unveiling the importance of distributive preferences for incentivization decisions, and (iii) by assessing whether beliefs about the source of inequality are shaped by the incentives we face.

### 3 Empirical methods used in this thesis

This thesis relies on *laboratory experiments* to improve our understanding of attitudes towards inequality. Experimental economics is an empirical method that studies behavior in a *controlled* environment. Experiments as an empirical method gained in popularity during the 1980s and have established themselves as a relatively small but influential field within economics. Laboratory experiments are also increasingly used as a complement to, or in combination with, other empirical methods (Roth, 2015). The main difference of laboratory experiments relative to other empirical methods is that it allows the experimenter to exert complete control over the data generating process (Jacquemet and L’Haridon, 2018). This facilitates the identification of causal relationships between the explanatory variables and outcomes of interest. It also permits a more direct mapping from choices to preferences.

Attitudes towards inequality have a long tradition of being studied in the laboratory. This is partly because experiments were influential in breaking the self-regarding paradigm in economics (see Güth et al., 1982; Kahneman et al., 1986b, for early examples) and fed directly into the development of models of other-regarding preferences. Indeed, seminal models of social preferences either use existing experimental data to motivate and test their model (see e.g. Fehr and Schmidt, 1999) or they provide experimental evidence along with their theory to either calibrate the parameters or test its predictions (see e.g. Charness and Rabin, 2002).

Furthermore, as discussed in the literature review, several motives behind attitudes towards inequality have been put forward. These motivations are often simultaneously present and relevant in the field. For example, non-experimental survey data may show that individuals with high incomes oppose a reduction in inequality. This could be because they believe that inequal-

ities are due to differences in hard work, due to their unwillingness to pay for a reduction of inequality through redistribution, or due to a genuine preference for maintaining high inequalities. The laboratory allows us to create a unique setting that enables us to disentangle these motivations. One popular way of doing this is by asking subjects to choose between different income distributions while varying the set of possible income distributions ( $\Omega$ ) to disentangle the different motives. This is the approach I use in Chapters 1 and 2, where I create situations where experimental subjects have to make consecutive choices between different income distributions or piece-rate schemes. The choices are designed such that some motives that may be relevant for the acceptance or rejection of inequality are muted in some decisions but become relevant in other decisions. Thereby, I can disentangle selfish concerns of income inequality from other-regarding concerns and ultimately identify preferences over income distributions.

Laboratory experiments are, furthermore, an effective method for studying how the economic environment shapes inequality attitudes. Identification is achieved by observing how behavior changes after varying one specific aspect of the experimental setting while holding the other characteristics constant. In all chapters, I employ this approach by varying crucial aspects of the environment. In Chapter 3, for example, I am interested in assessing how incentives shape luck-effort beliefs. To identify the relationship between incentives and beliefs, I create counterfactual situations that are identical, *except* for the fact that in one setting a decision-maker expects to face incentives while this is not the case in the other situation. Having such a clear identification is key for understanding how attitudes towards inequality may change from one situation to another.

What do I aim to achieve through the use of experiments? Following the—now classic—definition by Roth (1988), experiments can be used to “*speak to theorists*”, “*search for facts*”, and “*whisper in the ears of princes*.” The first approach uses experiments to test the predictive power of theory by creating an environment that resembles the context characterized in the theory and then compares observed behavior with the theoretical prediction. The second approach uses experiments to explore empirical regularities that generate new knowledge which may then build the basis of new theories. It can also be seen as the *scientific view* (Camerer, 2015), as the results from laboratory experiments help us to better understand behavior outside the laboratory. The third approach uses experimental evidence as a foundation for advice, for example on policy questions. The chapters of my thesis fall for the most part in the second category: The goal of these chapters is to unveil behavioral regularities that can then be used to inform more theory or better understand behavior in the field. Chapter 2, for example, shows that distributive preferences are important determinants for the use of incentives. Thus, future models of the labor market should account for the importance of other-regarding preferences on the choice of an incentive scheme. Chapter 3 also speaks to the first purpose by testing a mechanism that was previously proposed by theory.

What are the limits of the empirical methods that I use in my thesis? One of the main limitations of “searching for facts” experiments is their external validity. One should, thus, ask oneself whether the results and insights of my experiments replicate in other contexts. Given the inherent loss of characterizing all the details from the real world that comes with the necessary

abstraction of a laboratory experiment, external validity is impossible to be fully tested empirically (Jacquemet and L’Haridon, 2018). Nonetheless, it merits a discussion. One concern is that the treatment effects that I identify in the laboratory do not translate to real-world settings. This may be because there is heterogeneity in the treatment effect across different populations. While this can never be completely ruled out, there is now considerable evidence that the sign of the correlations found with standard subject pools in the laboratory replicate in more diverse subject pools (Snowberg and Yariv, 2021). Furthermore, I want to highlight that some of the effects documented in my thesis have previously been identified in the non-experimental empirical literature but lacked a causal interpretation. In this sense, my experiments serve as a powerful complement that should be used as a proof of concept nourishing the suggestive findings made in the previous literature. One other dimension in how external validity matters is not through differences in correlations but differences in preferences, i.e. levels. This is, arguably, harder to fulfill and the insights of my second chapter are particularly susceptible to this problem. Managers studied in the lab may be very different from those in firms, due to the self-selection of individuals into positions and occupations. In the concluding section of Chapter 2, we will therefore discuss this limitation in more detail.

## 4 Summary of each chapter

The following pages present a detailed summary of each chapter.

### **Preferences over income distribution: Evidence from a choice experiment**

The first chapter of the thesis is joint work together with Sophie Cêtre, Claudia Senik, and Thierry Verdier. This chapter contributes to answering the broader research question by assessing how attitudes towards inequality depend on the context of the choice. In the chapter, we focus on three aspects that have never been combined in one experiment studying inequality within small groups: (i) the Pareto-dominance criterion, i.e. whether an income distribution allows everyone to be weakly better off compared to the other distribution, (ii) whether choices are made behind the veil of ignorance or with the position known, and (iii) whether relative payoffs are based on merit or luck. We use a choice experiment framed as a series of choices between two projects that lead to different “bonus” distributions. More precisely, our design asks subjects to make a series of incentivized binary choices between two payoff distributions for a group of five individuals (the subject and four additional anonymous participants in the lab). Between subjects, we vary the origin of people’s position within the distribution (either based on luck or a real effort task). Within subjects, we vary whether one distribution is Pareto-dominant as compared to the other or not. We also ask subjects to choose successively behind the veil of ignorance, hence not knowing their future rank and payoff, and then with information about their position within their group. The series of binary choices that subjects have to make can be split into two categories. In the first category of choices, the total payoff is the same in the two proposed projects, but one distribution is more unequal and has higher top incomes and lower

bottom incomes. In the second category of choices, the more unequal project Pareto-dominates the more equal one, i.e. it makes all of the group members weakly better off in absolute terms. Finally, we randomly assign subjects into a Merit and a Luck treatment. In the Merit treatment, people's position within their group of 5 is determined by their relative performance in an effort task to be performed after the choices are made behind the veil of ignorance. In the Luck treatment, the ranking is determined randomly.

Our main finding is that, behind the veil of ignorance, subjects unanimously prefer the higher inequality project when it is Pareto-dominant. In this case, it does not make any difference whether subjects belong to the Luck treatment or the Merit treatment. Unanimity only breaks once positions within the income distributions are fixed, i.e. when subjects know their own ranking before they choose. In that setting, about 75% of subjects prefer the Pareto-dominant distribution over a more compressed payoff distribution. The other 25% engage in money burning. They burn money at the top by choosing the low inequality project even if it does not improve the lot of the low earners. Furthermore, when subjects choose between two distributions that have the same efficiency (same total payoff), about 65% of them prefer the low-inequality distribution. When choosing behind the veil of ignorance, subjects are significantly more likely to embrace the high inequality distribution if they are in the Merit rather than the Luck treatment. This significant treatment effect disappears as soon as subjects learn about their rank, whereupon 70% of subjects prefer lower inequality when their own payoff is not affected. All subjects who are better off in the *low* inequality distribution choose the latter, but only 80% of subjects who would be better off in the *high* inequality distribution choose the latter. Hence, 20% of individuals are strongly inequality averse and act accordingly, even when this comes at a personal cost.

Thus, our findings illustrate in a controlled environment how dependent attitude towards inequality are on the economic environment. This is most strikingly illustrated through our main finding: Inequalities may be contested in an environment, where individuals have no prospects of gaining from inequalities even if they are Pareto-optimal but they are embraced once individuals have the possibility to benefit personally from these inequalities.

### **Principals' distributive preferences and the incentivization of agents**

In the second chapter, which is joint work with Sophie Cêtre, we ask whether principals' or managers' distributive preferences affect the allocation of incentives within firms. The starting point of our analysis is the conjecture that distributive preferences of managers may interfere with the implementation of incentive schemes that a manager may view as profit or output maximizing. For example, an egalitarian may be reluctant to implement high-powered tournament incentives, since these imply very large inequalities.

In our analysis, we first provide evidence of a robust correlation between the distributive preferences of executive managers and the incentive structures of their firms. We use a French survey of 4,000 employers and executive managers that includes an extensive set of questions related to workers' wage compensations. We show that when managers think that a policy of individualized wages may be unfair, they are less likely to implement performance pay. Of

course, reverse causality can explain the result and the correlation could also be driven by strategic concerns, instead of purely normative preferences. Workers may exert less effort in excessively competitive environments, and this can be anticipated by managers. We show that the relationship declines in strength but remains sizable and statistically significant when we include strategic motives for using or avoiding performance pay, such as the prevalence of unions, whether they think performance pay motivates workers, whether it is likely to create tensions, etc. This correlation is also robust to a wide array of manager and firm specific controls.

We then complement the evidence from the survey with a laboratory experiment. The goal of the experiment is to create a controlled environment that allows us to make a more granular assessment of the trade-off that principals face and avoid potential reverse-causality confounds that may arise due to managers self-selecting into firms based on their distributive preferences. To that end, we run a principal-agent experiment, randomizing subjects into manager (principal) or worker (agent) positions. Each principal is matched with two workers of differing ability levels. Both workers choose a costly effort level to produce output, and effort is non-contractible. Principals choose between a series of binary piece rate wage contracts for both workers. These piece rates generate a variable pay-for-performance share of labor income. We randomly allocate principals to either a Stakeholder group (principals' income is proportional to the output produced by the workers), or a Spectator group (fixed income). The Spectator group makes the moral hazard situation irrelevant since the principal no longer has an incentive to maximize output. Thus, Spectators can implement their preferred income distribution at no cost, which gives us a measure of the distribution of income principals believe is fair. In the Stakeholder group, principals must take into account workers' incentives if they want to increase joint output and maximize their own income. This gives us a measure of principals' willingness to pay for implementing their preferred distribution. The difference in behavior between these two groups isolates normative distributive preferences at the extensive margin. The comparison across treatment groups also characterizes the possible effects of institutional factors such as competitive pressure through market forces on the importance of distributional concerns in incentivization decisions.

Moreover, our framework allows us to pin down the relative importance of various fairness ideals (egalitarian, output maximizing, and equal-procedure) among principals. Piece rate wage contracts are an innovation compared to the existing literature because comparing the piece rates chosen for each worker, depending on their ability level, leads to direct classification into three distributive preferences types. Choosing to reward the high ability worker with a higher piece rate is evidence of being output oriented, since in our setting this approach is output-maximizing if workers best respond to wage contracts. Rewarding both workers with the same piece rate implies to paying them in proportion to the output they have produced. This leads to procedural fairness since both workers are treated equally with the same piece rate. Finally, giving the low ability worker a higher piece rate shows an egalitarian concern, since differences in productivity will be offset. We calibrate these egalitarian contracts in such a manner that if both workers exert the same level of effort, then they are paid the same final total wage. This corresponds to a common situation in real firms, in which both workers are paid the same final

wage, despite their different production levels.

We find that despite the firm-like framing and the moral hazard situation, principals do hold egalitarian concerns. On average, they are willing to accept a trade-off between higher output and reduced within-firm inequality. This willingness is significantly lower if principals are Stakeholders (extensive margin incentives) and it is also the case within treatments when there is a large trade-off between maximizing output and equality. Stakeholders are also more sensitive to these intensive margin incentives than Spectators. When the alternative to the output-maximizing (high-inequality) contract is the equal piece rate contract (rather than the egalitarian contract), principals are not more likely to choose it on average. This indicates that subjects are not more willing to sacrifice output to implement equality in outcomes compared to equality in procedure. This indicates that equality in procedure as such is not seen as a particularly attractive contract characteristic and principals are more interested in distributive outcomes. Nonetheless, subjects are significantly more likely to opt for a contract that permits equality in procedure if it is posited *directly* against a contract that provides equality in outcomes. We then calibrate a simple utility function that takes principals' other-regarding concerns into account. The estimates for the representative principal suggest that (i) intrinsic motives are 30% as strong as extrinsic motives in maximizing output and (ii) that principals are averse to extreme inequalities. We then use a finite mixture model to characterize heterogeneity in preferences. We quantify the importance principals attach to the payoff of high- and low-ability agents, allowing for the variation in these importances according to whether one agent is paid a higher or lower piece rate than the other agent. The Normalized Entropy Criterion (see McLachlan and Peel, 2000, p.214) recommends assigning principals to one of three types: (1) Output maximizers who always favor the contract that maximizes joint output. These principals do not attach any importance to agents' well-being. (2) Strong redistributors who always attach considerable importance to the low-ability agent's income, and (3) an intermediate group that attaches positive importance to the low-ability agent's income if the difference in piece rates becomes too great. We show that all principals in the *Spectator* treatment care to some extent about the distributive consequences of their decisions. On the contrary, 40% of *Stakeholder* principals are classified as output maximizers and are never willing to relinquish income to compress wages. This implies a sizable crowding out of inequality concerns through the provision of extensive margin incentives. Nevertheless, 60% of stakeholders are allocated to either type (2) or (3), suggesting that moral concerns persist on average, even if principals hold a stake in the workers' output. Counterfactual simulations that vary *workers'* other-regarding preferences show that egalitarian concerns are not always associated with a loss in profit for the firm. Sophisticated output-maximizing principals will mimic the behavior of egalitarian principals because they ultimately make the most efficient choices if agents are egalitarian. But when principals are naive and do not update their effort beliefs, then the egalitarian principals perform better for moderate agent inequality aversion levels.

This chapter contributes to answering the overarching research question by studying inequality attitudes in the context of a firm from the perspective of a manager or principal. The findings of the chapter illustrate that these attitudes matter for the provision of incentives in the firm.

Furthermore, we show that the incentive structure a manager faces leads to complete crowding out of distributive concerns for a substantial minority of managers but not for all. This implies that distributive concerns affect managerial decisions on average, even if they hold a stake in the firm's outcome.

### **Motivating beliefs in a just world**

The third chapter takes luck-effort belief in the focus of research. As I showed in the literature review, these beliefs are important predictors of inequality acceptance but our understanding of how they are formed is still limited. The chapter contributes to answering this broader research question by studying whether individuals distort luck-effort beliefs to motivate effort, a form of motivated beliefs that I call *motivating* beliefs. Bénabou and Tirole (2002) propose that individuals distort beliefs to counter an under-provision of effort due to self-control problems. These same authors created a model that shows how motivating beliefs affect preferences for redistribution (Bénabou and Tirole, 2006). Specifically, they showed that, if an economic agent with self-control problems expects low levels of redistribution, exerting low levels of effort can become very costly. This creates a demand to motivate future effort by distorting beliefs. To validate this model, it is important to provide empirical evidence of motivating belief distortion, because such evidence would show that luck-effort beliefs could be shaped by expectations about future levels of redistribution. This evidence would, thus, advance our understanding of the dynamic interaction between inequality, redistributive preferences, and beliefs about the importance of luck and effort. Such evidence also would imply that the causal relationship between luck-effort beliefs and redistribution runs both ways: Beliefs affect the demand for redistribution and expected levels of redistribution affect beliefs by shaping incentives.<sup>29</sup>

To test the prediction that future incentives distort beliefs about the importance of effort in achieving success I use an online experiment. In the experiment, whose stages are characterized by Figure 3, subjects begin by performing a cumbersome real effort task. This task is completed in an uncertain environment, where the payment rule depends on the state of the world (Environment) that was drawn at the beginning of the experiment. If the subject is in the Performance-Environment condition, the likelihood of winning a prize for completing the task is an increasing function of their performance on the task. Specifically, the subject participates in a noisy tournament against a randomly drawn competitor, where the chance of winning the prize is equal to 80% if the subject transcribes more images than her competitor, while the chance is equal to 20% if she transcribes fewer images. If the subject is in the Chance-Environment condition, her performance on the task has no effect on her likelihood of winning the prize; the subject wins the prize with 50% probability no matter how many words she transcribes.

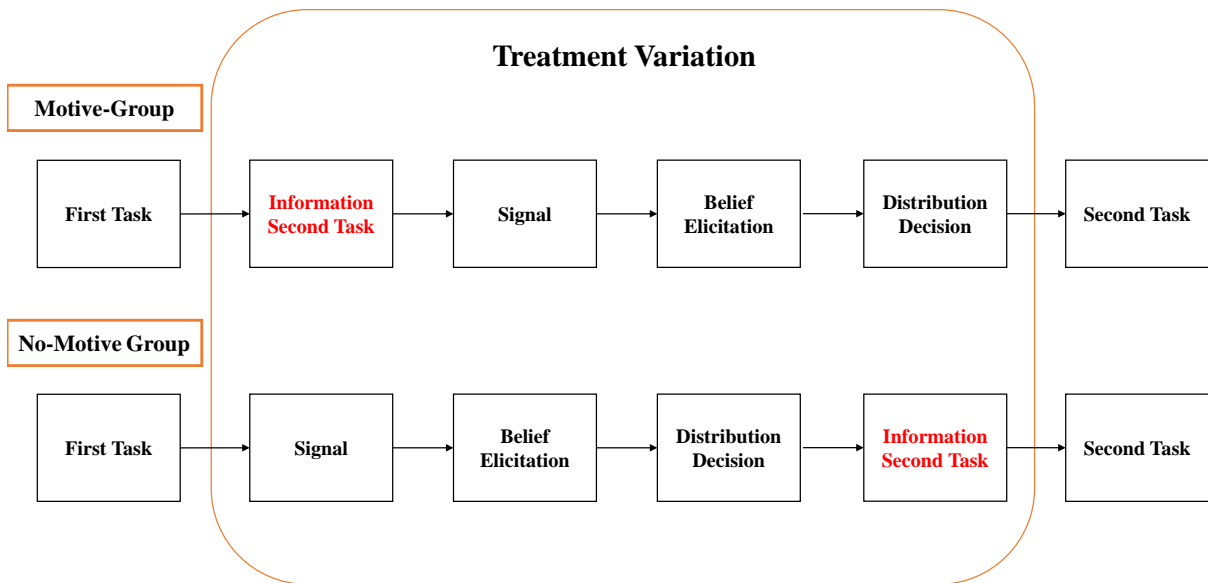
---

<sup>29</sup>Beliefs affect preferences for redistribution in several ways: First, there are selfish reasons to ask for less redistribution if one distorts beliefs to motivate future effort. An individual who believes that she will be a net contributor if her effort is reflected in the pre-tax income distribution would be less likely to support redistribution after engaging in motivating belief distortion. Second, motivating belief distortion may affect preferences for redistribution for other-regarding reasons. Meritocratic individuals who distort their beliefs in a motivating way are less likely to believe that an initially unequal distribution is due to luck and if they accept inequalities that reflect differences in effort, they should, hence, opt for less redistribution compared to a situation where they did not distort their beliefs to motivate future effort.



Upon completing the task, subjects receive a noisy signal informing them about which condition—Performance-Environment or the Chance-Environment—they were in. The aim of this signal is to induce variation in baseline beliefs regarding the state of the world. I induce these beliefs by giving subjects feedback about the outcome of the first task. The feedback comprises two pieces of information: (1) whether a subject won the prize and (2) whether she transcribed more or fewer images than her competitor. Using this information, subjects can form posterior beliefs about the likelihood of being in the Chance- or Performance-Environment. For example, a subject who learns that she transcribed more images than the competitor but did not win the prize is likely to perceive herself as having a high probability of being in the Chance-Environment, that is, the condition in which success is unrelated to effort. By contrast, if the same person learns that she won the prize, she should perceive herself as having a high probability of being in the Performance-Environment where effort does influence likelihood of success.<sup>30</sup> After subjects receive the signal, I elicit their probabilistic beliefs about the environment (Chance or Performance).

Figure 3: Schematic diagram characterizing the stages of the experiment



To test whether subjects distort baseline beliefs to motivate themselves to exert effort, I introduce a second task that subjects can complete at the end of the experiment and that serves as an incentive—and, hence, motive—to distort beliefs for motivating purposes. As in the first task, the payment rule depends on the environment that was drawn at the beginning of the experiment: If the subject is in the Performance-Environment, effort determines whether one receives

<sup>30</sup>The signal mimics real life experiences that people may use to infer the importance of effort for success in life: One’s colleague may get a promotion even though one considers oneself more talented and productive than the person who got the promotion; other people may get a position to which they applied, knowing that they only got the position because of their personal ties to the company’s CEO; still other individuals may win an award for their work knowing that they worked harder and performed better than the other people who were short-listed for the award.

a reward for performance on the second task, while for subjects in the Chance-Environment, effort has no effect on the likelihood of receiving a reward. Subjects who know about the second task may overestimate their likelihood of being in the Performance-Environment in order to motivate themselves to work hard on the second task.

To identify motivating belief distortion, I vary the point in time at which I inform subjects about the second task: Subjects who are assigned to the Motive treatment group are informed about the second task *before* belief elicitation and, hence, have an incentive to distort beliefs to motivate effort. Subjects in the No-Motive treatment group receive this information *after* belief elicitation. The latter subjects have no incentive to distort beliefs to motivate effort because they do not know that they will be completing a second task in the experiment. This variation allows me to test the main hypothesis of the experiment: Motive-group subjects, who know that they will be completing a second task, are, on average, more confident of being in the Performance-Environment than are No-Motive-group subjects who do not know that they have to exert effort in the future.

The design allows me to test for motivating belief distortion non-parametrically by comparing posterior beliefs across the two treatment groups. Nonetheless, I can go further and ask what type of signal leads to motivating belief distortion. First, I can ask whether subjects are more or less likely to engage in motivating belief distortion when receiving a signal that suggests that they are in the Performance-Environment rather than the Chance-Environment. Second, the design allows me to test whether events that are non-informative about the true state of the world affect beliefs. Specifically, I ask whether individuals are more likely to believe in the importance of effort if they won—rather than lost—a reward, holding the informational content of the event constant. This allows me to infer what type of events induce motivating belief distortion.

I further ask whether motivating belief distortion affects decisions about redistribution between two other individuals, as these beliefs are strong predictors of the demand for redistribution for meritocratic individuals. After the first phase of the experiment, in which subjects receive a signal about the environment to which they were assigned (i.e., Chance or Performance) and in which Motive-group subjects are informed about the second task, I give subjects the opportunity to redistribute an initially unequal bonus allocation between two uninvolved participants. These participants were previously recruited to perform the same first task as the decision maker herself. I truthfully tell participants that the initial allocation was determined by the same payment rule they themselves just faced. Subjects can then redistribute this initial allocation. By exploiting variation across treatment groups and signals, I can test (a) whether motivating belief distortion affects inequality-acceptance for other-regarding motives and (b) whether past experiences affect redistributive decisions above and beyond the experiences' informational content.

My results show that subjects distort beliefs to motivate future effort. Subjects who know they will perform another task in the same environment are significantly more confident (seven percentage points) of being in the Performance-Environment. This average effect masks heterogeneity by signal type. Motive-group subjects who received a disincentivizing signal indicating

that reward is unrelated to effort, i.e. that they were assigned to the Chance-Environment, are significantly more confident (nine percentage points) that they are in the Performance-Environment compared to control group subjects who received the same signal. My results do not show any difference in beliefs across treatment groups for subjects who received an incentivizing signal indicating that reward is a function of effort, i.e. that they were assigned to the Performance-Environment. This shows that motivating belief distortion is particularly frequent if people receive information that is disincentivizing, i.e., information that indicates that effort is not important for success. By exploiting independent variation in the event that leads to a given signal, I show that motivating belief distortion is particularly pronounced for individuals who know (or believe) that they would have done well in a world that actually rewards effort, i.e., people who learned that the outcome of the task was not justified by their relative performance. Overall, the results of my experiment provide strong evidence that individuals distort their own luck-effort beliefs to motivate themselves for the task they expect to face in the future.

Turning to the results on the distribution decision, I show that motivating belief distortion does not significantly affect distributive behavior. This suggests that beliefs may be instrumental for motivating future effort but this shift in beliefs is not strong enough to be reflected in aggregate distribution behavior. Even though subjects who are confident of being in the Performance-Environment are less likely to redistribute, I find that past experiences tend to matter a great deal for redistributive decisions in this context. Importantly, I find that subjects who did not win a prize and who performed worse than their competitor redistribute larger amounts than do other subjects, even though the former received a signal indicating a higher likelihood of being in the Performance-Environment. This result highlights the importance of taking into account event characteristics that are not informative about the relative importance of luck and effort when analyzing distributive behavior.

## 5 References

- Agranov, Marina and Thomas R. Palfrey. The effects of income mobility and tax persistence on income redistribution and inequality. *European Economic Review*, 123:103372, 2020.
- Akbaş, Merve, Dan Ariely, and Sevgi Yuksel. When is inequality fair? An experiment on the effect of procedural justice and agency. *Journal of Economic Behavior & Organization*, 161: 114–127, 2019.
- Akerlof, George A and Janet L Yellen. Fairness and Unemployment. *American Economic Review*, 78(2):44–49, 1988.
- Alesina, Alberto and George-Marios Angeletos. Fairness and redistribution: US vs. Europe. *American Economic Review*, 95(5):960–980, 2005.
- Alesina, Alberto and Paola Giuliano. Chapter 4 - Preferences for Redistribution. In Benhabib, Jess, Alberto Bisin, and Matthew O. Jackson, editors, *Handbook of Social Economics*, volume 1, pages 93–131. North-Holland, 2011.
- Alesina, Alberto and Eliana La Ferrara. Preferences for redistribution in the land of opportunities. *Journal of Public Economics*, 89(5-6):897–931, 2005.
- Alesina, Alberto, Edward L Glaeser, and Bruce Sacerdote. Why Doesn't the United States Have a European-Style Welfare State? *Brookings Papers on Economic Activity*, 2001(2):187–277, 2001.
- Alesina, Alberto, Stefanie Stantcheva, and Edoardo Teso. Intergenerational Mobility and Preferences for Redistribution. *American Economic Review*, 108(2):521–554, 2018.
- Alesina, Alberto, Armando Miano, and Stefanie Stantcheva. The Polarization of Reality. *AEA Papers and Proceedings*, 110:324–328, 2020.
- Almås, Ingvild, Alexander W Cappelen, Kjell G Salvanes, Erik Ø Sørensen, and Bertil Tungodden. Fairness and family background. *Politics, Philosophy & Economics*, 16(2):117–131, 2017.
- Almås, Ingvild, Alexander W. Cappelen, and Bertil Tungodden. Cutthroat Capitalism versus Cuddly Socialism: Are Americans More Meritocratic and Efficiency-Seeking than Scandinavians? *Journal of Political Economy*, 128(5):1753–1788, 2020.
- Almås, Ingvild, Alexander W Cappelen, Bertil Tungodden, and Erik Ø. Sørensen. Fairness Across the World: Preferences and Beliefs. *Working Paper*, 2021.
- Alvaredo, Facundo, Anthony B Atkinson, Thomas Piketty, and Emmanuel Saez. The Top 1 Percent in International and Historical Perspective. *Journal of Economic Perspectives*, 27(3): 3–20, 2013.

- Alvaredo, Facundo, Lucas Chancel, Thomas Piketty, Emmanuel Saez, and Gabriel Zucman. World Inequality Report 2018. *World Inequality Lab*, 2018.
- Amasino, Dianna, Davide D Pace, and Joël J. van der Weele. Fair Shares and Selective Attention. *Working Paper*, 2021.
- Andreoni, James and John Miller. Giving According To Garp: An Experimental Test of the Consistency of Preferences for Altruism. *Econometrica*, 70(2):737–753, 2002.
- Arneson, Richard J. Equality and Equal Opportunity for Welfare. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 56(1):77–93, 1989.
- Ashok, Vevikin, Ilyana Kuziemko, and Ebonya Washington. Support for redistribution in an age of rising inequality: New stylized facts and some tentative explanations. *Brookings Papers on Economic Activity*, pages 367–405, 2015.
- Ashraf, Nava, Iris Bohnet, and Nikita Piankov. Decomposing trust and trustworthiness. *Experimental Economics*, 9(3):193–208, 2006.
- Atkinson, Anthony B, Thomas Piketty, and Emmanuel Saez. Top Incomes in the Long Run of History. *Journal of Economic Literature*, 49(1):3–71, 2011.
- Babcock, Linda, George Loewenstein, Samuel Issacharoff, and Colin Camerer. Biased Judgments of Fairness in Bargaining. *The American Economic Review*, 85(5):1337–1343, 1995.
- Barr, Abigail, Luis Miller, and Paloma Ubeda. Moral consequences of becoming unemployed. *Proceedings of the National Academy of Sciences*, 113(17):4676–4681, 2016.
- Barron, Kai, Robert Stüber, and Roel van Veldhuizen. Motivated motive selection in the lying-dictator game. *Working Paper*, 2020.
- Bartling, Björn and Ferdinand A Von Siemens. The intensity of incentives in firms and markets: Moral hazard with envious agents. *Labour Economics*, 17(3):598–607, 2010.
- Bartling, Björn and Ferdinand A. von Siemens. Wage inequality and team production: An experimental analysis. *Journal of Economic Psychology*, 32(1):1–16, 2011.
- Bauer, Michal, Alessandra Cassar, Julie Chytilová, and Joseph Henrich. War’s Enduring Effects on the Development of Egalitarian Motivations and In-Group Biases. *Psychological Science*, 25(1):47–57, 2014.
- Bellemare, Charles, Sabine Kröger, and Arthur Van Soest. Measuring inequity aversion in a heterogeneous population using experimental decisions and subjective probabilities. *Econometrica*, 76(4):815–839, 2008.
- Bénabou, Roland and Efe A Ok. Social mobility and the demand for redistribution: The POUM hypothesis. *The Quarterly Journal of Economics*, 116(2):447–487, 2001.

- Bénabou, Roland and Jean Tirole. Self-Confidence and Personal Motivation. *The Quarterly Journal of Economics*, 117(3):871–915, 2002.
- Bénabou, Roland and Jean Tirole. Belief in a Just World and Redistributive Politics. *The Quarterly Journal of Economics*, 121(2):699–746, 2006.
- Bentham, Jeremy. *An Introduction to the Principles of Morals and Legislation*. T. Payne, 1789.
- Bjerk, David. In front of and behind the veil of ignorance: An analysis of motivations for redistribution. *Social Choice and Welfare*, 47(4):791–824, 2016.
- Bolton, Gary E and Axel Ockenfels. ERC: A theory of equity, reciprocity, and competition. *American Economic Review*, 90(1):166–193, 2000.
- Bracha, Anat, Uri Gneezy, and George Loewenstein. Relative Pay and Labor Supply. *Journal of Labor Economics*, 33(2):297–315, 2015.
- Breza, Emily, Supreet Kaur, and Yogita Shamdasani. The morale effects of pay inequality. *The Quarterly Journal of Economics*, 133(2):611–663, 2018.
- Bruhin, Adrian, Ernst Fehr, and Daniel Schunk. The many faces of human sociality: Uncovering the distribution and stability of social preferences. *Journal of the European Economic Association*, 17(4):1025–1069, 2018.
- Buser, Thomas, Gianluca Grimalda, Louis Putterman, and Joël van der Weele. Overconfidence and gender gaps in redistributive preferences: Cross-Country experimental evidence. *Journal of Economic Behavior & Organization*, 178:267–286, 2020.
- Camerer, Colin F. The Promise and Success of Lab–Field Generalizability in Experimental Economics: A Critical Reply to Levitt and List. In Fréchette, Guillaume R. and Andrew Schotter, editors, *Handbook of Experimental Economic Methodology*, pages 249–295. Oxford University Press, 2015.
- Cappelen, Alexander, John List, Anya Samek, and Bertil Tungodden. The Effect of Early-Childhood Education on Social Preferences. *Journal of Political Economy*, 128(7):2739–2758, 2020a.
- Cappelen, Alexander W. and Bertil Tungodden. Rewarding effort. *Economic Theory*, 39(3):425–441, 2009.
- Cappelen, Alexander W, Astri Drange Hole, Erik Ø Sørensen, and Bertil Tungodden. The pluralism of fairness ideals: An experimental approach. *The American Economic Review*, 97(3):818–827, 2007.
- Cappelen, Alexander W, James Konow, Erik Ø Sørensen, and Bertil Tungodden. Just luck: An experimental study of risk-taking and fairness. *The American Economic Review*, 103(4):1398–1413, 2013.

- Cappelen, Alexander W., Ranveig Falch, and Bertil Tungodden. Fair and Unfair Income Inequality. In Zimmermann, Klaus F., editor, *Handbook of Labor, Human Resources and Population Economics*, pages 1–25. Springer International Publishing, Cham, 2020b.
- Cappelen, Alexander W., Ranveig Falch, Erik Ø. Sørensen, and Bertil Tungodden. Solidarity and fairness in times of crisis. *Journal of Economic Behavior & Organization*, 186:1–11, 2021.
- Card, David, Alexandre Mas, Enrico Moretti, and Emmanuel Saez. Inequality at work: The effect of peer salaries on job satisfaction. *The American Economic Review*, 102(6):2981–3003, 2012.
- Carlsson, Fredrik, Dinky Daruvala, and Olof Johansson-Stenman. Are People Inequality-Averse, or Just Risk-Averse? *Economica*, 72(287):375–396, 2005.
- Cassar, Lea and Arnd H. Klein. A Matter of Perspective: How Failure Shapes Distributive Preferences. *Management Science*, 65(11):5050–5064, 2019.
- Charness, G. and M. Rabin. Understanding Social Preferences with Simple Tests. *The Quarterly Journal of Economics*, 117(3):817–869, 2002.
- Charness, Gary and Peter Kuhn. Does Pay Inequality Affect Worker Effort? Experimental Evidence. *Journal of Labor Economics*, 25(4):693–723, 2007.
- Charness, Gary and Peter Kuhn. Lab Labor: What Can Labor Economists Learn from the Lab? *Handbook of Labor Economics*, 4:229–330, 2011.
- Checchi, Daniele and Antonio Filippin. An experimental study of the POUM hypothesis. In *Inequality, Welfare and Income Distribution: Experimental Approaches*, pages 115–136. Emerald Group Publishing Limited, 2004.
- Cherry, Todd L, Peter Frykblom, and Jason F Shogren. Hardnose the Dictator. *The American Economic Review*, 92(4):1218–1221, 2002.
- Clark, Andrew E., Nicolai Kristensen, and Niels Westergård-Nielsen. Job satisfaction and co-worker wages: Status or signal? *The Economic Journal*, 119(536):430–447, 2009.
- Clark, Andrew E, David Masclet, and Marie Claire Villeval. Effort and comparison income: Experimental and survey evidence. *Industrial and Labor Relations Review*, 63(3):407–426, 2010.
- Cohn, Alain, Ernst Fehr, Benedikt Herrmann, and Frédéric Schneider. Social comparison and effort provision: Evidence from a field experiment. *Journal of the European Economic Association*, 12(4):877–898, 2014.
- Cooper, David J. and John H. Kagel. Other-Regarding Preferences: A selective Survey of Experimental Results. In Kagel, John H and Alvin E Roth, editors, *The Handbook of Experimental Economics*, Vol. 2, pages 2017–2282. Princeton University Press, Princeton, 2016.

- Cruces, Guillermo, Ricardo Perez-Truglia, and Martin Tetaz. Biased perceptions of income distribution and preferences for redistribution: Evidence from a survey experiment. *Journal of Public Economics*, 98:100–112, 2013.
- Cullen, Zoe and Ricardo Perez-Truglia. How Much Does Your Boss Make? The Effects of Salary Comparisons. *Working Paper*, 2018.
- Dalton, Hugh. The Measurement of the Inequality of Incomes. *The Economic Journal*, 30(119):348, 1920.
- Dana, Jason, Roberto A. Weber, and Jason Xi Kuang. Exploiting moral wiggle room: Experiments demonstrating an illusory preference for fairness. *Economic Theory*, 33(1):67–80, 2007.
- De Bruyn, Arnaud and Gary E. Bolton. Estimating the Influence of Fairness on Bargaining Behavior. *Management Science*, 54(10):1774–1791, 2008.
- Deffains, Bruno, Romain Espinosa, and Christian Thöni. Political self-serving bias and redistribution. *Journal of Public Economics*, 134:67–74, 2016.
- Di Tella, Rafael, Sebastian Galiani, and Ernesto Schargrotsky. The Formation of Beliefs: Evidence from the Allocation of Land Titles to Squatters. *The Quarterly Journal of Economics*, 122(1):209–241, 2007.
- Duesenberry, James S. *Income, Saving, and the Theory of Consumer Behavior*. Harvard University Press, first edition edition, 1949.
- Durante, Ruben, Louis Putterman, and Joël van der Weele. Preferences for Redistribution and Perception of Fairness: An Experimental Study. *Journal of the European Economic Association*, 12(4):1059–1086, 2014.
- Dworkin, Ronald. *Sovereign Virtue: The Theory and Practice of Equality*. Harvard University Press, Cambridge, 2002.
- Engel, Christoph. Dictator games: A meta study. *Experimental Economics*, 14(4):583–610, 2011.
- Engelmann, Dirk and Martin Strobel. Inequality aversion, efficiency, and maximin preferences in simple distribution experiments. *The American Economic Review*, 94(4):857–869, 2004.
- Engelmann, Dirk and Martin Strobel. Preferences over income distributions experimental evidence. *Public Finance Review*, 35(2):285–310, 2007.
- Englmaier, Florian and Achim Wambach. Optimal incentive contracts under inequity aversion. *Games and Economic Behavior*, 69(2):312–328, 2010.
- Epper, Thomas, Ernst Fehr, and Julien Senn. Other-Regarding Preferences and Redistributive Politics. *Working Paper*, 2020.



- Fehr, Dietmar and Martin Vollmann. Misperceiving Economic Success: Experimental Evidence on Meritocratic Beliefs and Inequality Acceptance. *Working Paper*, 2020.
- Fehr, E., G. Kirchsteiger, and A. Riedl. Does Fairness Prevent Market Clearing? An Experimental Investigation. *The Quarterly Journal of Economics*, 108(2):437–459, 1993.
- Fehr, Ernst and Klaus M Schmidt. A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, 114(3):817–868, 1999.
- Fehr, Ernst, Michael Naef, and Klaus M Schmidt. Inequality aversion, efficiency, and maximin preferences in simple distribution experiments: Comment. *The American Economic Review*, 96(5):1912–1917, 2006.
- Festinger, Leon. *A Theory of Cognitive Dissonance*. Stanford University Press, 1957.
- Fisman, Raymond, Shachar Kariv, and Daniel Markovits. Individual preferences for giving. *American Economic Review*, 97(5):1858–1876, 2007.
- Fisman, Raymond, Pamela Jakiela, and Shachar Kariv. How did distributional preferences change during the Great Recession? *Journal of Public Economics*, 128:84–95, 2015a.
- Fisman, Raymond, Pamela Jakiela, Shachar Kariv, and Daniel Markovits. The distributional preferences of an elite. *Science*, 349(6254), 2015b.
- Fisman, Raymond, Pamela Jakiela, and Shachar Kariv. Distributional preferences and political behavior. *Journal of Public Economics*, 155:1–10, 2017.
- Fisman, Raymond, Pamela Jakiela, Shachar Kariv, and Silvia Vannutelli. The Distributional Preferences of Americans. *Working Paper*, 2020.
- Fleurbaey, Marc. *Fairness, Responsibility, and Welfare*. Oxford University Press, 2008.
- Fong, Christina. Social preferences, self-interest, and the demand for redistribution. *Journal of Public Economics*, 82(2):225–246, 2001.
- Forsythe, Robert, Joel L. Horowitz, N.E. Savin, and Martin Sefton. Fairness in Simple Bargaining Experiments. *Games and Economic Behavior*, 6(3):347–369, 1994.
- Frank, Robert H. Are Workers Paid their Marginal Products? *The American Economic Review*, 74(4):549–571, 1984.
- Frank, Robert H. *Success and Luck: Good Fortune and the Myth of Meritocracy*. Princeton University Press, 2017.
- Frohlich, Norman, Joe A. Oppenheimer, and Cheryl L. Eavey. Choices of Principles of Distributive Justice in Experimental Groups. *American Journal of Political Science*, 31(3):606, 1987.
- Frohlich, Norman, Joe Oppenheimer, and Anja Kurki. Modeling other-regarding preferences and an experimental test. *Public Choice*, 119(1-2):91–117, 2004.

- Gächter, Simon and Christian Thöni. Social comparison and performance: Experimental evidence on the fair wage–effort hypothesis. *Journal of Economic Behavior & Organization*, 76(3):531–543, 2010.
- Gächter, Simon, Daniele Nosenzo, and Martin Sefton. The Impact of Social Comparisons on Reciprocity: Impact of social comparisons on reciprocity. *The Scandinavian Journal of Economics*, 114(4):1346–1367, 2012.
- Gärtner, Manja, Johanna Mollerstrom, and David Seim. Individual risk preferences and the demand for redistribution. *Journal of Public Economics*, 153:49–55, 2017.
- Gärtner, Manja, Johanna Möllerström, and David Seim. Income Mobility, Luck/Effort Beliefs, and the Demand for Redistribution: Perceptions and Reality. *Working Paper*, 2019.
- Giuliano, Paola and Antonio Spilimbergo. Growing up in a Recession. *The Review of Economic Studies*, 81(2):787–817, 2014.
- Godechot, Olivier and Claudia Senik. Wage comparisons in and out of the firm. Evidence from a matched employer-employee French database. *Journal of Economic Behavior & Organization*, 117:395–410, 2015.
- Greiner, Ben, Axel Ockenfels, and Peter Werner. Wage transparency and performance: A real-effort experiment. *Economics Letters*, 111(3):236–238, 2011.
- Grosfeld, Irena and Claudia Senik. The emerging aversion to inequality. *Economics of Transition*, 18(1):1–26, 2010.
- Gross, Till, Christopher Guo, and Gary Charness. Merit pay and wage compression with productivity differences and uncertainty. *Journal of Economic Behavior and Organization*, 117:233–247, 2015.
- Güth, Werner, Rolf Schmittberger, and Bernd Schwarze. An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior & Organization*, 3(4):367–388, 1982.
- Harsanyi, John C. Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *Journal of Political Economy*, 63(4):309–321, 1955.
- Hedegaard, Morten, Rudolf Kerschbamer, Daniel Müller, and Jean-Robert Tyran. Distributional preferences explain individual behavior across games and time. *Games and Economic Behavior*, 128:231–255, 2021.
- Hennig-Schmidt, Heike, Abdolkarim Sadrieh, and Bettina Rockenbach. In Search of Workers’ Real Effort Reciprocity—a Field and a Laboratory Experiment. *Journal of the European Economic Association*, 8(4):817–837, 2010.
- Hirschman, Albert O and Michael Rothschild. The Changing Tolerance for Income Inequality in the Course of Economic Development\* With A Mathematical Appendix. *The Quarterly Journal of Economics*, 87(4):544–566, 1973.

- Hoffman, Elizabeth, Kevin McCabe, Keith Shachat, and Vernon Smith. Preferences, property rights, and anonymity in bargaining games. *Games and Economic behavior*, 7(3):346–380, 1994.
- Jacquemet, Nicolas and Olivier L’Haridon. *Experimental Economics: Method and Applications*. Cambridge University Press, Cambridge, 2018.
- Jakiela, Pamela. How fair shares compare: Experimental evidence from two cultures. *Journal of Economic Behavior & Organization*, 118:40–54, 2015.
- Johansson-Stenman, Olof, Fredrik Carlsson, and Dinky Daruvala. Measuring Future Grandparents’ Preferences for Equality and Relative Standing. *The Economic Journal*, 112(479): 362–383, 2002.
- Kahneman, Daniel, Jack L. Knetsch, and Richard Thaler. Fairness as a Constraint on Profit Seeking: Entitlements in the Market. *The American Economic Review*, 76(4):728–741, 1986a.
- Kahneman, Daniel, Jack L. Knetsch, and Richard H. Thaler. Fairness and the Assumptions of Economics. *The Journal of Business*, 59(4):S285–S300, 1986b.
- Konow, James. A positive theory of economic fairness. *Journal of Economic Behavior & Organization*, 31(1):13–35, 1996.
- Konow, James. Fair Shares: Accountability and Cognitive Dissonance in Allocation Decisions. *American Economic Review*, 90(4):1072–1091, 2000.
- Konow, James. Which Is the Fairest One of All? A Positive Analysis of Justice Theories. *Journal of Economic Literature*, 41(4):1188–1239, 2003.
- Kosse, Fabian, Thomas Deckers, Pia Pinger, Hannah Schildberg-Hörisch, and Armin Falk. The Formation of Prosociality: Causal Evidence on the Role of Social Environment. *Journal of Political Economy*, 128(2):434–467, 2020.
- Kranton, Rachel, Matthew Pease, Seth Sanders, and Scott Huettel. Deconstructing bias in social preferences reveals groupy and not-groupy behavior. *Proceedings of the National Academy of Sciences*, 117(35):21185–21193, 2020.
- Kritikos, Alexander and Friedel Bolle. Distributional concerns: Equity-or efficiency-oriented? *Economics Letters*, 73(3):333–338, 2001.
- Krosnick, Jon A. and Duane F. Alwin. Aging and susceptibility to attitude change. *Journal of Personality and Social Psychology*, 57(3):416–425, 1989.
- Ku, Hyejin and Timothy C. Salmon. The Incentive Effects of Inequality: An Experimental Investigation. *Southern Economic Journal*, 79(1):46–70, 2012.
- Kuziemko, Ilyana, Michael I Norton, Emmanuel Saez, and Stefanie Stantcheva. How elastic are preferences for redistribution? Evidence from randomized survey experiments. *The American Economic Review*, 105(4):1478–1508, 2015.

- Ledyard, John O. Public Goods: A survey of Experimental Research. In Kagel, John H and Alvin E Roth, editors, *The Handbook of Experimental Economics*, pages 111–149. Princeton University Press, Princeton, 1995.
- Lefgren, Lars J., David P. Sims, and Olga B. Stoddard. Effort, luck, and voting for redistribution. *Journal of Public Economics*, 143:89–97, 2016.
- Lerner, Melvin J. The Belief in a Just World. In Lerner, Melvin J., editor, *The Belief in a Just World: A Fundamental Delusion*, Perspectives in Social Psychology, pages 9–30. Springer US, Boston, MA, 1980.
- Luttmer, Erzo F P. Group loyalty and the taste for redistribution. *Journal of political Economy*, 109(3):500–528, 2001.
- Luttmer, Erzo F. P and Monica Singhal. Culture, Context, and the Taste for Redistribution. *American Economic Journal: Economic Policy*, 3(1):157–179, 2011.
- McLachlan, Geoffrey J. and David Peel. *Finite Mixture Models*. Wiley Series in Probability and Statistics. Applied Probability and Statistics Section. Wiley, New York, 2000.
- Meltzer, Allan H. and Scott F. Richard. A Rational Theory of the Size of Government. *Journal of Political Economy*, 89(5):914–927, 1981.
- Michelbach, Philip A, John T Scott, Richard E Matland, and Brian H Bornstein. Doing Rawls justice: An experimental study of income distribution norms. *American Journal of Political Science*, 47(3):523–539, 2003.
- Mollerstrom, Johanna, Bjørn-Atle Reme, and Erik Ø Sørensen. Luck, choice and responsibility – An experimental study of fairness views. *Journal of Public Economics*, 131:33–40, 2015.
- Müller, Daniel and Sander Renes. Fairness views and political preferences: Evidence from a large and heterogeneous sample. *Social Choice and Welfare*, 56(4):679–711, 2021.
- Nielsen, Kai. *Equality and Liberty: A Defense of Radical Egalitarianism*. Rowman & Littlefield, 1985.
- Obama, Barack. Remarks by the President on Economic Mobility. *The White House Office of the Press Secretary*, 2013.
- Okun, Arthur M. *Equality and Efficiency: The Big Tradeoff*. Brookings Institution Press, 1975.
- Owens, Lindsay A. and David S. Pedulla. Material Welfare and Changing Political Preferences: The Case of Support for Redistributive Social Policies. *Social Forces*, 92(3):1087–1113, 2014.
- Piketty, Thomas. Social mobility and redistributive politics. *The Quarterly Journal of Economics*, 110(3):551–584, 1995.
- Piketty, Thomas. *Capital in the Twenty-First Century*. Harvard University Press, Cambridge, 2014.

- Piketty, Thomas. *Capital and Ideology*. Harvard University Press, Cambridge, 2020.
- Rawls, John. *A Theory of Justice*. Harvard university press, Cambridge, 1971.
- Roemer, John E. *Equality of Opportunity*. Harvard University Press, Cambridge, 1998.
- Romer, Thomas. Individual welfare, majority voting, and the properties of a linear income tax. *Journal of Public Economics*, 4(2):163–185, 1975.
- Roth, Alvin E. Laboratory Experimentation in Economics: A Methodological Overview. *The Economic Journal*, 98(393):974, 1988.
- Roth, Alvin E. Is Experimental Economics Living Up to Its Promise? In Fréchette, Guillaume R. and Andrew Schotter, editors, *Handbook of Experimental Economic Methodology*, pages 13–40. Oxford University Press, 2015.
- Roth, Christopher and Johannes Wohlfart. Experienced inequality and preferences for redistribution. *Journal of Public Economics*, 167:251–262, 2018.
- Ruffle, Bradley J. More is better, but fair is fair: Tipping in dictator and ultimatum games. *Games and Economic Behavior*, 23(2):247–265, 1998.
- Saez, Emmanuel and Gabriel Zucman. The Rise of Income and Wealth Inequality in America: Evidence from Distributional Macroeconomic Accounts. *Journal of Economic Perspectives*, 34(4):3–26, 2020.
- Sandel, Michael J. *The Tyranny of Merit: What’s Become of the Common Good?* Farrar, Straus and Giroux, 2020.
- Schildberg-Hörisch, Hannah. Is the veil of ignorance only a concept about risk? An experiment. *Journal of Public Economics*, 94(11):1062–1066, 2010.
- Sen, Amartya. Utilitarianism and Welfarism. *The Journal of Philosophy*, 76(9):463–489, 1979.
- Senik, Claudia. When information dominates comparison: Learning from Russian subjective panel data. *Journal of Public Economics*, 88(9):2099–2123, 2004.
- Shayo, Moses. A Model of Social Identity with an Application to Political Economy: Nation, Class, and Redistribution. *American Political Science Review*, 103(2):147–174, 2009.
- Snowberg, Erik and Leeat Yariv. Testing the Waters: Behavior across Participant Pools. *American Economic Review*, 111(2):687–719, 2021.
- Stantcheva, Stefanie. Understanding Tax Policy: How Do People Reason? *Working Paper*, 2020.
- Tyran, Jean-Robert and Rupert Sausgruber. A little fairness may induce a lot of redistribution in democracy. *European Economic Review*, 50(2):469–485, 2006.

Valero, Vanessa. Redistribution and Beliefs about the Source of Income Inequality. *Working Paper*, 2020.

Veblen, Thorstein. *The Theory of the Leisure Class*. Courier Corporation, 1899.

Yaari, M. E. and M. Bar-Hillel. On Dividing Justly. *Social Choice and Welfare*, 1(1):1–24, 1984.

# Chapter 1

## Preferences over Income Distribution: Evidence from a Choice Experiment

co-written with Sophie Cêtre (IRSN), Claudia Senik (PSE, University Paris-Sorbonne) and Thierry Verdier (PSE, Ecole des Ponts-ParisTech, PUC-Rio and CEPR London)

### Abstract

Using a choice experiment in the lab, we assess the relative importance of different motives behind attitudes to income inequality. We elicit subjects' preferences regarding pairs of payoff distributions within small groups, in a firm-like setting. We find that distributions that satisfy the Pareto-dominance criterion attract unanimous suffrage: all subjects prefer larger inequality provided it makes everyone weakly better off. This is true no matter whether payoffs are based on merit or luck. Unanimity only breaks once subjects' positions within the income distribution are fixed and known ex-ante. Even then, 75% of subjects prefer Pareto-dominant distributions, but 25% of subjects engage in money burning at the top in order to reduce inequality, even when it does not make anyone better off. A majority of subjects embrace a more equal distribution if their own income or overall efficiency is not at stake. When their own income is at stake and the sum of payoffs remains unaffected, 20% of subjects are willing to pay for a lower degree of inequality.

This chapter has been published:

Cêtre, Sophie, Max Lobeck, Claudia Senik and Thierry Verdier. Preferences over Income Distribution: Evidence from a Choice Experiment. *Journal of Economic Psychology*, 74, 102202, 2019.

---

We are grateful for financial support from the European Research Council (ERC TECTACOM), CEPREMAP, and the French National Research Agency, through the program Investissements d'Avenir, ANR-10-LABX-93-01. We are grateful to two anonymous referees for their insightful comments and suggestions. Special thanks go to Maxim Frolov for assisting in the implementation of the experimental sessions. For helpful comments we thank Nicolas Jacquemet, Beatrice Boulu-Reshef, and seminar participants at PSE, Sciences Po, GATE (Université Lumière Lyon 2), and conferences in Maastricht (MBEES), Cologne (EEA 2018), Essen (RGS Doctoral conference), Brighton (RES Junior Symposium) and Paris (ENS, Paris-1, AFSE 2018.)

## 1 Introduction

One of the most important questions in economics is how to divide the social surplus and whether income inequality is acceptable or not. This question is relevant not only in a societal context but also within smaller groups, such as firms and organizations (Card et al., 2012; Breza et al., 2018). Several motives behind attitudes to income distribution have been unearthed and discussed by an abundant literature. These include purely self-regarding motives, whereby people focus on their own current income, or income gaps vis-à-vis other relevant groups, as well as other-regarding motives, such as the fairness of the income generation process (merit versus luck) (Konow, 2000), pure aversion to income differences or to the topmost incomes (Fehr and Schmidt, 1999; Yang et al., 2016), or the Rawlsian pro-poor difference principle and maximin preferences (Charness and Rabin, 2002). These different motives are more or less salient depending on the setting in which people have to make their choice, e.g. whether they make the choice behind the veil of ignorance or not.

This chapter builds upon this existing literature and assesses systematically how these motives depend on the context of the choice, focusing on three aspects that have never been combined in one experiment studying inequality within small groups: 1) the Pareto-dominance criterion, i.e. whether an income distribution allows everyone to be weakly better off compared to the other distribution, 2) whether choices are made behind the veil of ignorance or with the position known, and 3) whether relative payoffs are based on merit or luck. We use a choice experiment framed as a series of choices between two projects that lead to different “bonus” distributions. More precisely, our design asks subjects to make a series of incentivized binary choices between two payoff distributions for a group of five individuals (the subject and four additional anonymous participants in the lab). Between subjects, we vary the origin of people’s position within the distribution (either based on luck or a real effort task). Within subjects, we vary whether one distribution is Pareto-dominant as compared to the other or not. We also ask subjects to choose successively behind the veil of ignorance, hence not knowing their future rank and payoff, and then with information about their position within their group. The series of binary choices that subjects have to make can be split into two categories. In the first category of choices, the total payoff is the same in the two proposed projects, but one distribution is more unequal and has higher top incomes and lower bottom incomes. In the second category of choices, the more unequal project Pareto-dominates the more equal one, i.e. it makes all of the group members weakly better off in absolute terms. Finally, we randomly assign subjects into a Merit and a Luck treatment. In the Merit treatment, people’s position within their group of 5 is determined by their relative performance in an effort task to be performed after the choices are made behind the veil of ignorance. In the Luck treatment, the ranking is determined randomly.

Our main finding is that, behind the veil of ignorance, subjects unanimously prefer the higher inequality project when it is Pareto-dominant. In this case, it does not make any difference whether subjects belong to the Luck treatment or the Merit treatment. Unanimity only breaks once positions within the income distributions are fixed, i.e. when subjects know their own ranking before they choose. In that setting, about 75% of subjects prefer the Pareto-



dominant distribution over a more compressed payoff distribution. The other 25% engage in money burning. They burn money at the top by choosing the low inequality project even if it does not improve the lot of the low earners. Furthermore, when subjects choose between two distributions that have the same efficiency (same total payoff), about 65% of them prefer the low-inequality distribution. When choosing behind the veil of ignorance, subjects are significantly more likely to embrace the high inequality distribution if they are in the Merit rather than the Luck treatment. This significant treatment effect disappears as soon as subjects learn about their rank, whereupon 70% of subjects prefer lower inequality when their own payoff is not affected. All subjects who are better off in the *low* inequality distribution choose the latter, but only 80% of subjects who would be better off in the *high* inequality distribution choose the latter. Hence, 20% of individuals are strongly inequality averse and act accordingly, even when this comes at a personal cost.

Our results contribute to the vast literature on distributive preferences. One part of this literature focuses on various distributive motives. Engelmann and Strobel (2004, 2007) use a multi-player dictator game where they let subjects choose between three different payoff distributions affecting them and two other players. Their main finding is that inequality aversion does not play a major role in explaining behavior, as compared to maximin preferences, efficiency concerns, and selfishness. Findings from an earlier experiment by Kritikos and Bolle (2001) are in line with Engelmann and Strobel findings. Bolton and Ockenfels (2006) comment on Engelmann and Strobel’s study by exploiting a similar choice experiment but find on the contrary that equity concerns are stronger than preferences for efficiency. In these papers, subjects choose without uncertainty about their future payoff and position in the distribution. Kamas and Preston (2012) find that distributive preferences predict behavior in games that involve reciprocity. Our experiment also relates directly to papers testing Rawls’ theory experimentally, such as Michelbach et al. (2003) and Frohlich et al. (1987). Michelbach et al. (2003) create 9 different income distributions that vary in terms of 4 allocation principles: equality, efficiency, need, and merit. Subjects make hypothetical choices as impartial spectators: they appear to care about both equality and efficiency and seem to be doing their best to strike a balance between those two principles. Choices also vary a lot across socio-demographic characteristics. Beckman et al. (2002) marks one of the few experimental contributions that test explicitly whether subjects are more likely to vote in favor of Pareto-efficient distribution behind the veil of ignorance rather than when their position is revealed. Like us, they find evidence that “envy” is more pronounced if subjects choose with known rank rather than behind the veil of ignorance.

We also complement other studies that focus more closely on the Merit versus Luck hypothesis, a distinction that is often considered as a criterion of fairness (Overlaet, 1991). Those who believe that the rich and the poor owe their situation to luck rather than effort may want to correct these “unfair” differences through income redistribution. Such theories of desert (Konow, 2003) have been documented empirically (Fong, 2001). They have been used to explain international differences in the demand for income redistribution and the extent of fiscal redistribution, in particular, the divide between European countries and the United States (Alesina and La Ferrara, 2005; Alesina and Angeletos, 2005). One strand of the experimental literature

studies whether impartial spectators are willing to distribute income equally or, instead, proportionally to their production (Konow, 2000; Cappelen et al., 2007). Sharma (2015) explicitly studies the role of gender in dictator games with a preceding output phase showing that men keep a greater share of the surplus than women. Other experimental studies use two-player dictator or ultimatum games to look at the effect of fairness concerns on altruistic behavior. They often use a contest, with a real effort task, and find that agents behave selfishly if the role of the first mover was earned rather than received without effort (Hoffman et al., 1994). Cherry et al. (2002) find a similar result for dictator games: dictators are less generous if their endowment was earned rather than simply received.<sup>1</sup> Dengler-Roscher et al. (2018) test the malleability of fairness ideals in a setting where income is generated in a production phase. They find that making an impartial allocation decision before (rather than after) playing a dictator game affects inexperienced subjects' degree of selfishness. Other experiments looked at multi-player versions of these games (Krawczyk, 2010; Durante et al., 2014; Bjerk, 2016; Lefgren et al., 2016), focusing mostly on redistribution and preferences over taxation rather than pure distributive preferences. They often conclude that merit matters, but Durante et al. (2014) and Bjerk (2016) look at preferences over taxation and find that the source of income does not affect behavior if individuals know their rank within the distribution.

## 2 Design of the experiment

Our laboratory experiment is designed to test how preferences over different income distributions depend on the three arguments described earlier: uncertainty about one's position, Pareto dominance and the Merit versus Luck hypothesis.

### 2.1 Overview

Before turning to the details of the design, let us focus first on a chronological summary shedding light on what information subjects hold at each step of the experiment. They first perform simple lottery choices to elicit their risk aversion. They are then randomly allocated into groups of 5 and are informed that their identity and that of the other members of their group will remain secret throughout the experiment. All subjects are asked to imagine that they must carry out a project within a firm or an organization with their group.<sup>2</sup>

In the first part of the core of the experiment, subjects are instructed that they will have to make a series of binary choices between two payoff (called "bonus") distributions, Project A and Project B, for their group. They are informed that they have to make their decision without knowing their position within the group ("behind the veil of ignorance" choice). In

---

<sup>1</sup>Ruffle (1998) also shows that if the winner of a contest contributes more to the total payoff, she is also awarded a higher split by the dictator. Similarly, Oxoby and Spraggon (2008) find that if the dictator earns the total payoff, she behaves selfishly, but if the receiver earns it, she allocates on average a significant amount to the receiver. Finally, in this kind of effort-based ranking experiment, preference for equality can be trumped by the will to obtain performance feedback (Alós-Ferrer et al., 2018).

<sup>2</sup>The exact framing is *Imagine that you are in a group with which you are carrying out a project, within a firm or an organization. You will choose between several projects that give each member of your group different bonuses.*

the Luck treatment, we explicitly inform subjects that their positions will be drawn randomly, while in the Merit treatment, they know that their positions will be based on a simple task to be performed later in the experiment.<sup>3</sup> In both treatments, this information on Merit or Luck is revealed at the same moment: right before they start making the binary choices behind the veil of ignorance (see more details in section 2.3). All subjects also go through a comprehension test to show them examples of the choices they will have to make.<sup>4</sup>

After they make their choices, subjects perform the real effort task. Then they are informed about their position, which is not going to change until the end of the experiment. In the second part, they make six of the eight choices, but this time, they know where they stand in the income distribution. The choices we drop in the second part are Choice 4 in Table A6 and the sanity check Choice 8.<sup>5</sup> These variations match real-life situations that can occur within a firm: choosing behind the veil of ignorance corresponds to a situation where positions within the firm are open; by contrast, fixed positions in the distribution evoke a situation where there is no prospect of mobility.

Why use such a within-subject design? One could argue that subjects may be tempted to stick to the same decisions both behind the veil of ignorance and with known positions in order to avoid cognitive dissonance. On the contrary, subjects could also overreact to the new setting due to an experimenter effect. We alleviate both concerns by randomizing the order of the choices and the labeling (Project A or Project B) in order to make it harder for them to make the same decisions blindly throughout. While acknowledging these limitations, we nevertheless think that this within-subject design is insightful for learning more about the conflict between people's fairness ideal and their payoff maximization concerns. Choices made behind the veil of ignorance can be interpreted as a measure of subjects' underlying preferences over payoff distribution, while choices with known rank confront these preferences with a reality principle. The within-subject design serves to determine how people switch or stick to their own underlying preferences, even if this may be costly to them.

## 2.2 The Choices

The binary choices belong to two categories: 4 Constant Efficiency Choices and 3 Pareto Comparable Choices, i.e., where one distribution (A) Pareto-dominates the other (B). Within each group of payoff distributions, subjects face the same type of tradeoff, but the numbers are slightly modified so as to test the robustness of the choices to marginal changes in the distributions.

---

<sup>3</sup>The subjects are informed step-by-step about what comes next in the experiment. Hence, they are not instructed when exactly they will do this task.

<sup>4</sup>We ask them basic questions regarding the payoff each member of the group would obtain depending on the hypothetical choices they could make (see Figure C5).

<sup>5</sup>See the end of Section 2.2 for an explanation of the sanity check. We did not include Choice 4 and Choice 8 in Part 2 (known rank situation) because we used these choices for non-incentivized hypothetical decisions that were taken after all incentivized decisions were made. Those questions were intended to study how subjects trade off their own rank and inequality (Choice 4) or efficiency (Choice 8). Subjects could choose to be either Person 3 (5) in Project A or Person 5 (1) in Project B for Choices 4 (8). These choices could not be incentivized because they were not aligned with the real effort task ranking and are, hence, not included in the main text. You can find the results of these hypothetical decisions in Appendix 1.B.

Note that, throughout this chapter, Project B is always the project with the lowest degree of inequality and it is the same throughout Choices 1, 2, 5, 6 and 7.<sup>6</sup>

Table 1.1 displays the Constant Efficiency Choices. The sum of the payoffs is constant across Project A and Project B, but Project A involves a higher degree of inequality than Project B. Hence, choosing Project B over Project A favors bottom players, to the detriment of top players.<sup>7</sup> The pairs of choices differ by the degree of inequality (e.g. the difference in standard deviation between Project A and B is higher for Choice 1 as compared to Choice 2 and 3), as well as by the rank affected by the choice (e.g. top ranked players are not directly affected by the decision in Choice 3 but their payoff does vary in the other choices). Furthermore, this generates within-subject variation in the tradeoffs subjects face in the second part of the experiment. A player ranked third faces no monetary tradeoff with inequality minimization in Choice 1 but does in Choice 2.

Table 1.1: Constant Efficiency Choices

	<b>Choice 1</b>		<b>Choice 2</b>		<b>Choice 3</b>		<b>Choice 4</b>	
	<b>A</b>	<b>B</b>	<b>A</b>	<b>B</b>	<b>A</b>	<b>B</b>	<b>A</b>	<b>B</b>
Person 1	1400	1000	1300	1000	1000	1000	1400	1100
Person 2	900	800	850	800	800	800	1200	800
Person 3	600	600	800	600	600	500	600	700
Person 4	300	500	200	500	350	300	350	650
Person 5	100	400	150	400	150	300	300	600
<i>Information below not shown to the subjects</i>								
Total	3300	3300	3300	3300	2900	2900	3850	3850
Std. dev.	513	241	484	241	340	311	502	199
GINI	0.39	0.18	0.36	0.18	0.29	0.21	0.32	0.12

Table 1.2 displays the Pareto Comparable Choices, where Project A always Pareto-dominates Project B and presents a more unequal income distribution. Within this category of choices, we vary the ranks that benefit from choosing Project A. For example, in Choice 7 all but the bottom-ranked player benefit from Project A, while only the first-ranked player benefits in Choice 6. We study Pareto dominance, as we are interested in documenting whether subjects are willing to reduce inequalities when this implies burning money without making anyone better off. This creates a situation that isolates envy from a maximin motive.

<sup>6</sup>Choices appear in random order, as does the letter of the project. For some subjects, what is presented as Project A, on the left-hand side, as the first choice, was presented to others as Project B, on the right-hand side, for a later choice.

<sup>7</sup>The screenshot of a choice, as presented in the laboratory, can be seen in Figure C1.

Table 1.2: Pareto Comparable Choices. Project A is Pareto-Dominant

	Choice 5		Choice 6		Choice 7	
	A	B	A	B	A	B
Person 1	1400	1000	3000	1000	1200	1000
Person 2	900	800	800	800	1000	800
Person 3	600	600	600	600	800	600
Person 4	500	500	500	500	700	500
Person 5	400	400	400	400	400	400
<i>Information below not shown to the subjects</i>						
Total	3800	3300	5300	3300	4100	3300
Std. Dev.	404	241	1095	241	303	241
Gini	0.23	0.18	0.52	0.18	0.19	0.18

Choice 8 is a sanity check, as it does not involve any kind of tradeoff: behind the veil of ignorance, everybody is weakly better off choosing A {1200; 1100; 1000; 900; 800} than B {800; 700; 600; 500; 400} and inequality is constant across both projects. Reassuringly, behind the veil of ignorance, 97% (309) of subjects actually choose Project A. We do not use the data from Choice 8 to produce the graphs and regressions since this choice does not involve any tradeoff.

All choices are incentivized by combining the random dictator approach with the random problem selection method. More precisely, for each part (behind the veil of ignorance, and decisions with known rank), subjects are told that the experimenter will randomly choose one person in each group of five subjects, and one out of the pairs of choices. The project that will have been chosen by that person in that round will become payoff-relevant for herself and for the other members of her group. We use the following exchange rate to convert experimental units into euros: 200 points = €1.

Before letting the subjects choose between the different projects, they do a training round. To make sure that all subjects understood the procedure, we get them to answer a short questionnaire. Figure C5 in the Appendix shows screenshots of the training choice and the questionnaire.

### 2.3 Merit versus Luck

Between subjects, 150 subjects were randomly allocated to the Luck treatment and 170 to the Merit treatment. In the Luck treatment, positions within a payoff distribution are based on random draws. In the Merit treatment, positions are determined by the relative performance of subjects in a task. The task was designed to elicit effort and not innate talent. It consists in typing as many five-letter strings as possible in 30 seconds. Members of the group are then ranked in decreasing order based on the number of words that they were able to type in 30 seconds. If two subjects type the same number of words, their ranking is based on the time of completion of the last string.<sup>8</sup> We chose this real effort task as it yields a quasi-continuous relative measure of effort that is used to break ties. Further, it mitigates concerns about procedural

<sup>8</sup>Before performing the real task, subjects carried out a 15-second training task.

fairness, as performance is not based on innate talent or other uncontrollable features. Given that the subject pool was relatively young, we believe that all subjects were equally able to use a computer. This task has previously been used by Jung et al. (2018) and is similar to other tasks where subjects perform repetitive activities in a given time interval, such as Dickinson (1999) (typing paragraphs), Erkal et al. (2019) (encryption), or Charness and Villeval (2009) (solving anagrams). In order to maintain the procedural balance between treatments, we also have subjects in the Luck treatment perform the task. Needless to say, we instruct them that the task will have no impact on their payoff or on the rest of the experiment.

Within the Merit treatment, before making choices behind the veil of ignorance, subjects are informed that the ranking within their group will be determined by their relative performance in a task. However, we do not inform them about the nature of the task. Letting them know such information *ex-ante* would tear the veil of ignorance somewhat, as subjects would be able to form expectations about their performance. Nevertheless, we ask subjects to predict the position they expect to achieve after the task (before it is described) in order to obtain a subjective measure of self-confidence that we will analyze in the robustness check section. To avoid hedging, this prediction is not incentivized. A screenshot of the real effort task as presented in the laboratory can be found in Appendix 1.C and is displayed in Figure C4. Concerning the Luck treatment, subjects are informed before they make choices behind the veil of ignorance that their positions will be determined randomly within their group.

## 2.4 Control variables

The binary choices between two payoff distributions can also be interpreted as risky bets. De facto, risk aversion and inequality aversion are two closely related measures (e.g. Harsanyi, 1953). In order to disentangle the two motives, we elicit risk aversion using two methods. First, we use an incentivized elicitation method introduced by Eckel and Grossman (2002) (henceforth, the Eckel-Grossman method). This method requires subjects to pick one out of six lotteries. The expected values and variance (riskiness) of the lotteries are jointly increasing.<sup>9</sup> Second, we let subjects choose between two different lotteries, depicted in Table A6 (Appendix 1.A.1), which correspond to the same payoffs as Choice 1. Subjects have the choice between a relatively safe lottery (Lottery B) and a relatively risky lottery (Lottery A). The expected value is the same for both lotteries. A payoff is drawn randomly from the chosen lottery. For both lotteries, the probability of getting any of the five payoffs is 0.2. As the lottery has the same values as Choice 1, a person who does not have any social preferences should make the same decision when faced with the choice behind the veil of ignorance and this lottery. Conversely, any difference between a subject's choice in the lottery setting and in the group-payoff setting is likely to denote a pure preference for certain payoff distributions. We elicit risk aversion *before* the subjects choose behind the veil of ignorance. The subjects are informed about the result of the risk-aversion tasks at the *end* of the experiment.<sup>10</sup>

---

<sup>9</sup>The sixth lottery is an exception. Here, the variance is increasing compared to the fifth one, but not its expected value.

<sup>10</sup>Additionally, we elicited beliefs about other's behavior for 75 subjects *after* they went through the main parts of the experiment. See beginning of Appendix 1.D for more details.

Table A1 in the Appendix displays the socio-economic characteristics of the subject pool elicited through a post-experimental questionnaire. As usual in lab experiments, a large proportion of the sample consists of students (67% of the sample) and the participants are relatively young (25 years old on average). The table also shows that the Merit and Luck groups are indistinguishable along various observed characteristics.

## 2.5 Implementation

Participants entered the laboratory and randomly drew place cards that assigned them to a computer in the laboratory where they found consent forms. After every participant got seated, the instructions were read out loud orally to establish common knowledge of the instructions. After the instructions had been read out loud, subjects read the instructions themselves on the computer screen. This procedure was identical for all parts of the experiment.

The experiment was computerized and coded in C#. All sessions were conducted in the Experimental Economics Laboratory of Paris (LEEP). 18 sessions took place between March and May 2017 and 4 in February 2018, with a total of 320 subjects. Participants were recruited using the online recruitment system ORSEE (Greiner, 2015), in sessions of 20, 15 or 10 people. The variations in the number of subjects per session was due to unforeseeable differences in participation rates across sessions. We always invited more than 20 subjects but we could not predict how many people would actually show up. We had to constitute groups of exactly 5 individuals. When the total number of subjects was not a multiple of 5, we paid the last-arriving superfluous participants a show-up fee of €7 and explained that they could not participate in the experiment.<sup>11</sup> All the subjects participating to the experiment received a fixed participation fee of €3.50 on top of the variable payoff they would obtain from the experimental games. Average earnings (including the participation fee) were €16 and the experiment lasted on average 33 minutes.

## 3 Results

We present the results for each category of choices (Constant Efficiency and Pareto Comparable ones) along two dimensions: choices behind the veil of ignorance versus known position, and Merit versus Luck treatments.

### 3.1 Choices behind the veil of ignorance

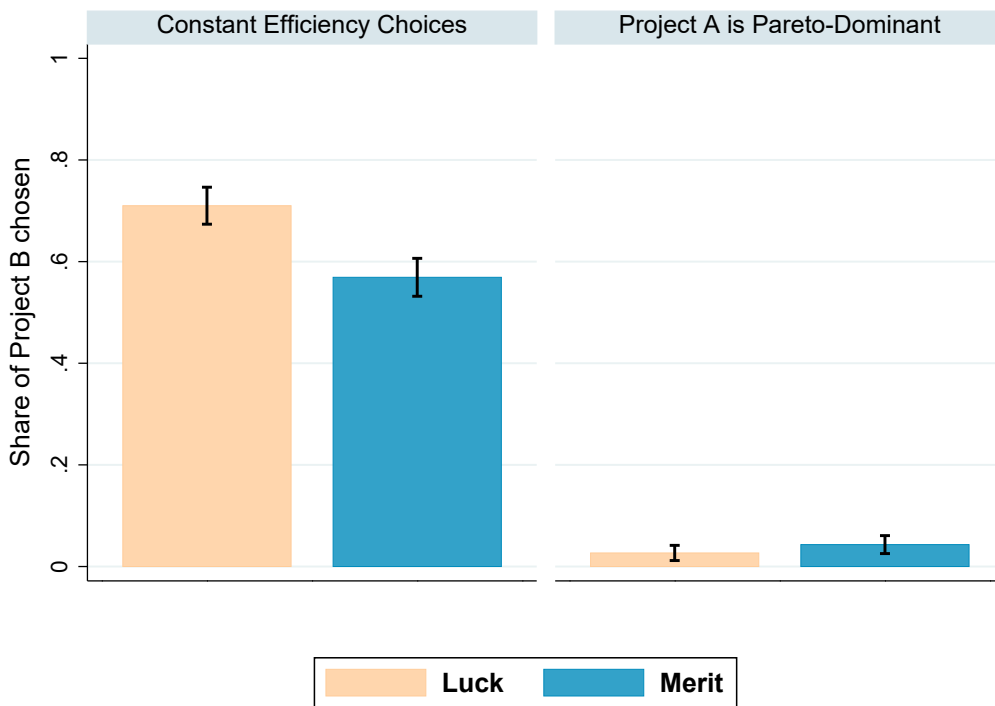
Pareto-dominance turns out to be the most important criterion when subjects make their choice behind the veil of ignorance. Figure 1.1 displays the pooled results of the choices between Project A (high inequality project) and Project B (low inequality project). Observations are at the choice-subject level: there are 4 choices per subject in the Constant Efficiency Choices category (left panel) and 3 choices per subject in the Pareto Comparable Choices category (right panel).

---

<sup>11</sup>This show-up fee was necessary to comply with the lab rules.

Regarding Constant Efficiency Choices, behind the veil of ignorance, we observe a high degree of heterogeneity. On average across both treatments, in 63% of cases subjects prefer the low-inequality Project B. But subjects are less likely to choose Project B in the Merit treatment than in the Luck treatment: 57% of observations choose Project B in the Merit treatment against 71% in the Luck treatment (a Wilcoxon-Mann-Whitney-Test rejects equality in means at the 1% level). This pattern of a 30/70 split for the Luck Treatment and a 40/60 split for the Merit Treatment across Projects A and B is fairly robust across all the Constant Efficiency Choices, as shown in Figure A3. This result is in line with the literature on the Merit versus Luck hypothesis. However, this is the only time in the entire experiment that the Luck and Merit treatments lead to different behaviors. In all the other contexts under consideration, i.e., Pareto Comparable Choices and choices with known rank, both the Luck and the Merit treatment groups behave similarly.

Figure 1.1: Choices Behind the Veil of Ignorance by Treatment



*Notes:* The bars show the share of low-inequality Project B chosen across the types of choices (Constant Efficiency Choices on the left and Choices where Project A is Pareto-Dominant on the right) and the treatment groups (Luck or Merit), with 95% confidence intervals. On the left panel, the p-value for the difference in means across treatment groups using a Wilcoxon-Mann-Whitney-Test is 0.00, while it is 0.17 for the right panel.

Moving to the right panel of Figure 1.1, which pools the results across the three choices where A is Pareto-dominant, we can see that when the more unequal distribution of Project A makes everyone at least weakly better off compared to Project B, then Project A is chosen almost unanimously and nearly reaches a consensus. Surprisingly, the origin of inequality no longer matters: the results are not statistically different across the Luck and Merit treatments (the p-value for the difference in means across treatment groups using a Wilcoxon-Mann-Whitney-Test is 0.17).



Here again, the results are not driven by one particular choice. The distribution of decisions for all Pareto Comparable Choices is displayed in Figure A4. Specifically, Choice 6, with its large top payoff of 3000 units, is not driving the results. In all of the 3 binary choices in this category, at least 95% favor the more unequal Project A. Hence, behind the veil of ignorance, subjects do not seem to be bothered about top income inequality: Pareto dominance is key.

These results are confirmed in a Logit regression (columns 1-3 in Table 1.3) pooling all of the 7 choices and displaying the marginal effects. We regress a dummy variable equal to 1 if Project B (low inequality) is chosen on the variables describing the context of the choice: a dummy equal to 1 if the subject is under the Merit treatment and 0 under the Luck treatment, and a dummy for the category of the choice (Constant Efficiency type of choice). The latter is what matters most: moving from a Pareto Comparable Choice to a Constant Efficiency type of choice increases the probability of choosing the low-inequality Project B by 56.6 percentage points on average (column 1). The origin of the distribution matters, but much less: being in the Merit treatment reduces the probability of choosing Project B by 7.3 percentage points, but this only holds for Constant Efficiency Choices: when the interaction term between the Merit dummy and the Constant Efficiency dummy is introduced in the regression, it attracts a negative coefficient but kills the statistical significance of the Merit main effect (columns 2 and 3).

### 3.2 Choices with known position

Thus far, we have shown that, when choices are made behind the veil of ignorance, it is possible to reach a consensus, and the main condition is Pareto-dominance. But what happens when subjects know their own positions in the distribution and have to make the same choices? Figure 1.2 shows how subjects choose across projects for the Constant Efficiency Choices depending on the tradeoffs they face. The left panel displays the pooled results when subjects attained a sufficiently high rank to be strictly better off by choosing the high-inequality Project A. The middle panel shows the choice of subjects whose earnings are the same in Projects A and B. Finally, the right panel shows how subjects react when they are worse off in the high-inequality Project A, which corresponds to situations where the player achieved a fairly low rank.<sup>12</sup>

As one might expect, choices are largely driven by own-payoff maximization. Project A is selected in more than 80% of cases by subjects whose payoff is higher with this choice, but it is chosen in less than 3% by subjects who stand to lose by selecting it. When subjects face the same payoff in Project A and Project B, their choices are very similar to what happens behind the veil of ignorance, with the important difference that here, choices do not depend on the Luck versus Merit treatment.

Nevertheless, not all players are selfish payoff maximizers: almost 20% of them choose the more equal Project B that favors the bottom-ranked individual, even though this implies giving up a higher payoff for themselves. The willingness to sacrifice one's own income in order to decrease inequality is not significantly related to Luck versus Merit treatments (see Table A3).

Instead of pooling the decisions by tradeoff type, one can also look at decisions for each choice

---

<sup>12</sup>Note that the tradeoff a player faces differs across choices; e.g. top-ranked players face a tradeoff in Choices 1 and 2 but not in Choice 3.

Table 1.3: Drivers of Inequality Aversion (Choice of Project B) - Logit Regressions on Pooled Data

	Veil of ignorance			Known position		
	(1)	(2)	(3)	(4)	(5)	(6)
<i>Dep. Var. = 1 if Project B (low inequality) is chosen</i>						
Merit	-0.0732*** (0.0227)	0.0718 (0.0646)	0.0721 (0.0594)	-0.00601 (0.0202)	0.0260 (0.0339)	0.0303 (0.0335)
Constant Efficiency	0.566*** (0.0194)	0.648*** (0.0474)	0.634*** (0.0430)	0.173*** (0.0185)	0.203*** (0.0300)	0.203*** (0.0296)
Merit * Constant Efficiency		-0.161** (0.0695)	-0.156** (0.0650)		-0.0562 (0.0422)	-0.0574 (0.0418)
Risk aversion (Eckel-Grossman)			0.0308*** (0.00663)			0.000991 (0.00606)
(Payoff B - Payoff A)/Payoff A				0.526*** (0.112)	0.525*** (0.112)	0.527*** (0.112)
Female dummy			0.0657*** (0.0206)			0.0205 (0.0201)
Higher education dummy			-0.0481** (0.0234)			-0.0312 (0.0261)
Age			0.000693 (0.00217)			0.000756 (0.00347)
Background in economics			-0.109*** (0.0207)			-0.0112 (0.0204)
Currently employed			-0.0366 (0.0301)			-0.0261 (0.0319)
Currently in a relationship			0.0480** (0.0224)			-0.00366 (0.0209)
Average probability to choose B	0.33	0.33	0.33	0.35	0.35	0.35
Week fixed effects	No	No	Yes	No	No	Yes
Session size effects	No	No	Yes	No	No	Yes
Observations	2240	2240	2240	1920	1920	1920
Pseudo $R^2$	0.343	0.346	0.395	0.356	0.357	0.361

Robust standard errors clustered at the individual level in parentheses; \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

*Notes:* The coefficients are marginal effects estimated using a logit model on pooled data. Hence, a coefficient equal to 0.1 means that a marginal change in the explanatory variable increases the probability of choosing Project B by 10 percentage points compared to the baseline. The unit of observation is the individual-choice level. *Merit* is a dummy variable equal to 1 if the subject is in the Merit treatment and 0 for the Luck treatment. *Constant Efficiency* equals 1 in the case of Constant Efficiency Choices (Choices 1 to 4) and 0 in the case of Pareto Comparable Choices (Choices 5 to 7). *Merit\*Constant Efficiency* is an interaction term between the two previous variables. *(Payoff B - Payoff A)/Payoff A* corresponds to the difference in payoff resulting from choosing Project B rather than Project A given the subject's position in the distribution. Risk aversion is elicited using the Eckel-Grossman method; a higher value in the Eckel-Grossman task is equivalent to an increase in risk aversion. Columns 1 to 3 pool decisions made behind the veil of ignorance and columns 4 to 6 pool decisions made with known position within the income distribution. The control variables include age, and dummy indicators for: female, background in economics, employed, currently being in a relationship, holding a higher degree. We also include "week of the session" dummies (the experiment took place in 4 different weeks) and "session size" dummies (the experiment had 10,15 or 20 participants).

Figure 1.2: Decisions with Known Position - Constant Efficiency Choices



*Notes:* Pooling over Constant Efficiency Choices, the bars show the share of low-inequality Project B chosen across the situation of the subject and the treatment groups (Luck or Merit), with 95% confidence intervals. On the left panel, we show the choices made by sufficiently high-ranked subjects so that they are better off with the high-inequality project A (the p-value for the difference in means across treatment groups using a Wilcoxon-Mann-Whitney-Test is 0.137 in that case). On the middle panel, we display the choices made by subjects that will receive the same payoff across Projects A and B (the p-value of the difference in means across treatments is 0.767). The right panel shows the results for subjects with a low rank such that they are worse off in the high-inequality project A (the p-value for the difference in means across treatments is 0.535).

(Figures A5-A7 in the appendix for Constant Efficiency Choices). The results are similar. More precisely, regarding players that are better off with the low-inequality Project B<sup>13</sup>, the share of them choosing Project B is never below 80% for each Choice. Conversely, subjects systematically choose Project B less often whenever they are ranked high enough to benefit from Project A<sup>14</sup>.

Figure 1.3 displays the pooled results of the Pareto Comparable Choices where Project A is Pareto-dominant and subjects know their position. Unsurprisingly, subjects still quasi-unanimously favor the high-inequality project when it is to their advantage, on the left panel (subjects have a sufficiently high rank to be better off with Project A). However, on the right panel, when their gain is the same in Projects A and B (subjects who achieved a lower ranking), about 23% of subjects act as money burners. They burn money at the top by choosing Project B even if it does not improve the lot of the low earners. Since Pareto Comparable choices ensure that the payoffs of the bottom players are constant across Projects, we can infer that money burners are not motivated by maximin concerns.

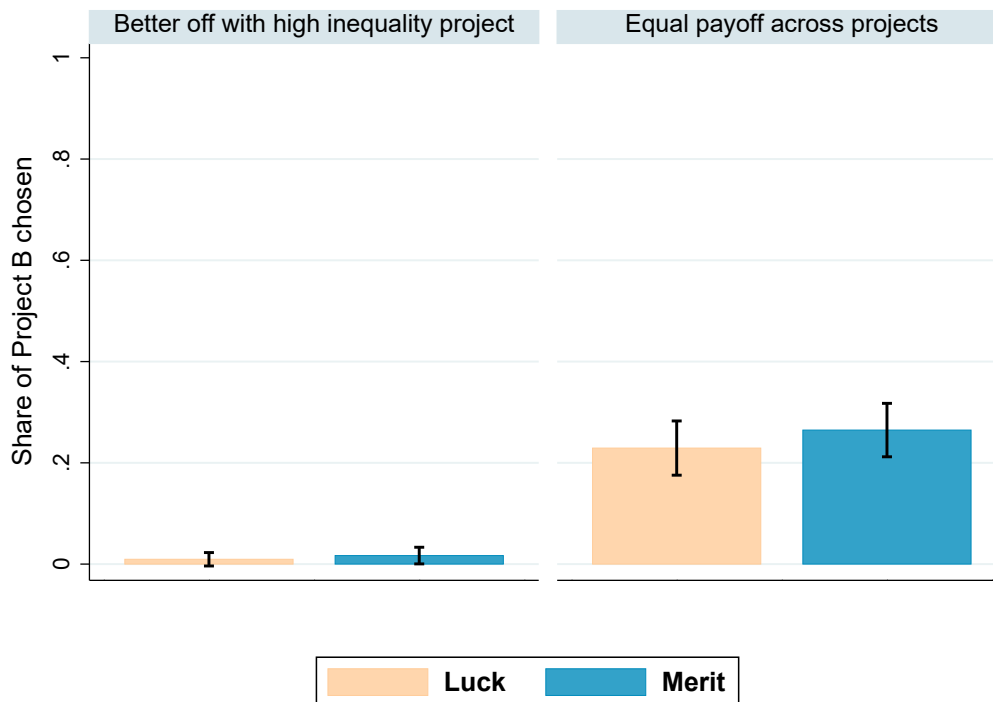
Figures A8-A10 display the distribution of choices made by subjects depending on their rank, for Pareto-comparable choices. As before, the position subjects hold in the distribution does not make a systematic and meaningful difference. Similarly, one can see that the propensity to choose a Pareto-dominant distribution is relatively invariant to the exact shape of the distribution.

The last 3 columns of Table 1.3 confirm these results. The only two significant predictors of

<sup>13</sup>Players 4 and 5 in Choices 1 and 2, and Player 5 in Choice 3

<sup>14</sup>Players 1 and 2 in Choice 1, players 1, 2 and 3 in Choice 2

Figure 1.3: Decisions with Known Position - Project A is Pareto-Dominant



*Notes:* Pooling over Pareto Comparable Choices, the bars show the share of low-inequality Project B chosen across the situation of the subject and the treatment groups (Luck or Merit), with 95% confidence intervals. On the left panel, we show the choices made by sufficiently high-ranked subjects so that they are better off with the high-inequality project A (the p-value for the difference in means across treatment groups using a Wilcoxon-Mann-Whitney-Test is 0.504 in that case). On the right panel, we display the choices made by subjects that will receive the same payoff across Projects A and B (the p-value for the difference in means across treatments is 0.353). There is no situation for which a subject is worse off in the high inequality project A, since A is Pareto-dominant.

choice are whether A is Pareto-dominant and the variation in personal payoffs between Project B and Project A. We should emphasize that these results come in spite of the high saliency of the effort task. Indeed, subjects went through the task right before making the choices with known positions. Finally, the choices are not significantly dependent on the subjects' rank after controlling for all variables included in Table 1.3, such as the difference in payoffs. In that case, subjects choose Project A (or B) with the same likelihood no matter what their rank is (see e.g. Tables A3 and A2).

### 3.3 Observable heterogeneity

Table 1.3 reports the marginal effects of observable characteristics such as gender or having a background in economics. We find that women are less likely to choose the high-inequality project, as compared to men, when choosing behind the veil of ignorance (after controlling for risk aversion). Note that this only holds for Constant Efficiency choices, as there is virtually no variation in the Pareto Comparable choices. Furthermore, having a background in economics and holding a tertiary degree decrease the probability of opting for the more equal project.

These significant effects, including gender, disappear as soon as subjects know their position. This finding is in line with the result in Ben-Ner et al. (2004), who show, using dictator games,

that gender only matters if the gender of the recipient is known. We are, however, not able to replicate the findings of Sharma (2015) that men tend to keep more of the total surplus than women in a dictator game with a preceding production phase. Note that women are no more likely than men to sacrifice their own income in order to decrease inequality, nor to engage in money burning (see Tables A2 and A3).

## 4 Within-subject variation

In this section, we exploit our within-subject design to study how people react when they end up in a position of conflict between the motive that drove their initial choice (behind the veil of ignorance) and own-payoff maximization concerns. Focusing on the top left panel of Figure 1.2, we want to know what motivated subjects to choose Project B over Project A: why do they decide to reduce their own income? Looking at their choices behind the veil of ignorance enables us to discard explanations based on irrationality or cognitive fatigue. It turns out that in 85% of cases, these subjects also chose the low-inequality Project B behind the veil of ignorance. The behavior of these subjects suggests that they hold strong normative preferences for egalitarianism and stick to them despite the loss in income that can be associated with it.

At the other end of the spectrum, we note that individuals who are identified as money burners (subjects choosing Project B in the Pareto-comparable choices once their rank is revealed) almost unanimously chose the high-inequality Project A behind the veil of ignorance. This position reversal may result from a form of disappointment, as these subjects were probably hoping to reach a higher payoff rank. Interpreted in a broader context, this result echoes a dynamic situation where individuals initially hope to climb up the meritocratic social ladder, but in the end realize that social positions are largely fixed and that upward mobility is not an option. Our results indicate that this situation is a source of frustration for a sizable share of the subjects, as shown in a recent working paper by Gangadharan et al. (2018).

## 5 Robustness Checks

### 5.1 Are subjects consistent within a category of choices?

Our main results are based on average choice frequencies, but it might be the case that subjects do not choose consistently throughout the experiment. We want to determine whether subjects are actually consistent within a category of choices, despite the fact that choices were displayed randomly.

We start by looking at the choices made behind the veil of ignorance (Figures A3 and A4). Obviously, subjects choosing within the category of Pareto Comparable Choices are highly consistent since there is quasi-unanimity in all choices (about 92% of subjects always select the same project within this category of choice). What about Constant Efficiency Choices behind the veil of ignorance? If we consider as consistent subjects who choose the same Project (A or B) on each of the four occasions, then we identify that 50.3% of subjects prefer the same type of

project. If we define as consistent subjects who choose the same Project (A or B) on 3 occasions out of 4, then 85% stick to the same category.

Turning to behavior with known rank (Figures A5 to A10), we observe that only 41% of subjects always stick to the same Project (A or B) within the Constant Efficiency Choice category. This is explained by the payoff maximization behavior described in Section 3.2. Concerning Pareto Comparable Choices with known positions, 78% of subjects always choose the same project. We have shown above that a significant number of subjects burn money at the top of the distribution if they do not stand to gain from the unequal distribution. 88% of subjects that burn money at least once in the experiment do so every time they have the opportunity.

In summary, subjects behave in a fairly consistent way in the sense that they follow one distributive principle within each category of choices.

## **5.2 Is choosing behind the veil of ignorance the same as choosing between two lotteries?**

If individuals do not have any distributive preferences, the choice between two income distributions is equivalent to a choice between two lotteries. Choices should then be interpreted as reflecting risk aversion rather than social preferences. Prior literature suggests that this is not the case (Schildberg-Hörisch, 2010; Johansson-Stenman et al., 2002). To test this hypothesis we asked subjects, in the risk elicitation part of the experiment, to choose between two lotteries (Table A7) that are payoff-equivalent to Choice 1. Each lottery has five payoffs that are equally likely to be drawn. The only difference with Choice 1 is that each person chooses a lottery for herself only and her decision does not affect any other subjects. Figure A2 depicts the share of choices attracted by each lottery by treatment group and compares it with the decisions made by each treatment group in Choice 1. It turns out that the Luck treatment group opts more for the equal project B (71.3%) than for the safer lottery (64.7%), although both choices are payoff-equivalent. Conversely, subjects in the Merit treatment group are slightly less likely to opt for Project B (54.1%) than for the safe lottery (59.4%). These differences cast doubt on the hypothesis that inequality aversion is indistinguishable from risk aversion.

To test further whether choices behind the veil of ignorance are completely explained by risk aversion, we regress Choice 1 on the choice between the two lotteries and other covariates. As shown in Table A4, choosing the safe lottery increases the likelihood of choosing project B by 27 percentage points (the overall share of subjects who chose B is 62%). However, the inclusion of this predictor does not entirely explain all of the choice variations, and other covariates such as Luck or Merit treatment or having a background in economics, remain statistically significant.

## **5.3 Does the significant treatment effect pick up on over-confidence?**

The influence of the Merit treatment in the Constant Efficiency Choices could be driven by subjects who are over-confident, and who thus choose Project A because they believe that they will perform well at the task and achieve a top position. This would imply that they choose Project A for self-regarding motives instead of fairness motives. This is unlikely given that subjects have no information about the nature of the task when they make their choices. Nevertheless,

in order to capture people’s expectations, we asked subjects to estimate the position they would achieve after the task. It turns out that most subjects were relatively optimistic: 97.06% of those in the Merit treatment predicted that they would at least achieve third position, which implies that they would be weakly better off in Project A in the Constant Efficiency Choices. However, only 45.9% of subjects actually chose Project A. If self-confidence were really driving the results, a much higher share of subjects should choose Project A.

Furthermore, overconfidence would imply that predicted positions are a strong predictor of actual choices within the Merit treatment.<sup>15</sup> Table A5 tests this hypothesis. We include both predicted position and rank to control jointly for confidence and over-confidence. If confidence is driving the result, we should observe a correlation between future position and choice; if overconfidence is driving the results, we should see a relationship between predicted position and choice that persists after controlling for rank. We do find that individuals who predict themselves to be ranked third are significantly more likely to choose Project B in Choice 3 only, but this runs counter to the overconfidence hypothesis because they would actually lose part of their own payoff if they chose Project B and were actually ranked third. There is no significant correlation for all the other choices. We summarize this finding by pooling the data over Choices 1, 2 and 4. None of the “Predicts rank” coefficients are statistically significant in that case.<sup>16</sup>

## 6 Conclusion

This experiment studies in a systematic way the importance of three main institutional settings that can affect individuals’ preferences regarding the way wages are distributed within a firm. We vary the shape of the income distribution, the uncertainty regarding people’s own rank and the origin of inequalities (Merit vs Luck). We shed light on the conditions under which unanimity over payoff distributions in small groups can be reached and when it breaks. It turns out that Pareto-dominant distributions are likely to reach a near-consensus, even if they come with higher inequality. But this is true only to the extent that choices are made behind the veil of ignorance. Once positions are fixed and known ex-ante, a non-negligible proportion of individuals engage in money burning at the top of the distribution. Our within-subject design illustrates the frustration generated by fixed ranks, as opposed to open positions.

Despite the salience of the effort task in our experiment, the relevance of the Merit versus Luck hypothesis appears to be rather weak in our context. This is at odds with many experiments (Hoffman et al., 1994; Cherry et al., 2002; Fong, 2001; Oxoby and Spraggon, 2008; Krawczyk, 2010), but is consistent with several recent studies that focus on taxation and follow a structure that is similar to ours (Durante et al., 2014; Bjerck, 2016). We do not make the claim that the merit criterion does not matter, but our experiment shows that other aspects may crowd it out. Moving from the context of a small working group to the level of society as a whole, our findings suggest that people might tolerate inequality, independently of its cause (whether luck or merit), if it came as a joint product of income growth for everyone - a condition that is far from being met in the context of the early 21st century (Alvaredo et al., 2018).

---

<sup>15</sup>We did not ask subjects in the Luck treatment about their predicted position, in order to avoid confusion.

<sup>16</sup>Choice by choice regressions do not yield statistically significant results (regressions not show)

## 7 References

- Alesina, Alberto and George-Marios Angeletos. Fairness and redistribution: US vs. Europe. *American Economic Review*, 95(5):960–980, 2005.
- Alesina, Alberto and Eliana La Ferrara. Preferences for redistribution in the land of opportunities. *Journal of Public Economics*, 89(5-6):897–931, 2005.
- Alós-Ferrer, Carlos, Jaume García-Segarra, and Alexander Ritschel. Performance curiosity. *Journal of Economic Psychology*, 64:1–17, 2018.
- Alvaredo, Facundo, Lucas Chancel, Thomas Piketty, Emmanuel Saez, and Gabriel Zucman. World Inequality Report 2018. *World Inequality Lab*, 2018.
- Beckman, Steven R, John P Formby, W James Smith, and Buhong Zheng. Envy , malice and Pareto efficiency : An experimental examination. *Social Choice and Welfare*, 19(2):349–367, 2002.
- Ben-Ner, Avner, Fanmin Kong, and Louis Putterman. Share and share alike? Gender-pairing, personality, and cognitive ability as determinants of giving. *Journal of Economic Psychology*, 25(5):581–589, 2004.
- Bjerk, David. In front of and behind the veil of ignorance: An analysis of motivations for redistribution. *Social Choice and Welfare*, 47(4):791–824, 2016.
- Bolton, Gary E and Axel Ockenfels. Inequality Aversion, Efficiency, and Maximin Preferences in Simple Distribution Experiments: Comment. *American Economic Review*, 96(5):1906–1911, 2006.
- Breza, Emily, Supreet Kaur, and Yogita Shamdasani. The morale effects of pay inequality. *The Quarterly Journal of Economics*, 133(2):611–663, 2018.
- Cappelen, Alexander W, Astri Drange Hole, Erik Ø Sørensen, and Bertil Tungodden. The pluralism of fairness ideals: An experimental approach. *The American Economic Review*, 97(3):818–827, 2007.
- Card, David, Alexandre Mas, Enrico Moretti, and Emmanuel Saez. Inequality at work: The effect of peer salaries on job satisfaction. *The American Economic Review*, 102(6):2981–3003, 2012.
- Charness, G. and M. Rabin. Understanding Social Preferences with Simple Tests. *The Quarterly Journal of Economics*, 117(3):817–869, 2002.
- Charness, Gary and Brit Grosskopf. Relative payoffs and happiness: An experimental study. *Journal of Economic Behavior & Organization*, 45(3):301–328, 2001.
- Charness, Gary and Marie Claire Villeval. Cooperation and competition in intergenerational experiments in the field and the laboratory. *American Economic Review*, 99(2):956–978, 2009.



- Cherry, Todd L, Peter Frykblom, and Jason F Shogren. Hardnose the Dictator. *The American Economic Review*, 92(4):1218–1221, 2002.
- Dengler-Roscher, Kathrin, Natalia Montinari, Marian Panganiban, Matteo Ploner, and Benedikt Werner. On the malleability of fairness ideals: Spillover effects in partial and impartial allocation tasks. *Journal of Economic Psychology*, 65:60–74, 2018.
- Dickinson, David L. An Experimental Examination of Labor Supply and Work Intensities. *Journal of Labor Economics*, 17(4):638–670, 1999.
- Durante, Ruben, Louis Putterman, and Joël van der Weele. Preferences for Redistribution and Perception of Fairness: An Experimental Study. *Journal of the European Economic Association*, 12(4):1059–1086, 2014.
- Eckel, Catherine C and Philip J. Grossman. Sex differences and statistical stereotyping in attitudes toward financial risk. *Evolution and human behavior*, 23(4):281–295, 2002.
- Engelmann, Dirk and Martin Strobel. Inequality aversion, efficiency, and maximin preferences in simple distribution experiments. *The American Economic Review*, 94(4):857–869, 2004.
- Engelmann, Dirk and Martin Strobel. Preferences over income distributions experimental evidence. *Public Finance Review*, 35(2):285–310, 2007.
- Erkal, Nisvan, Lata Gangadharan, and Boon Han Koh. By chance or by choice? Biased attribution of others’ outcomes. *Working Paper*, 2019.
- Fehr, Ernst and Klaus M Schmidt. A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, 114(3):817–868, 1999.
- Fong, Christina. Social preferences, self-interest, and the demand for redistribution. *Journal of Public Economics*, 82(2):225–246, 2001.
- Frohlich, Norman, Joe A Oppenheimer, and Cheryl L Eavey. Laboratory Results on Rawls’ s Distributive Justice. *British Journal of Political Science*, 17(1):1–21, 1987.
- Gangadharan, Lata, Philip J. Grossman, and Joe Vecci. Moving on up: The impact of income mobility on antisocial behaviour. *Working Paper*, 2018.
- Greiner, Ben. Subject Pool Recruitment Procedures: Organizing Experiments with ORSEE. *Journal of the Economic Science Association*, 1(1):114–125, 2015.
- Harsanyi, John. Cardinal utility in welfare economics and in the theory of risk-taking. *Journal of Political Economy*, 61(5):434–435, 1953.
- Hoffman, Elizabeth, Kevin McCabe, Keith Shachat, and Vernon Smith. Preferences, property rights, and anonymity in bargaining games. *Games and Economic behavior*, 7(3):346–380, 1994.

- Johansson-Stenman, Olof, Fredrik Carlsson, and Dinky Daruvala. Measuring Future Grandparents' Preferences for Equality and Relative Standing. *The Economic Journal*, 112(479): 362–383, 2002.
- Jung, Se Eun, Chung Choe, and Ronald L. Oaxaca. Gender wage gaps and risky vs. secure employment: An experimental analysis. *Labour Economics*, 52:112–121, 2018.
- Kamas, Linda and Anne Preston. Distributive and reciprocal fairness: What can we learn from the heterogeneity of social preferences? *Journal of Economic Psychology*, 33(3):538–553, 2012.
- Konow, James. Fair Shares: Accountability and Cognitive Dissonance in Allocation Decisions. *American Economic Review*, 90(4):1072–1091, 2000.
- Konow, James. Which Is the Fairest One of All? A Positive Analysis of Justice Theories. *Journal of Economic Literature*, 41(4):1188–1239, 2003.
- Krawczyk, Michal. A glimpse through the veil of ignorance: Equality of opportunity and support for redistribution. *Journal of Public Economics*, 94(1):131–141, 2010.
- Kritikos, Alexander and Friedel Bolle. Distributional concerns: Equity-or efficiency-oriented? *Economics Letters*, 73(3):333–338, 2001.
- Lefgren, Lars J., David P. Sims, and Olga B. Stoddard. Effort, luck, and voting for redistribution. *Journal of Public Economics*, 143:89–97, 2016.
- Michelbach, Philip A, John T Scott, Richard E Matland, and Brian H Bornstein. Doing Rawls justice: An experimental study of income distribution norms. *American Journal of Political Science*, 47(3):523–539, 2003.
- Overlaet, Bert. Merit criteria as justification for differences in earnings. *Journal of Economic Psychology*, 12(4):689–706, 1991.
- Oxoby, Robert J and John Spraggon. Mine and yours: Property rights in dictator games. *Journal of Economic Behavior & Organization*, 65(3):703–713, 2008.
- Ruffle, Bradley J. More is better, but fair is fair: Tipping in dictator and ultimatum games. *Games and Economic Behavior*, 23(2):247–265, 1998.
- Schildberg-Hörisch, Hannah. Is the veil of ignorance only a concept about risk? An experiment. *Journal of Public Economics*, 94(11):1062–1066, 2010.
- Sharma, Smriti. Gender and distributional preferences: Experimental evidence from India. *Journal of Economic Psychology*, 50:113–123, 2015.
- Yang, Yang, Sander Onderstal, and Arthur Schram. Inequity aversion revisited. *Journal of Economic Psychology*, 54:1–16, 2016.

# Appendices

## 1.A Tables and Figures

### 1.A.1 Tables

Table A1: Balance Table

	(1)	(2)	(3)	(4)	(5)
	Luck	Merit	Overall	Diff.	p-value t-test difference in means
Age	24.88	25.38	25.15	-0.502	0.425
Secondary Degree	0.713	0.771	0.744	-0.057	0.243
Female	0.533	0.541	0.537	-0.008	0.889
Employed	0.173	0.212	0.194	-0.038	0.387
Student	0.687	0.647	0.666	0.040	0.455
Economics background	0.413	0.406	0.409	0.007	0.893
In a relationship	0.313	0.365	0.341	-0.051	0.335
Political Orientation	3.811	3.532	3.656	0.278	0.335
Risk loving (6 lotteries)	3.693	3.735	3.716	-0.042	0.816
Risk loving (2 lotteries)	0.647	0.594	0.619	0.053	0.336
Subjective risk measure	6.193	6.229	6.213	-0.036	0.883
<i>N</i>	150	170	320	320	

*Notes:* This table presents the summary statistics of individual characteristics of the whole sample. *Treated* is an indicator variable equal to 1 if the subject was in a group where performance rather than luck determined rank. *Secondary degree* is an indicator equal to 1 if the subject has a higher degree than the French Baccalaureate. *Employed* is an indicator variable equal to 1 if the subject is currently employed and not self-employed or completing studies. *Student* is an indicator equal to 1 if the subject is currently a student. *Economics background* is an indicator variable equal to 1 if the subject has an academic background in economics or a related subject; i.e. she is either an economics student or has studied it in the past. *In a relationship* is an indicator equal to 1 if the subject is currently in a civil relationship (the subject pool did not contain a married subject). *Political orientation* is a variable ranging from 0-10. 0 indicates that a subject identifies very much as being left wing and 10 indicates that a subject identifies as being very right wing. Risk loving (6 lotteries) corresponds to the choices made in the Eckel-Grossman task. A score closer to 1 means higher risk aversion. Risk loving (2 lotteries) is 0 if the subject is risk averse and 1 otherwise. The subjective risk measure goes from 0 to 10 where 0 means extreme reported risk aversion. The precise questions asked are presented in Appendix 1.D.

Table A2: Who are the money burners?

	(1)	(2)	(3)	(4)	(5)	(6)
<i>Dep. Var. = 1 if subject chooses to burn money</i>						
Merit	0.0455 (0.0522)	0.0465 (0.0523)			0.0442 (0.0580)	0.0529 (0.0545)
Risk aversion (Eckel-Grossman)	0.0115 (0.0152)	0.0110 (0.0151)	0.0171 (0.0234)	0.0180 (0.0231)	0.0143 (0.0174)	0.00735 (0.0165)
Age	-0.00209 (0.00817)	-0.00252 (0.00825)	-0.0135 (0.00824)	-0.0136 (0.00872)	0.000596 (0.00661)	-0.00597 (0.00520)
Female dummy	0.0546 (0.0551)	0.0528 (0.0550)	0.00259 (0.0818)	-0.00599 (0.0819)	0.0740 (0.0613)	0.0760 (0.0577)
Currently employed	0.0327 (0.0531)	0.0347 (0.0528)	0.0704 (0.0759)	0.0700 (0.0758)	0.0408 (0.0675)	0.0682 (0.0616)
Background in economics	-0.0261 (0.0585)	-0.0244 (0.0586)	0.0104 (0.0785)	0.0130 (0.0784)	-0.0507 (0.0642)	-0.0754 (0.0608)
Currently in a relationship	0.0302 (0.0816)	0.0306 (0.0816)	0.113 (0.1000)	0.114 (0.100)	-0.0130 (0.0829)	0.0195 (0.0782)
Higher education dummy	-0.0835 (0.0691)	-0.0830 (0.0690)	0.0455 (0.104)	0.0424 (0.105)	-0.119 (0.0822)	-0.0855 (0.0701)
Rank		0.0168 (0.0239)				
Predicted Position			-0.0294 (0.0429)			
Rank - pred. pos.				-0.00441 (0.0306)		
Political Orientation					0.0417*** (0.0139)	0.0131 (0.0140)
Voted extreme right						0.608** (0.239)
Mean probability to burn money	0.25	0.25	0.26	0.26	0.25	0.25
<i>N</i>	512	512	272	272	385	385
Pseudo- <i>R</i> <sup>2</sup>	0.015	0.017	0.026	0.024	0.069	0.142

Standard errors clustered at the individual level in parentheses. \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

*Notes:* The coefficients are marginal effects estimated using a logit model on pooled decisions where subjects had the opportunity to burn money (decisions included in the right panel of Figure 1.3). These decisions were only made in part 2 of the experiment (known rank). Hence, a coefficient equal to 0.1 means that a marginal change in the explanatory variable increases the probability of choosing Project B by 10 percentage points compared to the baseline. The unit of observation is the individual-choice level. *Merit* is a dummy variable equal to 1 if the subject is in the Merit treatment and 0 for the Luck treatment. Risk aversion is elicited using the Eckel-Grossman method; a higher value in the Eckel-Grossman task is equivalent to an increase in risk aversion. The control variables include age, and dummy indicators for: female, background in economics, employed, currently being in a relationship, holding a higher degree, rank in the group, predicted rank, the difference between rank and predicted rank, political orientation (0 left wing, 10 right wing), voted for the far-right party in the last election.

Table A3: Who are the people that are willing to pay for more equality?

	(1)	(2)	(3)	(4)	(5)	(6)
<i>Dep. Var. = 1 if subject chooses to reduce inequality at her own expense</i>						
Merit	-0.0761 (0.0423)	-0.0762 (0.0416)	-0.0759 (0.0424)	-0.0764 (0.0417)	-0.0649 (0.0438)	-0.0633 (0.0437)
Risk aversion (Eckel-Grossman)	-0.00981 (0.0129)	-0.00738 (0.0128)	-0.0109 (0.0127)	-0.00844 (0.0126)	-0.00313 (0.0136)	-0.00319 (0.0135)
Age	0.00474 (0.00647)	0.00438 (0.00601)	0.00510 (0.00656)	0.00480 (0.00604)	0.00464 (0.00678)	0.00489 (0.00670)
Female dummy	0.0132 (0.0419)	0.0128 (0.0414)	0.0138 (0.0419)	0.0143 (0.0414)	0.0106 (0.0447)	0.0118 (0.0448)
Background in economics	-0.0608 (0.0458)	-0.0606 (0.0456)	-0.0556 (0.0466)	-0.0512 (0.0465)	-0.0248 (0.0486)	-0.0230 (0.0492)
Currently employed	-0.160** (0.0728)	-0.170** (0.0717)	-0.159** (0.0722)	-0.170** (0.0710)	-0.233*** (0.0876)	-0.234*** (0.0882)
Currently in a relationship	-0.0410 (0.0482)	-0.0346 (0.0478)	-0.0406 (0.0479)	-0.0319 (0.0478)	-0.00222 (0.0481)	-0.00538 (0.0492)
Higher education dummy	-0.00812 (0.0520)	-0.00978 (0.0506)	-0.00198 (0.0539)	0.000466 (0.0521)	-0.0262 (0.0507)	-0.0293 (0.0508)
(Payoff B - Payoff A)/Payoff A		0.486 (0.280)		0.581** (0.272)	0.633** (0.288)	0.643** (0.291)
Rank			-0.0184 (0.0208)	-0.0318 (0.0232)		
Political Orientation					-0.0379*** (0.0127)	-0.0401*** (0.0127)
Voted extreme right						0.110 (0.141)
Mean prob. to pay for equality	0.17	0.17	0.17	0.17	0.16	0.16
<i>N</i>	448	448	448	448	361	361
Pseudo- <i>R</i> <sup>2</sup>	0.045	0.056	0.048	0.063	0.125	0.126

Standard errors clustered at the individual level in parentheses. \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

*Notes:* The coefficients are marginal effects estimated using a logit model on pooled decisions where subjects had the opportunity to reduce inequality at a cost to themselves when knowing their rank *and* choose Project B in constant efficiency choices behind the veil of ignorance. Hence, a coefficient equal to 0.1 means that a marginal change in the explanatory variable increases the probability of choosing Project B by 10 percentage points compared to the baseline. The unit of observation is the individual-choice level. *Merit* is a dummy variable equal to 1 if the subject is in the Merit treatment and 0 for the Luck treatment. Risk aversion is elicited using the Eckel-Grossman method; a higher value in the Eckel-Grossman task is equivalent to an increase in risk aversion.  $(\text{Payoff } B - \text{Payoff } A)/\text{Payoff } A$  is the relative difference in payoff resulting from choosing Project B rather than Project A given the subject's position in the distribution. Other variables include age, and dummy indicators for: female, background in economics, employed, currently being in a relationship, holding a higher degree, rank in the group, predicted rank, political orientation (0 left wing, 10 right wing), voted for the far-right party in the last election.

Table A4: Is choice of Project B *only* explained by risk?

	(1)	(2)
<i>Dep. Var. = 1 if Project B (low inequality) of Choice 1 is chosen</i>		
Safe lottery is chosen	0.256*** (0.0447)	0.218*** (0.0451)
Merit		-0.163*** (0.0493)
Female dummy		0.0553 (0.0506)
Higher education dummy		-0.104* (0.0618)
Age		0.000128 (0.00502)
Background in economics		-0.166*** (0.0478)
Currently Employed		-0.0929 (0.0660)
Currently in a relationship		0.0985* (0.0540)
Mean Probability to choose B	0.62	0.62
Week FE	No	Yes
Session size FE	No	Yes
<i>N</i>	320	320
Pseudo- <i>R</i> <sup>2</sup>	0.058	0.135

Robust standard errors clustered at the session level in parentheses;  
 \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

*Notes:* The coefficients are marginal effects estimated using a logit model where the dependent variable is a dummy variable equal to 1 if the subject chose B in Choice 1 behind the veil of ignorance. Hence, a coefficient equal to 0.1 means that a marginal change in the explanatory variable increases the probability of choosing Project B by 10 percentage points compared to the baseline. The unit of observation is the individual level. *Safe lottery is chosen* equals 1 if the lottery equivalent to Project B is chosen. *Merit* is a dummy variable equal to 1 if the subject is in the Merit treatment and 0 for the Luck treatment. The control variables include age, and dummy indicators for: female, background in economics, employed, being in a relationship, holding a higher degree. We also include “week of the session” dummies (the experiment took place in 4 different weeks) and “session size” dummies (the experiment had 10,15 or 20 participants).

Table A5: Confidence and Over-Confidence do not Predict Choices

	Constant Efficiency (excl. Choice 3)			Choice 3		
	(1)	(2)	(3)	(4)	(5)	(6)
<i>Dep. variable = 1 if Project B (low inequality) is chosen</i>						
Predicts rank 2	0.0692 (0.0711)		0.0746 (0.0718)	0.000800 (0.0798)		0.00733 (0.0797)
Predicts rank 3	0.0523 (0.0744)		0.0629 (0.0734)	0.348*** (0.0936)		0.363*** (0.0904)
Predicts rank 4	-0.0578 (0.165)		-0.0609 (0.163)	0.103 (0.174)		0.107 (0.179)
Rank		-0.0215 (0.0203)	-0.0240 (0.0202)		-0.0208 (0.0246)	-0.0318 (0.0241)
Risk aversion (Eckel-Grossman)	0.0552*** (0.0182)	0.0575*** (0.0175)	0.0554*** (0.0178)	0.0492** (0.0205)	0.0538** (0.0216)	0.0488** (0.0199)
Female dummy	0.129** (0.0530)	0.141*** (0.0533)	0.129** (0.0531)	0.0906 (0.0694)	0.130* (0.0685)	0.0896 (0.0694)
Higher education dummy	-0.0971 (0.0692)	-0.0886 (0.0694)	-0.0874 (0.0702)	-0.0454 (0.0813)	-0.0282 (0.0834)	-0.0344 (0.0814)
Age	-0.00251 (0.00773)	-0.000655 (0.00778)	-0.000202 (0.00808)	-0.0136 (0.00890)	-0.0113 (0.00950)	-0.0106 (0.00951)
Background in economics	-0.196*** (0.0550)	-0.193*** (0.0531)	-0.194*** (0.0551)	-0.176*** (0.0616)	-0.214*** (0.0653)	-0.173*** (0.0617)
Currently employed	-0.0638 (0.0911)	-0.0884 (0.0915)	-0.0745 (0.0920)	0.124 (0.101)	0.115 (0.114)	0.108 (0.103)
Currently in a relationship	0.133** (0.0563)	0.135** (0.0559)	0.132** (0.0559)	0.0703 (0.0683)	0.0669 (0.0723)	0.0689 (0.0679)
Mean prob. to pay for equality	0.61	0.61	0.61	0.56	0.56	0.56
Week FE	Yes	Yes	Yes	Yes	Yes	Yes
Session size FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	510	510	510	170	170	170
Pseudo- $R^2$	0.118	0.117	0.121	0.210	0.142	0.218

Robust standard errors clustered at the individual level in parentheses (columns 1-3); robust standard errors in parentheses (columns 4-6); \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

*Notes:* The coefficients were estimated using a logit model on the subjects in the Merit treatment. Hence, a coefficient equal to 0.1 means that a marginal change in the explanatory variable increases the probability of choosing Project B by 10 percentage points compared to the baseline. *Rank* denotes the position in a group that the subject will achieve after the effort task. A larger value means a lower gain; *Predicts rank #* are dummies indicating the self-reported rank the subjects expect to attain. The omitted category is expecting first rank (highest gain). The dependent variable is the choice made by subjects behind the veil of ignorance, coded 1 if subject chooses Project B and 0 if subjects chose Project A. The unit of observation is the individual-choice level within a category of choices. Regressions in columns 1 to 3 pool individual choices for Choices 1,2 and 4. There are thus 3 observations per treated individual in these regressions. Columns 4 to 6 only consider Choice 3; hence, there is one observation per individual. The control variables include risk aversion (Eckel-Grossman), age, and dummies for: female, background in economics, employed, being in a relationship, and holding a higher degree. We also include “week of the session” dummies (the experiment took place in 4 different weeks) and “session size” dummies (the experiment had 10,15 or 20 participants). Results are similar without controls and without week and session size fixed effects.

**Eliciting risk aversion**

Table A6: Choice table to elicit risk aversion using Eckel-Grossman method (Lottery 1)

<b>Lottery</b>	<b>Low gain</b>	<b>High gain</b>	<b>Mean</b>	<b>St. Dev.</b>	$r$	<b>Choice share</b>
<b>Lottery 1</b>	140	140	140	0	$3.46 < r$	10.00%
<b>Lottery 2</b>	120	180	150	30	$1.16 < r < 3.46$	11.25%
<b>Lottery 3</b>	100	220	160	60	$0.72 < r < 1.16$	31.88%
<b>Lottery 4</b>	80	260	170	90	$0.5 < r < 0.72$	11.56%
<b>Lottery 5</b>	60	300	180	120	$0 < r < 0.5$	14.69%
<b>Lottery 6</b>	10	350	180	170	$r < 0$	20.63%

*Notes:* The second and third columns show the possible gains for each lottery. The probability of each gain being drawn is 0.5 in all lotteries. The sixth column displays the implied range of the coefficient of relative risk aversion denoted as  $r$  assuming a CRRA utility function ( $u(x) = x^{1-r}$ ). The probability of each payoff being chosen stands at 50% for all lotteries. The last column shows the percentage of subjects that chose this lottery.

Table A7: Choice between two lotteries

	<b>Lottery A</b>	<b>Lottery B</b>
	1400	1000
	900	800
	600	600
	300	500
	100	400
Total	3300	3300
Standard deviation	512.84	240.83
GINI	0.39	0.18
Percent of choices	38.13%	61.88%
if luck	35.33%	64.67%
if merit	40.59%	59.41%

*Notes:* The columns show the possible payoffs of each lottery. The probability of any payoff in each lottery is 0.2. Subjects choose between Lotteries A and B.



1.A.2 Figures

Figure A1: Distribution of the Choices Made for Lottery 1

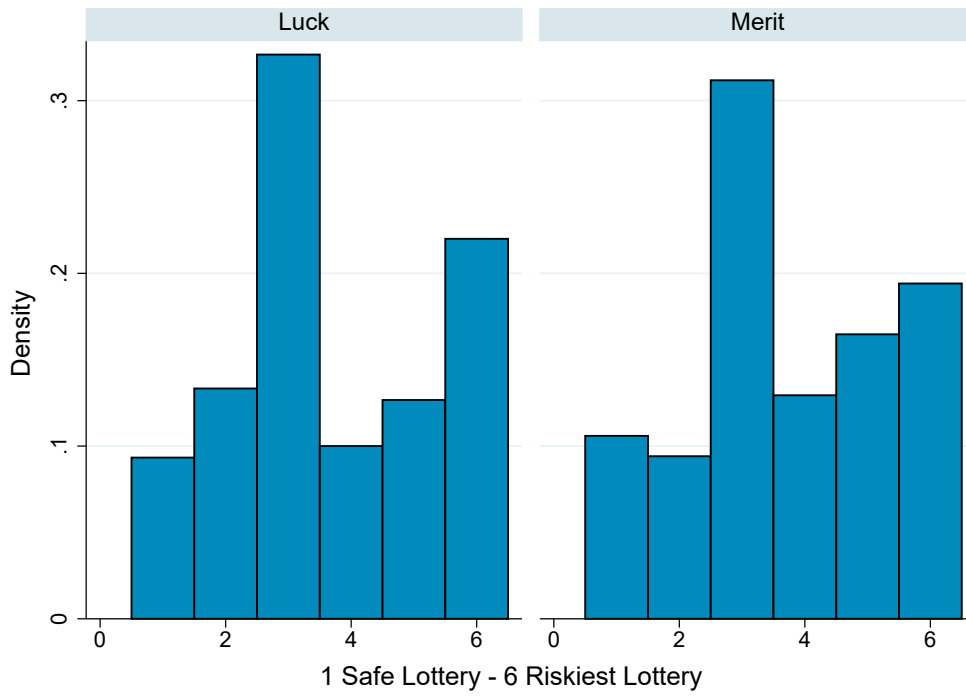
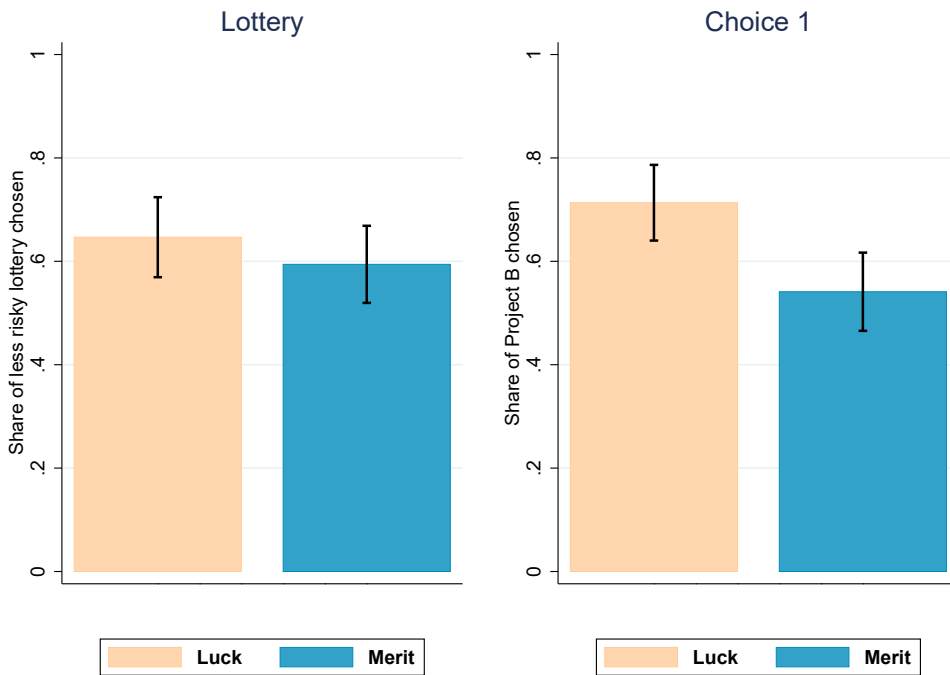
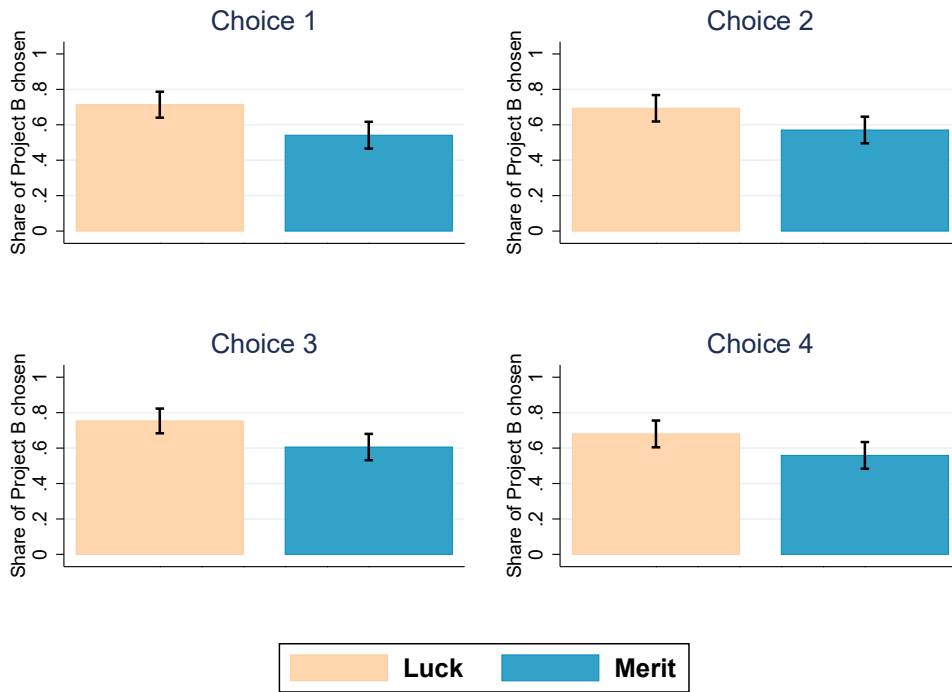


Figure A2: Differences between Choice 1 and the payoff-equivalent lottery



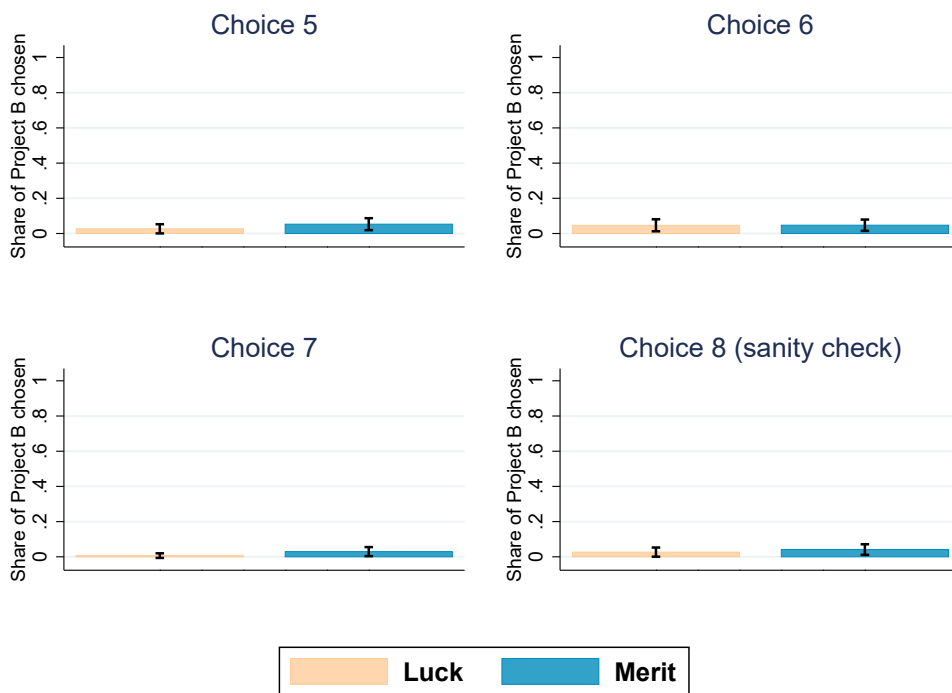
Notes: The lottery and Choice 1 are payoff-equivalent. In the lottery setting, subjects choose between two lotteries that have 5 equally likely payoffs. These payoffs are identical to the ones of Project A and Project B in Choice 1. We display 95% confidence intervals.

Figure A3: Distribution of the Choices Made Behind the Veil of Ignorance - **Constant Efficiency Choices**



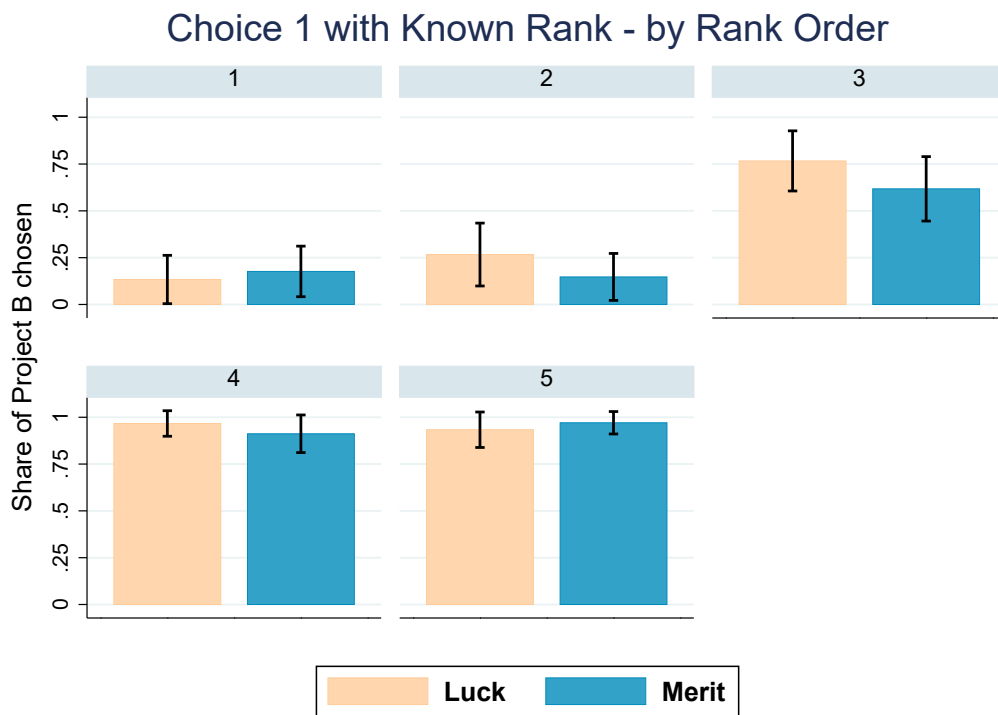
Notes: We display 95% confidence intervals.

Figure A4: Distribution of the Choices Made Behind the Veil of Ignorance - **A Pareto-Dominant project**



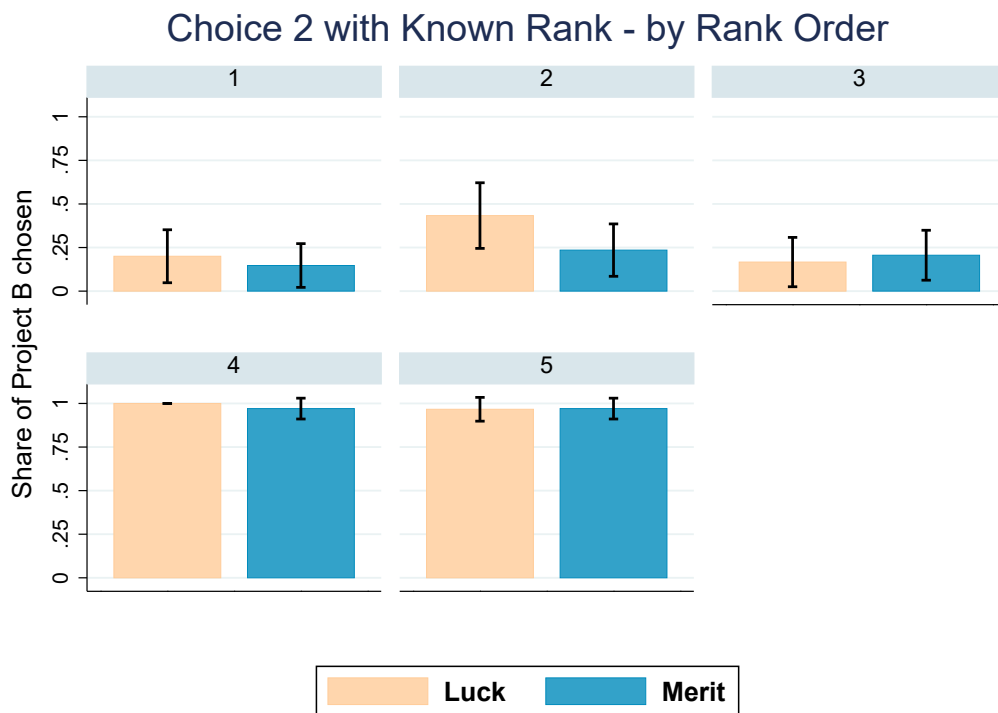
Notes: We display 95% confidence intervals.

Figure A5: Choice 1 (Constant Efficiency Choice) with Known Rank - by Rank Order



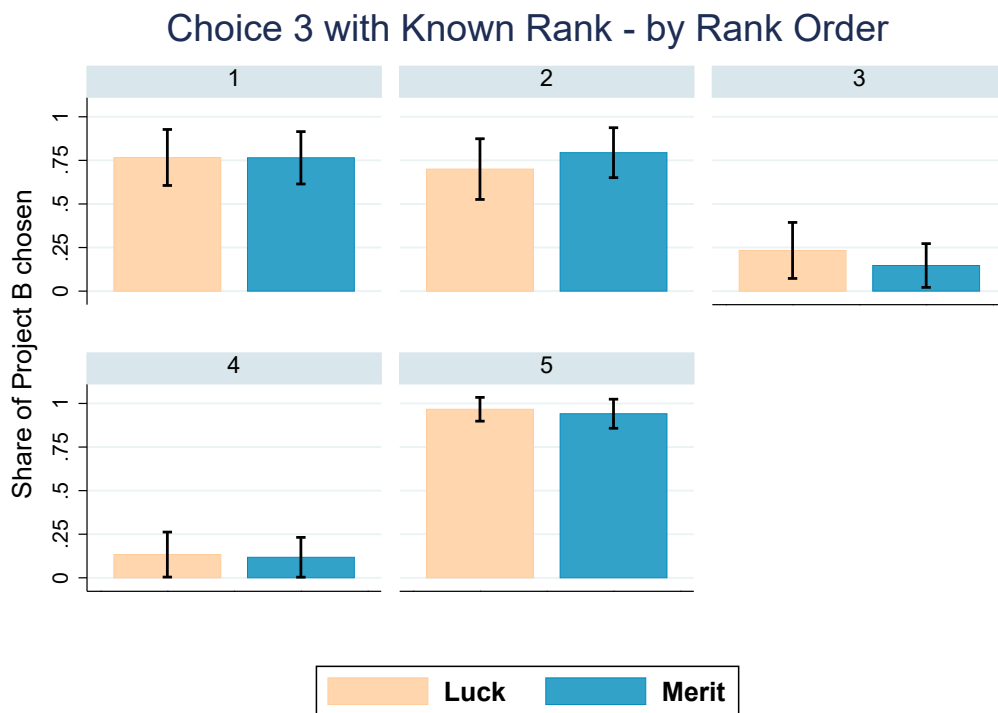
Notes: The graph shows the share of Project B chosen in Choice 1 depending on Treatment groups (Luck vs Merit) and the rank that the subject achieved after the real effort task, with 95% confidence intervals. The top left panel shows the choices made by players that achieved the top rank and the bottom right panel aggregates the choices made by the bottom-ranked players.

Figure A6: Choice 2 (Constant Efficiency Choice) with Known Rank - by Rank Order



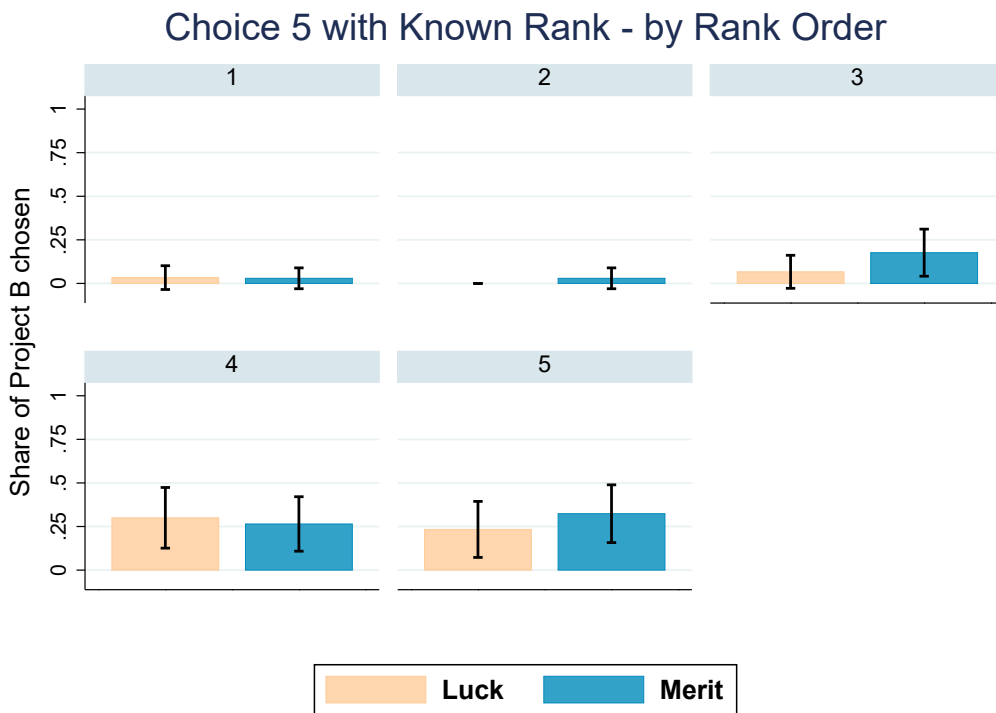
Notes: The graph shows the share of Project B chosen in Choice 2 depending on Treatment groups (Luck vs Merit) and the rank that the subject achieved after the real effort task, with 95% confidence intervals. The top left panel shows the choices made by players that achieved the top rank and the bottom right panel aggregates the choices made by the bottom-ranked players.

Figure A7: Choice 3 (Constant Efficiency Choice) with Known Rank - by Rank Order



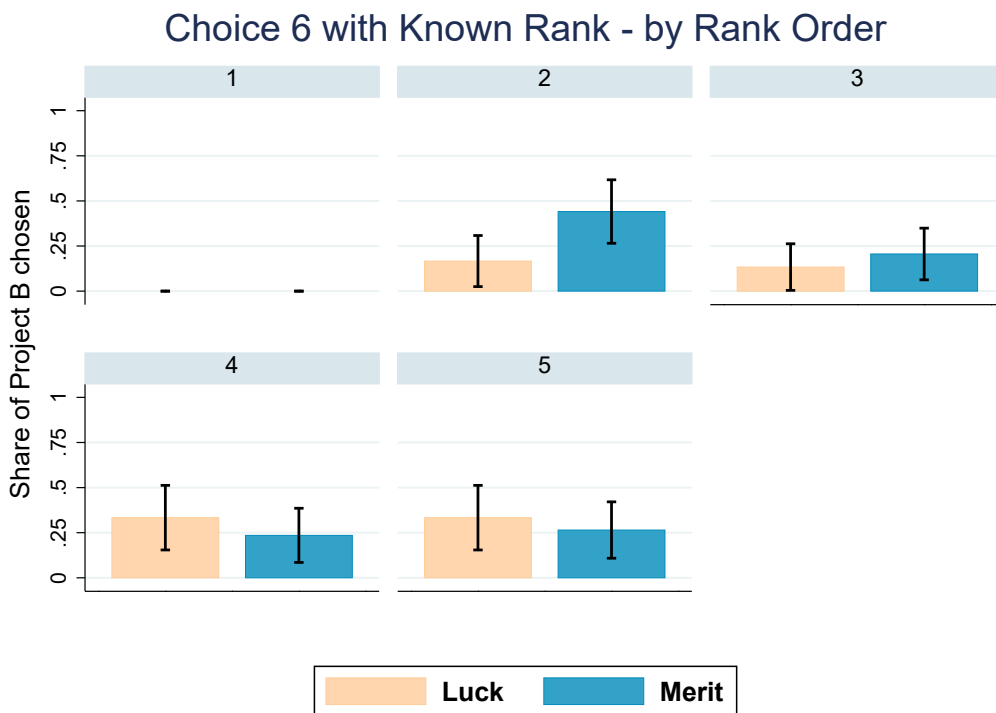
Notes: The graph shows the share of Project B chosen in Choice 3 depending on Treatment groups (Luck vs Merit) and the rank that the subject achieved after the real effort task, with 95% confidence intervals. The top left panel shows the choices made by players that achieved the top rank and the bottom right panel aggregates the choices made by the bottom-ranked players.

Figure A8: Choice 5 (A Pareto-Dominant) with Known Rank - by Rank Order



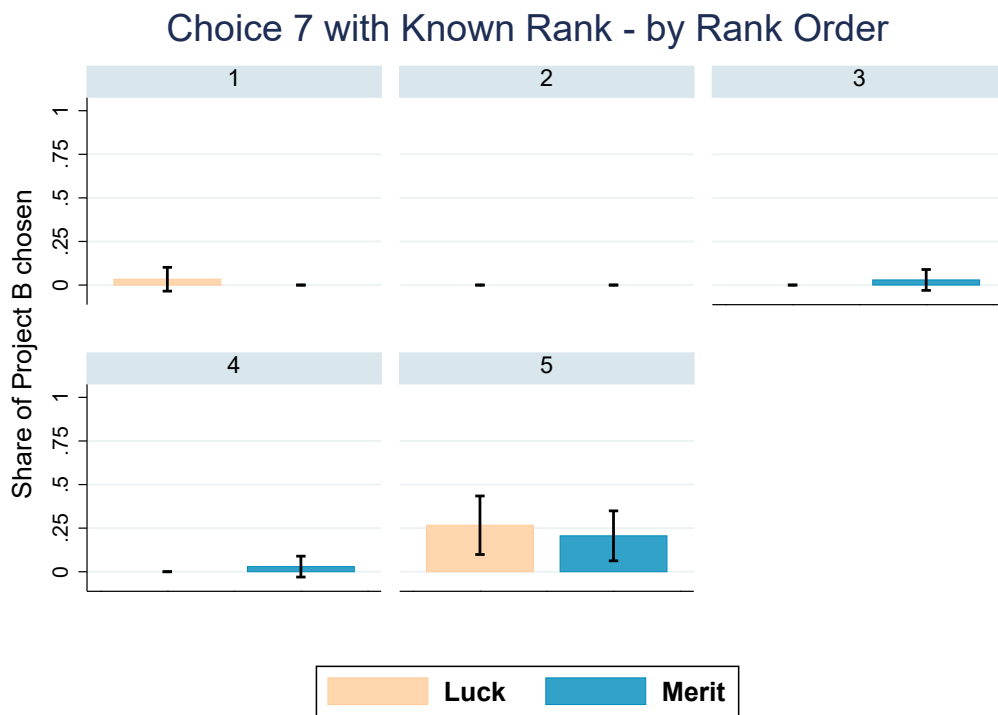
Notes: The graph shows the share of Project B chosen in Choice 5 depending on Treatment groups (Luck vs Merit) and the rank that the subject achieved after the real effort task, with 95% confidence intervals. The top left panel shows the choices made by players that achieved the top rank and the bottom right panel aggregates the choices made by the bottom-ranked players. Choice 4 was not used in the known rank part of the experiment.

Figure A9: Choice 6 (A Pareto-Dominant) with Known Rank - by Rank Order



Notes: The graph shows the share of Project B chosen in Choice 6 depending on Treatment groups (Luck vs Merit) and the rank that the subject achieved after the real effort task, with 95% confidence intervals. The top left panel shows the choices made by players that achieved the top rank and the bottom right panel aggregates the choices made by the bottom-ranked players.

Figure A10: Choice 7 (A Pareto-Dominant) with Known Rank - by Rank Order



Notes: The graph shows the share of Project B chosen in Choice 7 depending on Treatment groups (Luck vs Merit) and the rank that the subject achieved after the real effort task, with 95% confidence intervals. The top left panel shows the choices made by players that achieved the top rank and the bottom right panel aggregates the choices made by the bottom-ranked players.



## 1.B Hypothetical Decisions

Hypothetical choices are summarized in Table A8. In Choice 4, the subject is Person 3 in Project A, Person 5 in Project B and makes 600 in both cases. We find that 65% of subjects choose the low inequality Project B in this setting, despite implying a lower ranking. There is a slight difference between the Luck (70% choose B) and the Merit group (61% choose B) that is significant at the 10% level (p-value = 0.094).

Regarding Choice 8, the subject is Person 5 in Project A and Person 1 in Project B. 19% choose Project B in that case (without significant differences across treatment groups). A majority of people value more efficiency that benefit to everyone except them (Project A) than achieving a high rank (Project B). But when we compare to the results behind the veil of ignorance, virtually everyone choose Project A when positions are not known. So this result also indicates that about one fifth of subjects are money burners and are willing to make everyone worse off to achieve a higher rank.

Table A8: Hypothetical situations

	Choice 4		Choice 8	
	A	B	A	B
Person 1	1400	1100	1200	<b>800</b>
Person 2	1200	800	1100	700
Person 3	<b>600</b>	700	1000	600
Person 4	350	650	900	500
Person 5	300	<b>600</b>	<b>800</b>	400
<i>Information below not shown to subjects</i>				
Total	3850	3850	4100	3300
Std. dev.	502	199	303	241
GINI	0.32	0.12	0.19	0.18

*Notes:* Bold numbers correspond to the rank subjects are supposed to occupy once positions are revealed in the hypothetical situations.

## 1.C Screenshots of the Experiment

Figure C1: Screenshot of Choice Behind the Veil of Ignorance(Choice 2)

(a) Luck group

Partie 2 - Etape 1

**Faites un choix entre deux projets : A et B**

**Rappel :** Votre place dans le groupe sera tirée au sort.

Membres du groupe	Gains avec projet A	Gains avec projet B
Personne 1	1000	1300
Personne 2	800	850
Personne 3	600	800
Personne 4	500	200
Personne 5	400	150

*Appuyez pour afficher les instructions*

**Quel est votre choix ?**

Projet A     Projet B

Appuyez sur OK pour continuer :

(b) Merit group

Partie 2 - Etape 1

**Faites un choix entre deux projets : A et B**

**Rappel :** Votre position dans le groupe dépendra d'une tâche simple que vous effectuerez à la fin de l'expérience.

Membres du groupe	Gains avec projet A	Gains avec projet B
Personne 1	1300	1000
Personne 2	850	800
Personne 3	800	600
Personne 4	200	500
Personne 5	150	400

*Appuyez pour afficher les instructions*

**Quel est votre choix ?**

Projet A     Projet B

Appuyez sur OK pour continuer :

Figure C2: Screenshot of Risk Aversion Elicitation

(a) Eckel-Grossman Task

Partie 1

**Choisissez une loterie parmi les 6 loteries suivantes**

Vous avez une chance sur deux de tirer le petit gain et une chance sur deux de tirer le gros gain.

	Petit gain	Gros gain
<input type="radio"/> Loterie 1	140	140
<input type="radio"/> Loterie 2	120	180
<input type="radio"/> Loterie 3	100	220
<input type="radio"/> Loterie 4	80	260
<input type="radio"/> Loterie 5	60	300
<input type="radio"/> Loterie 6	10	350

*Appuyez pour afficher les instructions*

Appuyez sur OK pour continuer :

(b) Decision between two lotteries equivalent to Choice 1

Partie 1

**Faites un choix entre la loterie 1 et la loterie 2. Chaque nombre a la même chance d'être tiré.**

Gains avec loterie 1	Gains avec loterie 2
1400	1000
900	800
600	600
300	500
100	400

*Appuyez pour afficher les instructions*

**Choisissez une des deux loteries**

Loterie 1    Loterie 2

Appuyez sur OK pour continuer :

Figure C3: Screenshot of Choice with Known Position (Choice 6, Rank 2)

Partie 3 - Etape 6

Faites un choix entre deux projets : A et B, sachant que votre propre bonus est de **800** points dans le projet A et **800** points dans le projet B car vous êtes la personne **2**.

Membres du groupe	Gains avec projet A	Gains avec projet B
Personne 1	1000	3000
Personne 2	<b>800</b>	<b>800</b>
Personne 3	600	600
Personne 4	500	500
Personne 5	400	400

*Appuyez pour afficher les instructions*

**Quel est votre choix ?**

Projet A     Projet B

*Appuyez sur OK pour continuer :*

Figure C4: Screenshot of the Real Effort Task

Tâche : récopier les mots

kezch, lfata, njhsg, cclit, vqyrc, zcbiw, trcbh, mzoam, mqqbx, lblch, wjncu, zenpy, qbvbv, iuepp, hxlfx, grkle, skeqb, vxrbl, ujgrt, mhtmu, sugti, efwxi, pstvu, jgmjr, npxy, fgqhx, bbojx, zxavb, tqnad, kjwsf, khgas, uhufu, ezlzk, osnwd, jedcx, rjvuz, zyfxb, gnwuv, fezce, gndyq, zskgo, xjarg, oqrtq, loizw, poeho, mbhsy, jzkkx, dnyno, wmxnm, ezvrr, gpuoe, dwrhj, qarww, ubzql, oyofn, nrhye, roitk, wodgx, kicmx, apdta, qxsju, jrwe, zbrtu, yibct, waury, eyocc, iduhv, pbdun, zsvhu, encew, xsthf, twyez, cmlvn, bbnsy, txdwy, jehao, lhwx, xutee, gizjd, fyzo, cwiyv, jbygj, xdaxb, moswq, qzbnv, eysnv, soyyv, jzylm, tiwcb, pfgis, svghb, njlvf, yntqm, rwnbu, ucwysy, yjvd, vbxfm, sxcid, gztqq, iqbtn, ovkrw, svhtu, seqji, klzjl, wisyu, xboyx, sabgq, fneix, fygfj, zflv, ctasv, wtggr, vdvr, tnlgt, issmy, agfmd, gnrm, fzbxk, gibls, dwaej, akvnf, dizyk, gbbha, okhmf, rpoj, wcdnm, istxt, gixke, gwwfp, vnfew, wgysb, ntuzy, arqfg, sgkdq, lqmdr, wdvda, dngpx, pgnxf, lqim, gspth, vjbd, mxqdc, qdjyn, pxlov, snrfw, xsgvt, jtdda, heoos, fhefv, zgozd

Veuillez recopier les mots ci-dessus

Il vous reste 18 secondes

Figure C5: Comprehension test

(a) Screen 1

Partie 2 - Exemple

**Faites un choix entre deux projets : A et B**

Membres du groupe	Gains avec projet A	Gains avec projet B
Personne 1	20	16
Personne 2	17	13
Personne 3	11	11
Personne 4	4	8
Personne 5	1	5

*Appuyez pour afficher les instructions*

**Quel est votre choix ?**

Projet A     Projet B

Appuyez sur OK pour continuer :

(b) Screen 2

Partie 2 - Exemple

- Vous avez choisi le projet **B**
- Imaginez qu'un participant de votre groupe de 5 personnes ait été tiré au sort et qu'il ou elle ait choisi le projet A. Votre position a été tirée au sort, vous avez obtenu la 3ème position dans cette exemple.
- La position que vous obtiendrez réellement peut bien sûr être différente
- Vous aurez autant de chance de vous retrouver à chacune des 5 positions à l'issue du tirage au sort.

Membres du groupe	Gains avec projet A	Gains avec projet B
Personne 1	20	16
Personne 2	17	13
<b>Personne 3</b>	<b>11</b>	<b>11</b>
Personne 4	4	8
Personne 5	1	5

Le projet A a été sélectionné :

Quel est le montant de votre bonus?

Quel est le bonus de la personne qui a tiré au sort la première position ?

Quel est le bonus de la personne qui a tiré au sort la cinquième position ?

Appuyez sur OK pour continuer :

## (c) Screen 3

Partie 2 - Exemple

Membres du groupe	Gains avec projet A	Gains avec projet B
Personne 1	20	16
Personne 2	17	13
<b>Personne 3</b>	<b>11</b>	<b>11</b>
Personne 4	4	8
Personne 5	1	5

Quel est le montant de votre bonus? - Vous avez raison, votre bonus serait de 11 points.

Quel est le bonus de la personne qui a tiré au sort la première position? - Vous avez raison, votre bonus serait de 20 points.

Quel est le bonus de la personne qui a tiré au sort la cinquième position? - Vous n'avez pas raison, votre bonus serait de 1 point et non 3.

Appuyez sur OK pour continuer :

OK

## 1.D Post-experimental Questionnaire

After the subjects made all their decisions, we asked them questions about their age, occupation, political orientation and attitudes toward redistribution. We informed the subjects that their answers were strictly anonymous (as in the whole experiment). All the questions were asked in French. The French wording of the questions is available upon request.

Additionally, we confronted 75 subjects with four choices they made earlier (Choices 1 and 6, both behind the veil of ignorance and with known rank) and they could explain the reasons for their choice. For these choices, we also elicited the beliefs (incentivized) of the same 75 subjects about other players' behaviour. We decided to elicit subjects' beliefs about others' behaviour after receiving feedback on the first results of this experiment. Knowing subjects' beliefs could be helpful to elicit behavioral mechanisms of choices. Findings from the belief elicitation task suggest that subjects believe that other people will behave in a similar way as they do themselves. Such a false consensus effect has already been documented in the literature in similar settings (Charness and Grosskopf, 2001). Therefore, we decided not to include this result in the chapter. This difference in the post-experimental questionnaire across subjects could not affect the core results of this chapter since we did not inform the subjects that they will have to guess what other participants would do. They are only informed about the belief elicitation task after they have done their Project choices. The experiment is identical for everyone (with the Luck and Merit treatment variations) for all subjects up to the post-experimental questionnaire.

Before they started the questionnaire we asked the subjects to make two *hypothetical* choices. The distributions are identical to Choices 4 and the sanity check, with the difference that payoff is fixed but rank is varied. Subjects could thus choose to be ranked third (fifth) in Project A, or ranked fifth (first) in Project B. The results from these hypothetical decisions are available on request.

**Question 1 - Life satisfaction** Answered on a scale from 0 to 10.

- All things considered, how satisfied are you with your life as a whole these days?

**Question 2 - Risk taking** Answered on a scale from 0 (not at all willing to take risks) to 10 (very willing to take risks).

- Are you generally a person who is willing to take risks or do you try to avoid taking risks?

**Question 3 - Attitudes towards inequality** Answered on a scale from 0 (do not agree) to 10 (totally agree).

- Do you think that inequalities should be reduced?

**Question 4 - Attitudes towards the deservingness of income** Answered on a scale from 0 (do not agree) to 10 (totally agree).

- Do you think that the rich deserve their revenue?



**Question 5 - Attitudes towards poor peoples' effort** Answered on a scale from 0 (do not agree) to 10 (totally agree).

- Do you think that poor people do not make enough effort to improve their current situation?

**Question 6 - Political orientation** Answered on a scale from 0 (far left) to 10 (far right). People could say that they do not know, or that they do not want to answer this question.

- In political matters, people sometimes talk of “the left” and “the right”. Using this card, where would you place yourself on this scale, where 0 means the left and 10 means the right?

**Question 6 - Voting behavior**

- Which party do you feel closest to?
  - Parti Socialiste
  - Les Républicains
  - Front National
  - France Insoumise
  - Parti Communiste Français
  - Europe Ecologie – Les verts
  - En Marche
  - Autres : précisez
  - Ne sait pas
  - Ne se prononce pas

**Question 7 - Gender** Answered by opting between ‘man’ and ‘woman’

- Are you a man or a woman?

**Question 8 - Age** Answered by typing their age

- How old are you?

**Question 9 - Marital status** Answered by opting between ‘in a relationship’, ‘single’, ‘separated/divorced’ and ‘widowed’

- What is your marital status?

**Question 10 - Labor market status** Answered by opting between ‘employed’, ‘self-employed’, ‘homemaker’, ‘retired’, ‘unemployed’, ‘student’, ‘working student’ and ‘other’

- What is your current employment situation?

**Question 11 - Academic background** Answered by opting between disciplines with the option of entering one not on the list

- What discipline are/were you enrolled in as a student?

**Question 12 - Degree** Answered by opting between no degree, primary school, middle school (brevet des collèges), high school (baccalauréat) and university degree (enseignement supérieur)

- What is your highest diploma?

**Question 13 - Income** Answered by opting between different monthly and yearly income intervals (For monthly: less than €1100, €1101-1420, €1421-1715, €1716-2050, €2051-2450, €2451-2880, €2881-3400, €3401-4100, €4101-5300, more than €5301)

- What are your available resources (after taxes, all sources included and also that of your parents)? In which interval would you locate yourself?

## 1.E Experimental Instructions

The following section will include the experimental instructions. To establish common knowledge the instructions were read out loud and the subjects were explicitly asked to also read them on their computer screen.

Each title in the instructions symbolizes a new screen. For each part of the experiment, we will present a choice table to give the reader an idea of how the subjects experienced the experiment in the laboratory. A simulation of the experiment is available on request. Instructions in French, the language in which the experiment was conducted, are available on request.

### 1.E.1 Introduction of the experiment

#### Description of the experiment (*Luck and Merit groups*)

- This experiment is divided into 3 parts.
- The experiment is anonymous and your identity is never known to the experimenter or to other participants.
- You will receive a participation fee equal to 3.50 euros
- For each of the three parts, you will receive a remuneration that depends on your own decisions and those made by other participants.
- Your total gains will be paid out in cash at the end of the experiment.
- At the end of the experiment, we will additionally ask you to complete an anonymous questionnaire.

### 1.E.2 Part 1

#### Description of the first part (*Luck and Merit groups*)

- We will ask you to choose one out of six lotteries in which you would like to participate.
- In every lottery, you have the same chance of winning either the high or the low prize.
- We will give you the chance to participate in your preferred lottery (free of charge). You will win the amount that is drawn randomly.

#### Your gains in part 1 (*Luck and Merit groups*)

- The amount drawn randomly will be converted into euros.
- The conversion rate is  $\text{€}1 = 200$  points.
- The amount that you win in the lottery will be paid out in cash at the end of the experiment.

- You can now reread the instructions on your computer. Do not hesitate to raise your hand if you have any open questions.

*The subjects will proceed by making their first choice. The screen is depicted in Figure C2a.*

### **Continuation of part 1 (*Luck and Merit groups*)**

- We will now ask you to choose between two lotteries that both include 5 numbers.
- We will allow you to participate in one lottery (free of charge). You will win a randomly drawn amount.
- In other words, one number of that lottery will be chosen randomly and you will win these points. All numbers have the same chance of being drawn.
- You can now reread the instructions on your computer. Do not hesitate to raise your hand if you have any open questions.

*The subjects will proceed by choosing a lottery. The screen is depicted in Figure C2b.*

### **1.E.3 Part 2**

#### **Part 2 (*Luck and Merit groups*)**

- From now on until the end of the experiment, you will be assigned to a group with 4 other participants. The composition of this group will remain fixed throughout the experiment.
- Imagine that you are in a group with which you are carrying out a project, within a firm or an organization. You will choose between several projects that give each member of your group different bonuses.
- The group composition will be determined randomly by the computer.
- You will never know the identity of the other group members and they will never know your own identity.
- Your choice will never be announced to the other group members. You are the only one who knows the choice you have made.
- You will choose eight times between two different projects.
- Each project results in a different bonus for the people in your group.

#### ***Only shown to the Luck group***

- You will choose your preferred project in each of the eight decisions. Then we will choose one of the eight decisions to be implemented. One of the members of your group will be chosen randomly (all members, including yourself, have the same likelihood of being chosen). The preferred project of that player will be implemented.

- Your position in the project will be drawn randomly.

*Only shown to the Merit group*

- You will choose your preferred project in each of the eight decisions. Then we will choose one of the eight decisions to be implemented. One of the members of your group will be chosen randomly (all members, including yourself, have the same chance of being chosen). The preferred project of that player will be implemented.
- Your bonus will be determined by your performance in a simple task. You and the other group members will conduct this task later in the experiment. The best-performing group member will get the highest bonus, the second-best performing group member will get the second-highest bonus, and so on.

*Shown to both groups*

**Part 2 - Example (*Luck and Merit groups*)**

- Before coming to the real choices, we will show you an example. This will give you the opportunity to familiarize yourself with the setting
- You can now reread the instructions on your computer. Do not hesitate to raise your hand if you have any open questions.

*The subjects proceed by making hypothetical practice decisions. A screenshot of the practice choice and the resulting questionnaire can be seen in Figure C5.*

*Translation of the practice choice and the resulting questionnaire:*

- Screen 1
  - Make a decision between the two projects: A and B
  - Click here for instructions
  - What is your choice?
  - Click here to continue
- Screen 2 (assuming the subject chose project B (A))
  - You have chosen project B (A)
  - Imagine that a randomly chosen participant of your group has chosen project A.
  - *If luck group:* Your position was chosen randomly. You have obtained third position in this example.
  - *If luck group:* The position you obtain in reality can, of course, be different.
  - *If luck group:* You will have an equal chance of attaining any given position within the group.

- *If merit group* Imagine you had the third-best performance in the task. You thus hold third position.
  - What is the amount of your bonus?
  - *If luck group*: What is the bonus of the person that got assigned the first rank?
  - *If merit group*: What is the bonus of the best-performing group member?
  - *If luck group*: What is the bonus of the person in fifth position?
  - *If merit group*: What is the bonus of worst-performing group member?
  - Click o.k. to continue
- Screen 3 (in this simulation questions 1 and 2 were answered correctly but question 3 was false)
    - What is the amount of your bonus? Your answer is correct, your bonus would be 11 points.
    - *If luck group*: What is the bonus of the person in the first position? You are right, her bonus would be 20 points.
    - *If merit group*: What is the bonus of the best-performing group member? You are right, her bonus would be 20 points.
    - *If luck group*: What is the bonus of the person in fifth position? Your answer is not correct, her bonus would be 1 point and not three points.
    - *If merit group*: What is the bonus of the worst-performing group member? Your answer is not correct, her bonus would be 1 point and not three points.

## Partie 2 (*Luck and Merit groups*)

- You will now make the real decisions.

### Belief about future position in the group (merit group only)

- At the end of this part, you will do the task that determines your position within the group. Where do you believe you will rank within your group of five?
  - First place
  - Second place
  - Third place
  - Fourth place
  - Fifth place

### Decisions *Luck and Merit groups*

*The subjects proceed by making the 8 choices behind the veil of ignorance. A screenshot of one of those choices can be seen in Figure C1.*

*Text in Figure C1a (luck group):*

- Make a choice between two projects: A and B
- Reminder: Your position in the group will be drawn randomly.
- Click here for instructions
- What is your choice?
- Click here to continue

*Text in Figure C1b (merit group):*

- Make a choice between two projects: A and B
- Reminder: Your position in the group will be determined by a simple task at the end of the experiment.
- Click here for instructions
- What is your choice?
- Click here to continue

#### **1.E.4 Real effort task**

*Only shown to the luck group*

**Simple task to do on your computer (*Luck group*)**

- You and the other group members will do a simple task on your computer to stay focused.
- This task will have no effect on the rest of the experiment.
- You are asked to copy, with the help of your keyboard, as many of the “words” that appear on your screen as possible. You will have 30 seconds to complete this task. Please separate the words by either a space or comma. These “words” make no sense and are just a row of letters. The order of these “words” is not important.
- You will first have the chance to test the task.
- You are asked to copy as many “words” as possible in 15 seconds.
- You can now reread the instructions on your computer. Do not hesitate to raise your hand if you have any open questions.

*The subjects do the trial task. If they type nothing, we ask if they have understood the assignment.*

- You will now perform the real task in 30 seconds.

*The subjects perform the real task. Afterward, they will be told what their payoffs for Part 2 were.*

*Only shown to the merit group*

**Simple task to do on your computer (*Merit group*)**

- You and your group will do a simple task on your computer to stay focused.
- Your performance in this task will determine your position in the randomly drawn project.
- The person with the highest performance will receive the highest bonus. The least-performing group member will receive the lowest bonus.
- You are asked to copy, with the help of your keyboard, as many of the “words” that appear on your screen as possible. You will have 30 seconds to complete this task. Please separate the words by either a space or comma. These “words” make no sense and are just a row of letters. The order of these “words” is not important.
- The best-performing group member is the person that writes the highest number of “words” in thirty seconds. In the event of ties, we will use the exact time of the last word that was typed. Whoever finished typing the last word first will get first position. The same procedure applies to break all ties.
- You will first have the chance to test the task.
- You are asked to copy as many “words” as possible in 15 seconds.

*The subjects do the trial task. If they type nothing, we come to them directly to ask them if they have understood the assignment.*

- You will now perform the real task in 30 seconds.

*The subjects perform the real task. Right after the task is completed, a screen appears with the within-group position they achieved and their Part 2 payoffs.*

**1.E.5 Part 3**

*Shown to both groups*

**Partie 3 (*Luck and Merit groups*)**

- Your position in the group is known.
- You have received a certain position in part 2. You will keep this position for the third part of the experiment.
- We will ask you to make decisions between two projects for your group.
- You will now know your position and hence your bonus within each project (your position in the table will be highlighted in blue).
- You will choose six times between the two projects.



**Your gains in the third part (*Luck and Merit groups*)**

- You will choose your preferred project in each of the six decisions. Then we will choose one of the six decisions to be implemented. One of the five members of your group will be chosen randomly (all members, including yourself, have the same chance of being chosen).
- The preferred project of that player will be implemented.
- Every member of your group will keep their position that was determined at the end of part 2. You will thus receive the bonus that is linked to your position in the chosen project.
- You can now reread the instructions on your computer. Do not hesitate to raise your hand if you have any open questions.

*The subjects proceed by making the six choices with known rank and known payoff. A screenshot of this decision is found in Figure C3.*

*Text in Figure C3 (both groups):*

- Make a decision between two projects: A and B, knowing that your bonus in project A is 800 points and 800 points in project B because you are person 2.
- [Click here for instructions.](#)
- What is your decision?
- [Click here to continue.](#)

**1.E.6 Questionnaire (*Luck and Merit groups*)**

- You will now answer several questions.
- The first part of the questionnaire consists in making two decisions between two projects.

*The subjects answer the questionnaire with the hypothetical choices and then they answer the questions in the questionnaire. The phrasing of the questions asked in the questionnaire can be found in appendix 1.D.*

## Chapter 2

# Principals' Distributive Preferences and the Incentivization of Agents

co-written with Sophie Cêtre (IRSN)

### Abstract

Do principals' distributive preferences affect the allocation of incentives within firms? We document a robust relationship between French employers' fairness preferences and the incentive contracts they choose for their workers. To establish causality, we run a Principal-Agent lab experiment, framed as a firm setting. In the experiment, subjects are randomized in the principal or worker position. Principals must choose piece rate wage contracts for two workers that differ in terms of ability. Workers have to choose an effort level that is non-contractible. Principals are either paid in proportion to the output produced (Stakeholder treatment) or paid a fixed wage (Spectator treatment). We study how principals make trade-offs between incentive concerns (motivating workers to maximize output) and their own normative distributive preferences. We find that, despite the firm-frame and the moral hazard situation, principals do hold egalitarian concerns, but they are sensitive to both extensive and intensive margin incentives. We characterize the heterogeneity in distributive preferences by positing a utility function that incorporates the principal's other-regarding preferences and we estimate it using a finite mixture model and show that principals can be categorized as performance-maximizing, inequality targeting, and egalitarian.

---

The research leading to the results of this study has received funding from the CEPREMAP and EUR grant ANR-17-EURE-0001. We are immensely grateful to Nicolas Jacquemet, Shachar Kariv and Claudia Senik for particularly helpful comments. We also thank Ghazala Azmat, Béatrice Boulu-Reshef, Andrew Clark, Stefano DellaVigna, Jana Friedrichsen, Jeanne Hagenbach, Dorothea Kubler, Fabrice LeLec and Alvin Roth for the insightful discussions. We also thank seminar participants at PSE, ESSEC and Stanford, and conference participants at FAIR-UCSD Spring School in Behavioral Economics, at the European and North American meetings of the Economic Science Association 2019, at the Economic Science Association world meeting 2020 and at ASFEE 2019 for their feedback and comments.

## 1 Introduction

Employers and managers are, first and foremost, citizens with views about what is fair or not. Are these personal preferences interfering with their managerial choices? Several studies suggest that managers’ social preferences play an important role in the organization of firms and more specifically in the way incentives are allocated among workers (Bertrand and Schoar, 2003; Bastos and Monteiro, 2011; Cronqvist and Yu, 2017). However, the extent to which these preferences affect firms’ performance, and the context in which they are revealed and used to take managerial decisions, remain unclear. Understanding the relationship between managers’ fairness preferences and their managerial decisions is important because there are still substantial variations in management practices that are insufficiently understood. These variations cause persistent gaps in total factor productivity across firms, within and between countries (Bloom et al., 2014).

We provide evidence of a robust correlation between the distributive preferences of executive managers and the incentive structures of their firms. We use a French survey of 4,000 employers and executive managers that includes an extensive set of questions related to workers’ wage compensations. We show that when managers think that a policy of individualized wages is unfair, they are less likely to implement performance pay. Of course, reverse causality can explain the result and the correlation could also be driven by strategic concerns, instead of purely normative preferences. Workers may exert less effort in excessively competitive environments, and this can be anticipated by managers. We show that the relationship declines in strength but remains sizeable and statistically significant when we include strategic motives for using or avoiding performance pay, such as the prevalence of unions, whether they think performance pay motivates workers, whether it is likely to create tensions, etc. This correlation is also robust to a wide array of manager and firm specific controls.

Establishing causality in such a context is complicated. Ideally, we would need a random allocation of managers to firms – to ensure that their normative preferences vary exogenously – and to then measure the type of incentive schemes they subsequently implement. A more realistic approach is to consider exogenous shocks on managers’ preferences or their disclosure. For instance, some managers may face stronger incentives to maximize output than others because their pay is indexed to the company’s performance. This implies that an inequality-averse manager would face a stronger conflict between her normative preferences and incentivization concerns, thereby reducing the influence of her preferences. However, incentive schemes for managers vary non-exogenously across firms and self-sorting of managers into firms leads to a reverse causality problem.

To work around these issues, we run a principal-agent lab experiment, randomizing subjects into manager (principal) or worker (agent) positions. Each principal is matched with two workers of differing ability levels. Both workers choose a costly effort level to produce output, and effort is non-contractible. Principals choose between a series of binary piece rate wage contracts for both workers. These piece rates generate a variable pay-for-performance share of labor income. We randomly allocate principals to either a Stakeholder group (principals’ income is proportional

to the output produced by the workers), or a Spectator group (fixed income). The Spectator group makes the moral hazard situation irrelevant since the principal no longer has an incentive to maximize output. Thus, Spectators can implement their preferred income distribution at no cost, which gives us a measure of the distribution of income principals believe is fair. In the Stakeholder group, principals must take into account workers’ incentives if they want to increase joint output and maximize their own income. This gives us a measure of principals’ willingness to pay for implementing their preferred distribution. The difference in behavior between these two groups isolates normative distributive preferences at the extensive margin. The comparison across treatment groups also characterizes the possible effects of institutional factors such as competitive pressure through market forces on the importance of distributional concerns in incentivization decisions.

Moreover, our framework allows us to pin down the relative importance of various fairness ideals (egalitarian, output maximizing, and equal-procedure) among principals. Piece rate wage contracts are an innovation compared to the existing literature because comparing the piece rates chosen for each worker, depending on their ability level, leads to direct classification into three distributive preferences types. Choosing to reward the high ability worker with a higher piece rate is evidence of being output oriented, since in our setting this approach is output-maximizing if workers best respond to wage contracts. Rewarding both workers with the same piece rate implies to paying them in proportion to the output they have produced. This leads to procedural fairness since both workers are treated equally with the same piece rate. Finally, giving the low ability worker a higher piece rate shows an egalitarian concern, since differences in productivity will be offset. We calibrate these egalitarian contracts in such a manner that if both workers exert the same level of effort, then they are paid the same final total wage. This corresponds to a common situation in real firms, in which both workers are paid the same final wage, despite their different production levels.

The analysis crucially depends on (i) whether or not agents optimally respond to piece rates and (ii) whether principals anticipate such behavior. Before asking principals to choose their preferred wage contracts, we elicit their beliefs concerning workers’ responses to piece rates. This provides control over the output-equality trade-off that principals believe they face before workers start working.

We find that despite the firm-like framing and the moral hazard situation, principals do hold egalitarian concerns. On average, they are willing to accept a trade-off between higher output and reduced within-firm inequality. This willingness is significantly lower if principals are Stakeholders (extensive margin incentives) and it is also the case within treatments when there is a large trade-off between maximizing output and equality. Stakeholders are also more sensitive to these intensive margin incentives than Spectators. When the alternative to the output-maximizing (high-inequality) contract is the equal piece rate contract (rather than the egalitarian contract), principals are not more likely to choose it on average. This indicates that subjects are not more willing to sacrifice output to implement equality in outcomes compared to equality in procedure. This indicates that equality in procedure as such is not seen as a particularly attractive contract characteristic and principals are more interested in distributive

outcomes. Nonetheless, subjects are significantly more likely to opt for a contract that permits equality in procedure if it is posited *directly* against a contract that provides equality in outcomes. We then calibrate a simple utility function that takes principals’ other-regarding concerns into account. The estimates for the representative principal suggest that (i) intrinsic motives are 30% as strong as extrinsic motives in maximizing output and (ii) that principals are averse to extreme inequalities.

Furthermore, we are interested in examining different profiles of principals and identifying which types actually generate inefficiencies in the allocation of incentives. We use a finite mixture model to characterize heterogeneity in preferences. We quantify the weight principals attach to the payoff of high- and low-ability agents, allowing for the variation in these weights according to whether one agent is paid a higher or lower piece rate than the other agent. The Normalized Entropy Criterion (see McLachlan and Peel, 2000, p.214) recommends assigning principals to one of three types: (1) Output maximizers who always favor the contract that maximizes joint output. These principals do not attach any importance to agents’ well-being. (2) Strong redistributors who always attach considerable importance to the low-ability agent’s income, and (3) an intermediate group that attaches positive importance to the low-ability agent’s income if the difference in piece rates becomes too great. We show that all principals in the *Spectator* treatment care to some extent about the distributive consequences of their decisions. On the contrary, 40% of *Stakeholder* principals are classified as output maximizers and are never willing to relinquish income to compress wages. This implies a sizable crowding out of inequality concerns through the provision of extensive margin incentives. Nevertheless, 60% of stakeholders are allocated to either type (2) or (3), suggesting that moral concerns persist on average, even if principals hold a stake in the workers’ output. Counterfactual simulations that vary *workers’* other-regarding preferences show that egalitarian concerns are not always associated with a loss in profit for the firm. Sophisticated output-maximizing principals will mimic the behavior of egalitarian principals because they ultimately make the most efficient choices if agents are egalitarian. But when principals are naive and do not update their effort beliefs, then the egalitarian principals perform better for moderate agent inequality aversion levels.

We contribute to the large and growing body of literature that explores the role of social preferences and inequality in the workplace. Managers’ preferences have rarely been the main focus in the theoretical, empirical and experimental literature, despite the important consequences of managerial decisions on wage inequality and firm performance. Our main contribution to this literature is to study the trade-off that managers face between implementing wage contracts that satisfy their distributive preferences and maximizing the firm’s performance. We first present correlational evidence from rich survey data that shows that fairness concerns affect managerial practices. We then use the controlled environment of the laboratory to analyze and identify the trade-offs that principals face in their incentivization decision. We thus document the consequences of distributive concerns by mapping them to decisions regularly made in firm contexts.

More precisely, we contribute to the experimental literature on social preferences and distributive fairness. This literature studies distributional preferences using relatively abstract

dictator games to infer whether subjects’ allocation decisions are guided by concerns about selfishness, efficiency, inequality, or maximin preferences (e.g. Engelmann and Strobel, 2004; Fisman et al., 2007). Similarly, allocation games have been used to infer whether subjects are primarily concerned about inequality, or rather inequity (Konow, 2000; Cappelen et al., 2007; Almås et al., 2020). These studies do involve the (re)allocation of income *after* a production stage. Therefore, they do not consider the role played by distributional preferences in contract creation that is decided *before* production occurs. Furthermore, Balafoutas et al. (2013) study the conflict between equality, equity, and incentives using a public goods game.

The theoretical literature on social preferences in the workplace has incorporated social preferences into principal-agent models with a focus on team production. Bartling and Von Siemens (2010), Englmaier and Wambach (2010), and von Siemens (2011) incorporated workers’ envy and social comparisons into the derivation of optimal contracts, and found that this affects the optimal incentive structure. However, principals are modelled as output-maximizers.

Field and lab experiments have shown that agents compare their income horizontally (e.g. Clark et al., 2010; Bandiera et al., 2005; Breza et al., 2018; Cohn et al., 2014; Gross et al., 2015; Eisenkopf et al., 2013; Abeler et al., 2010), and that they care about being treated equally (Gagnon et al., 2020). Similarly, workers may have social preferences towards principals and reciprocate high unconditional wages with high effort, as shown in the gift-exchange literature (Bellemare and Shearer, 2009; DellaVigna et al., 2016; Fehr et al., 1993).

Few papers study how other-regarding concerns may affect the allocation of incentives within a firm. Existing work shows that principals’ incentives affect how they allocate their supervision (Bandiera et al., 2007). Principals take into account fairness concerns in a context in which they are matched with a single agent (Fehr and Schmidt, 2004; Fehr et al., 2007). Brandts et al. (2019) study principals’ distributive concerns in a gift-exchange setting, where principals’ strategic motives are muted. Kocher et al. (2013) show that social preferences correlate with preferences concerning managerial leadership styles. Cabrales et al. (2010) also document a correlation between social preferences and choices concerning contracts, but in a setting in which principals have to compete for workers.

The remainder of the chapter is structured as follows: Section 2 characterizes the trade-off that we want to study in more detail; Section 3 presents survey evidence; Section 4 introduces the design; Section 5 presents our main results; Section 6 presents the structural model; Section 7 concludes.

## 2 Trading off distributive and incentivization concerns

This chapter studies the trade-off principals face between the implementation of an incentive scheme that they view as *fair* and the implementation of an incentive scheme that they may view as *optimal* for maximizing the outcome of the firm. Before detailing the empirical analysis of this trade-off, we clarify the theoretical foundations of this trade-off.

A manager has to choose an incentive scheme  $\mathbf{w} = (w_1, w_2, \dots, w_n)$ , which assigns incentives  $w_i$  to each of her  $N$  workers that may differ in terms of their ability. This incentive scheme

can be any wage policy, ranging from tournament incentives, to flat monthly wages, hourly wages, or a piece rate. Incentives can also be characterized by a combination of the above; for example, workers may benefit from a flat base-rate and additionally compete for a bonus through a tournament. The incentives chosen by the manager affects the outcome of the firm  $\pi(\mathbf{w})$ , as well as the income of the workers, denoted by the vector  $\mathbf{y}(\mathbf{w})$ .

Managers are aware that some incentivization methods are more successful than others to maximize  $\pi(\mathbf{w})$  (denoted as  $\pi^{max}$ ), which is a function of the incentive schemes that induces this optimal outcome  $\mathbf{w}^{max}$ . Importantly,  $\mathbf{w}^{max}$  depends on the *beliefs* that managers hold about the effectiveness of these incentivization methods. For example, a manager may believe that tournament incentives are less effective in motivating workers than a flat pay-scheme because they can lead to anti-social behavior among employees (Charness et al., 2013), others may believe that differences in monthly wages are detrimental for worker morale and performance if their employees view them as unjustified (Breza et al., 2018). This implies that  $\mathbf{w}^{max}$  may take very different forms across individuals, which must be accounted for in the empirical design.

The innovation of this chapter is to ask whether managers care about  $\mathbf{w}$  because of the distributive consequences of the incentive scheme. To capture this, we assume that managers care about the distribution of income within their firm ( $\mathbf{y}$ ) and that they hold a distribution of income within this firm that they perceive as fair ( $\mathbf{y}^{fair}$ ). This distribution of income is made possible with the incentive scheme  $\mathbf{w}^{fair}$ .

In this chapter, we are interested in asking whether  $\mathbf{w}^{fair}$  is relevant to the principal’s choice of incentives, denoted as  $\mathbf{w}^*$ . This situation may arise if there is a conflict between the incentives that a principals views as optimal and those that yield a fair distribution of income, i.e.  $\mathbf{w}^{max} \neq \mathbf{w}^{fair}$ . For example, if a principal holds egalitarian fairness views and wants to implement an incentive scheme such that all workers are paid the same at the end of the month, but also believes that tournament incentives are the most effective way to motivate workers and maximize the profit of the firm, then we have a situation where  $\mathbf{w}^{fair} \neq \mathbf{w}^{max}$ . This conflict in objectives would, however, be non-existent if she believes that a flat wage-scheme is the most effective way to motivate workers, because tournament incentives are viewed as counter-productive to firm productivity. Alternatively, some managers may believe that it is optimal to reward the most productive workers with very high performance bonuses. This sort of contract could lead to a wage gap even wider than the actual productivity gap. In this situation, managers would obviously face a trade-off if  $\mathbf{w}^{fair}$  is a flat payment scheme, but they would also be conflicted if they hold as fair an earning gap that is proportional to the ability or productivity gap across workers within the firm.

Furthermore, even if a manager faces a situation where  $\mathbf{w}^{fair} \neq \mathbf{w}^{max}$ , the relevance of her distributive preferences on incentivization choices depends on her willingness to trade-off  $\pi$ , as well as her own income, in order to implement  $\mathbf{w}^{fair}$  or minimize the discrepancy between  $\mathbf{w}^{fair}$  and the implemented incentive scheme,  $\mathbf{w}^*$ . Some managers may hold distributive preferences over  $\mathbf{w}$  but the relative importance of these preferences may be secondary compared to their own payoff or the well-being of the firm.

In this chapter, we will empirically analyze this conflict from two perspectives. First, we will

use survey data to study whether  $\mathbf{w}^{fair}$  actually correlates with incentive schemes used in real firms. Then, we will use an experiment to take a more granular view of the shape of  $\mathbf{w}^*$ , as well as the strength of these distributive concerns. In Section 6, we use the experimental data to estimate a simple utility function that captures distributive concerns of managers. We show that managers can be categorized as strongly egalitarian, output maximizing, and inequality targeting (subjects that accept intermediate levels of inequality). In Appendix 2.D, we simulate  $\mathbf{w}^*$  for each type of principal, while varying (a) social preferences from the agents and (b) whether principals anticipate the social preferences of the workers. The analysis reveals interesting insights of how  $\mathbf{w}^{fair}$  affects the implemented incentive scheme. For example, we show that egalitarian principals face a strong trade-off between  $\mathbf{w}^{fair}$  and  $\mathbf{w}^{max}$  if workers are not inequality averse themselves. In that case, an egalitarian principal will always perform worse than the other two types of principals. Principals that are concerned with disadvantaging a more talented worker, however, face a much weaker trade-off, because their concern of not making a talented worker worse off than the other workers is often in line with profit maximizing motives. Furthermore, we show that principals that do not attach any importance to their distributive preference when choosing incentives will always implement incentives that are profit maximizing, as long as they are able to correctly anticipate the social preferences workers hold. This, however, changes if workers are egalitarian and their preferences are not correctly anticipated by the principal. In that case an egalitarian principals face no real trade-off between  $\mathbf{w}^{fair}$  and  $\mathbf{w}^*$  and the incentives may be more profit-maximizing than those chosen by a selfish principal who, falsely, assumes that agents are also selfish.

### 3 Managers’ preferences: survey evidence

We use the two waves (2011 and 2017) of a French survey on Professional Relationships and Firm Negotiations (REPONSE), that has been conducted every six years since 1993.<sup>1</sup> The survey was administered in 4,023 firms in 2011 and 4,364 in 2017<sup>2</sup> and three types of questionnaires are distributed, one for a representative of the executive managerial positions<sup>3</sup>, one for a personnel representative<sup>4</sup>, and one for employees of the firm.

We use the questionnaire dedicated to managers in which they are asked whether workers benefited from individualized pay rises and also whether they received bonuses related to individual performance in 2010 and 2016, for the 2011 and 2017 waves, respectively. The survey asks both questions to white- and blue-collar workers.<sup>5</sup> We use these outcome variables as indicators of whether the firm engaged in pay-for-performance and thus wage (or bonus) differentiation

<sup>1</sup>We only had access to the last two waves via the *Réseau Quetelet*, as the earliest ones are no longer available.

<sup>2</sup>To be more precise, the survey is conducted at the plant (*établissement*) level, unless the firm has an independent status. We use “firm“ in the text for simplicity. These plants are representative of 196,434 plants with 11 employees or more, and approximately 10 million employees in France, according to the 2017 wave documentation.

<sup>3</sup>Either CEO, Secretary General, Plant Director, Head of Human Resources, or another top managerial position.

<sup>4</sup>Either a union representative or a staff representative.

<sup>5</sup>*Cadres* and *Non Cadres* in French.



based on effort or ability, for both types of workers. In our experiment, we proxy this type of choices by the series of decisions between two binary piece rate wage contracts.

Regarding our main explanatory variable, we use a question to proxy principals’ distributive preferences: the questionnaire asks whether the manager believes that individualized wage raises are unfair. Managers who agree with this statement can be categorized as averse to inequalities among their employees.

Columns 1 and 4 of Table 2.1 use a logit specification to show that there is a strong negative correlation between both variables. Managers who think that individualized wage rises are unfair are 20 (20.9) percentage points less likely to run a company that implements individualized wage rises among white-collar (blue-collar) workers. Obviously, this correlation is likely to suffer from reverse causality or self-selection since managers are not randomly allocated across firms. An omitted variable bias is also likely: this correlation may capture other motives. Principals may answer the fairness question by considering what workers think is fair instead of their own personal distributive preferences. For instance, principals may believe that individualized wages generate tensions among their employees, and might therefore avoid implementing them in order to maintain levels of production and profit. They may declare that individualized wages are unfair, by considering their employees’ opinions rather than their own preferences.

Table 2.1: Individualized wage raises and managers’ distributive preferences

	White-collar workers			Blue-collar workers		
	(1)	(2)	(3)	(4)	(5)	(6)
<b>Dep var = Did white/blue-collar workers benefit from individualized wage raises</b>						
Individualized wage raises are unfair	-0.200*** (0.0127)	-0.103*** (0.0140)	-0.0830*** (0.0151)	-0.209*** (0.0118)	-0.138*** (0.0130)	-0.119*** (0.0145)
Individualized wage raises create tension		-0.146*** (0.0102)	-0.0993*** (0.0110)		-0.116*** (0.0101)	-0.0881*** (0.0113)
Individualized wage raises motivate		0.0678*** (0.0165)	0.0579*** (0.0178)		0.0425*** (0.0156)	0.0317* (0.0174)
Individualized wage raises are subjective		-0.111*** (0.0120)	-0.0418*** (0.0131)		-0.0926*** (0.0114)	-0.0650*** (0.0128)
Wave dummy	Yes	Yes	Yes	Yes	Yes	Yes
Individual controls	No	No	Yes	No	No	Yes
Firm controls	No	No	Yes	No	No	Yes
Observations	7666	7566	5771	8139	8033	6156
Pseudo $R^2$	0.026	0.069	0.185	0.028	0.055	0.104

Robust standard errors in parentheses

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

The Table displays marginal effects from logit specifications. We regress a binary variable for whether white-collar workers benefited from individualized wage raises in columns (1) to (3) (blue-collar workers in columns (4) to (6)) on binary variables for whether the manager answering the survey thinks that individualized wage raises are unfair, whether they create tensions, motivate, or are subjective. A 2017 wave dummy is added in all regressions. We additionally control for individual and firms’ controls in columns (3) and (6). See Appendix Table A1 for a description of all the variables.

Fortunately, the survey is extensive enough to control for such beliefs. To isolate normative preferences as much as possible, in the rest of the columns we control for other strategic concerns that may lead the firm to avoid or adopt performance pay, such as the belief that it motivates

workers, can create tensions, or is difficult to base on objective criteria.<sup>6</sup> In columns (3) and (6), we also control for the sociodemographic characteristics of the manager answering the survey and a large set of firm-related controls, in order to minimize the self-selection issue.<sup>7</sup>

We see that the normative distributive preference variable (individualized wage raises are unfair) still leads to a negative and significant coefficient, even after the inclusion of all these controls. It is similar in magnitude to the coefficient associated with the belief that individualized wages generate tensions within the firm.<sup>8</sup> Appendix Table A2 shows the same regressions but focuses on the implementation of performance-based bonuses, rather than individual pay raises, as the dependent variable. Results are similar in sign and significance.

This representative survey of French managers reveals a robust correlation between the implementation of individualized wage raises or bonuses and normative distributive preferences. With respect to our research question, the data reveals three important regularities that motivate the design of our experiment: First, a substantial share of managers (16%) does agree that individualized incentives are unfair. This implies that, for some managers, individualized incentives are considered as unfair. Second, the majority of managers is convinced that incentives are an effective to motivate workers (89%). Putting the two together indicates that there is a trade-off between the goal to maximize the profit of the firm and implementing incentive schemes that are in line with fairness views. This is further confirmed by the fact that fairness views are not co-linear with other beliefs that affect the effectiveness of individualized wages. Third, both, fairness views and beliefs about the effectiveness are important predictors of the use of individualized incentives. This shows that both convictions and beliefs matter for behavior. Hence, this survey indicates that managers in real situations are, at least in some instances, willing to abandon a subjectively efficient tool to increase production (performance pay) in order to avoid a conflict with their own normative distributive preferences.

## 4 Experimental design

We take the evidence from the survey as a starting point for an experimental analysis of this topic. The goal of the experiment is to make a more granular assessment of the trade-off that principals face and test some of the predictions that are hard to answer using the survey data. First, we are interested in targeting the trade-off between inequality concerns and incentivization decisions. To that end, we need more control over workers’ behavior, workers’ ability, and principals’ beliefs than we can achieve through the survey. Second, we want to study a wider set of fairness ideals. Principals may for example tolerate small levels of inequality, or they may value more equality of opportunity than equality in wage outcomes. Third, we are interested in causally assessing how these fairness concerns are crowded out by incentives. The experiment

<sup>6</sup>Appendix Table A1 describes all these variables, including an English translation of the original questions.

<sup>7</sup>Of course, we do not claim that such an empirical strategy is sufficient to establish a causal relationship.

<sup>8</sup>Whether managers believe that individualized wages motivate workers is smaller in magnitude and only significant at the 10% level for blue-collar workers. This can be explained by the fact that 90% of the sample agrees with this affirmation. Hence, there is practically no variation in the answer to this question. It also highlights that managers generally believe in the motivating power of monetary incentives. This is also in line with a more general view that good management practices include rewarding high individual performance Bloom et al. (2014)

will permit us to assess whether the effect of fairness concerns on incentivization decisions is likely to disappear once managers’ own remuneration is directly and saliently affected by their incentivization decision.

## 4.1 Lab setting

Each session of our laboratory experiment consists of 18 to 24 subjects that are randomly assigned as either an agent or a principal at the beginning of the session. Furthermore, each principal is randomly matched with two agents and the groups and roles are fixed throughout the experiment. The experiment is framed as an interaction in a firm, which is the most natural setting in which principal-agent interactions and wage distribution take place (see Alekseev et al., 2017, for a discussion on contextual instructions). Agents are called “workers” and principals are called “Managers”.<sup>9</sup> We inform all participants that the currency used during the experiment is the ECU with the following conversion rate:  $1\text{€} = 10\text{ ECU}$ . The detailed instructions (translated from French to English) are included in Appendix 2.E.

We ran the experiment at the Laboratoire d’Economie Experimentale de Paris between December 2018 and January 2019. All sessions were in French with French-speaking subjects who were recruited using ORSEE (Greiner, 2015). Sessions were computerized using zTree (Fischbacher, 2007), average payments were 15 € and sessions lasted 90 minutes, on average. Overall, 339 subjects were invited in groups of 18, 21 or 24 subjects.<sup>10</sup> 226 participants were randomly assigned to the worker role and 113 to the principal role.

## 4.2 Workers

**Production and cost functions** Workers are invited to make consecutive effort choices for a number of piece rates. Their income is composed of a fixed share of 90 ECU (9€) and a variable share that depends on the piece rate they are paid as well as their induced production function. They are informed that an (anonymous) principal will choose a piece rate for them that will determine the variable share of their wage. The latter takes the general form  $y_i = \alpha_i e_i$ , where  $e_i$  is the effort level chosen by the worker and  $\alpha_i$  is the marginal productivity which varies across workers ( $\alpha \in \{\alpha_H, \alpha_L\}$ ). In all sessions, we define  $\alpha_H = 60$  and  $\alpha_L = 40$ .

$\alpha_i$  is allocated according to the subjects’ performance at an aptitude test that the workers take after receiving the instructions about the workplace setting described above, and after completing a comprehension test.<sup>11</sup> They are informed that performing better at the aptitude test will increase their chances of having higher productivity (a high  $\alpha$ ). Using an aptitude

---

<sup>9</sup>We use the French word “*gérant*” rather than “manager”, which is also frequently used, in order to avoid any confusion stemming from the possible negative connotations of the word “manager” in French (it is sometimes related to being “bossy”). “*Gérant*” is the French translation of manager and has a more neutral connotation. Moreover, the principal in our case is also an employee of the firm. Hence, using the words “employer” and “employees” could be misleading.

<sup>10</sup>Since the design of the experiment was based on a group composed of a principal matched with a pair of workers, the number of participants was a multiple of 3 in each session. Variation in participants per session stemmed from differences in the show-up rate.

<sup>11</sup>To ensure that all participants understand the experiment, they take an extensive comprehension test that asks them to explore the environment. The questions are designed to ensure that they understand the consequences of their decisions. Section 2.F describes this test further and how the subjects performed.

test to generate heterogeneity in productivity across agents in a stated effort experiment has been used in gift-exchange experiments to justify induced productivity differences (Bolton and Werner, 2016; Gross et al., 2015). The idea is to overcome a certain arbitrariness in productivity differences by creating a link between induced and real ability that would not exist under random ability allocation. Furthermore, we deliberately use an aptitude test that not only accounts for innate ability or willingness to work hard, but that may also depend on the education that the worker benefited from. We made this design choice to capture the fact that principals often deal with agents that have different ability levels precisely because they benefited from different education levels. This will then also be relevant for their evaluation of incentives schemes from a fairness perspective.

The aptitude test consists of nine questions: three logic questions, three French questions and three general knowledge questions. The French and logic questions were simplified versions of TAGE MAGE, a French equivalent of GMAT. Workers have 10 minutes to complete a practice test (same format but different questions) and then have five minutes to complete a test that will define their production function.<sup>12</sup> Ability is determined at the pair level. We assign  $\alpha_H$  to the worker with the best performance within the pair and  $\alpha_L$  to the other one.

The cost function is constant across agents, and it is convex in effort choices. Figure A1 in the appendix displays the production and cost function of both workers.

**Workers’ decisions** The agents make effort choices for *all* piece rates that can be chosen by the principal. As is common in the strategy method, they are informed that the principal will only choose one of their choices as payoff-relevant.<sup>13</sup>

Piece rates range from 0.3 to 0.70 ECU (for high-ability workers) and from 0.3 to 0.75 ECU (for low-ability workers) in increments of 0.05. It is possible that workers will react differently to a certain wage if the previous piece rate was higher or lower. Nonetheless, we decided not to completely randomize the order applied to the workers because it is unfeasible to robustly identify order effects under complete randomization (81 possible combinations would need to be compared). However, we test for order effects by looking at two benchmark cases: (1) ascending order of piece rates starting at 0.3 and ending at 0.7 ECU; and (2) descending where the order is reversed. One of the order is randomly assigned to each worker.

Workers choose effort levels from a discrete set between 0 and 5 ( $e \in \{0, 0.5, 1, \dots, 5\}$ ). We elicit effort decisions for *all* piece rates. The final income of the worker is  $\pi_i^w = pr^m \alpha_i e_i - c(e_i) + 9$  €, where  $pr^m$  is the piece rate chosen by the principal and  $c(\cdot)$  is the effort cost function.

A screenshot of agent B’s decision can be found in the appendix, Figure A2. For each piece rate, workers can view their production table showing how each effort level translates to production, effort cost and net variable income. To ease the cognitive burden, we show them

<sup>12</sup>Appendix 2.G includes the questions. The practice test is simply meant to allow them to evaluate the type of questions they will encounter and keep them occupied while principals progress through the experiment. Agents receive no feedback on this practice test.

<sup>13</sup>One could argue that workers may themselves form beliefs regarding which piece rate is more likely to become payoff-relevant. This is unlikely to happen in our setting since from the worker’s point of view, the principal’s objective function is unknown. First, they do not know that principals choose piece rates for two workers at the same time. Second, they are not informed about how principals are paid.

a simulation of the consequences of their decision when clicking on a particular effort level. For instance, when effort level 3 is selected, the screen shows the worker’s production output (180 units), the current piece rate (0.5 ECU), the cost (48 ECU) and the net income (42 ECU) associated with such an effort level.

**Workers’ information set** Workers are informed that the payoff-relevant piece rate will be chosen by a principal but they are not informed that this principal also chooses a piece rate for another worker. We chose this feature of the design to avoid horizontal wage comparisons among workers that could lead them to sabotage very unequal piece rates on the basis of their own fairness motives. Since we want to focus on the principals’ reaction to wage inequality among workers, we want to eliminate other, possibly confusing, factors from the principal’s decision, as far as possible.

Furthermore, workers are not informed how their decisions affect the principal in order to avoid vertical social preferences that have been documented in the field (Ashraf and Bandiera, 2018; DellaVigna et al., 2016). Since the remuneration of principals is our main treatment variation, we want workers’ effort decisions to be orthogonal to the treatment.

### 4.3 Principals

Each principal is matched at the beginning of the session with two workers, and different ability levels are assigned to them on the basis of the aptitude test. Each worker is randomly assigned a neutral label – “Worker A” or “Worker B” – and we present a table summarizing both workers’ characteristics in terms of productivity (how much output they can produce for a given effort level) and cost function (see Figure A1). Labels A and B are randomized and thus independent of the ability level.<sup>14</sup> This neutral labeling implies that we never tell the principal which subject is more productive; they can infer this on their own from the information disclosed in the tables.<sup>15</sup>

**Belief elicitation** Principals are invited to choose wages for the pair of workers they are matched with. Prior to making these decisions, we elicit their beliefs about the effort level chosen by the workers for *each* piece rate they could possibly implement. We elicit beliefs regarding each worker’s effort sequentially to avoid asking too many questions at once. The workers’ order of appearance (either Worker A or Worker B) is randomized at the principal level. At the end of the experiment, we randomly draw one guessed belief, and if the principal’s guess is correct she receives 10 ECU (1€).<sup>16</sup> The drawing of the payoff-relevant piece rate in the belief elicitation is completely independent of the drawing of the payoff-relevant choice in the latter part of the experiment in order to achieve independence in the decisions across the two parts.

<sup>14</sup>However, Worker A’s characteristics are always summarized in the left-hand table. Starting with Worker B on the left would have been puzzling for many subjects

<sup>15</sup>In the comprehension test, we asked them to find out which worker was the most productive in a hypothetical situation (table with completely different production and cost function). See Appendix 2.F for more details regarding the comprehension test.

<sup>16</sup>We are aware that this is a very simplistic way of eliciting beliefs and we measure the modal rather than the mean belief. However, we want to minimize complexity in the experiment and thus opt for a method of incentivizing beliefs that is easier for the subjects to understand.

Belief elicitation of workers’ effort choices plays a vital part in the experiment. It enables us to determine whether an egalitarian contract choice originates from normative distributive preferences or different beliefs regarding how workers should behave under each contract. Principals may believe that workers do not seek to maximize their own income and would choose different effort levels instead of the best responses. Under such a belief structure, an egalitarian contract may become optimal. In other words, eliciting beliefs enables us to determine whether our classification of output-maximizing contracts is also shared by principals or not.

Table 2.2: Set of decisions made by the principal assuming workers’ choose effort to maximize their own income

	piece rates		Income Contract 1		Income Contract 2		Joint output		
	Contract 1	Contract 2	$Worker_H$	$Worker_L$	$Worker_H$	$Worker_L$	Contract 1	Contract 2	$\Delta$ -output
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
<i>Egalitarian VS output-maximizing contract</i>									
N°									
1	0.4 – 0.6	0.55 – 0.45	25.5	25.5	51.5	13.5	250	270	20
2	0.4 – 0.6	0.6 – 0.4	25.5	25.5	63	10.5	250	300	50
3	0.5 – 0.65	0.55 – 0.45	42	30.5	51.5	13.5	280	270	-10
4	0.5 – 0.65	0.6 – 0.4	42	30.5	63	10.5	280	300	20
5	0.5 – 0.65	0.65 – 0.35	42	30.5	75.5	8	280	310	30
6	0.5 – 0.65	0.7 – 0.3	42	30.5	90	6	280	340	60
7	0.5 – 0.75	0.6 – 0.4	42	42	63	10.5	300	300	0
8	0.5 – 0.75	0.65 – 0.35	42	42	75.5	8	300	310	10
9	0.5 – 0.75	0.7 – 0.3	42	42	90	6	300	340	40
<i>Equal piece rate VS output-maximizing contract</i>									
10	0.5 – 0.5	0.55 – 0.45	42	17	51.5	13.5	260	270	10
11	0.5 – 0.5	0.6 – 0.4	42	17	63	10.5	260	300	40
12	0.55 – 0.55	0.6 – 0.4	51	20.5	63	10.5	290	300	10
13	0.55 – 0.55	0.65 – 0.35	51	20.5	75.5	8	290	310	20
14	0.55 – 0.55	0.7 – 0.3	51	20.5	90	6	290	340	50
<i>Egalitarian VS equal piece rate and output maximizing contract</i>									
15	0.4 – 0.6	0.5 – 0.5	25.5	25.5	42	17	250	260	10
16	0.5 – 0.65	0.55 – 0.55	42	30.5	51	20.5	280	290	10

*Notes:* This table shows the series of decisions principals are asked to make. All units in columns (1)–(6) are in ECUs. The units in columns (7)–(9) are production quantities. The first two columns display the piece rates that are associated with each contract. The left-hand piece rate is the piece rate for the most productive worker ( $Worker_H$ ) and right-hand piece rate is for the least productive worker ( $Worker_L$ ) of the pair. The decisions can be split into egalitarian vs output-maximizing and equal piece rate vs output-maximizing choices. Egalitarian contracts result in outcomes proportional to effort. The equal piece rate contracts result in outcomes proportional to output. Columns (3)–(6) correspond to the variable share of income and thus exclude the 90 ECU show-up fee, common to all workers. The variable income levels (columns (3)–(6)) and the joint output for each contract (columns (7)–(8)) are conditional on the workers best responding to the piece rate.

**Contract decisions** After the belief elicitation part of the experiment, principals make 16 binary decisions between two contracts, where each contract consists of two piece rates (one for the more productive worker  $Worker_H$  and one for the less productive worker  $Worker_L$ ). The choices are summarized in Table 2.2, showing the piece rates associated with each decision, as well as the distributive and productive consequences of each option (conditional on the workers best responding to the piece rate).

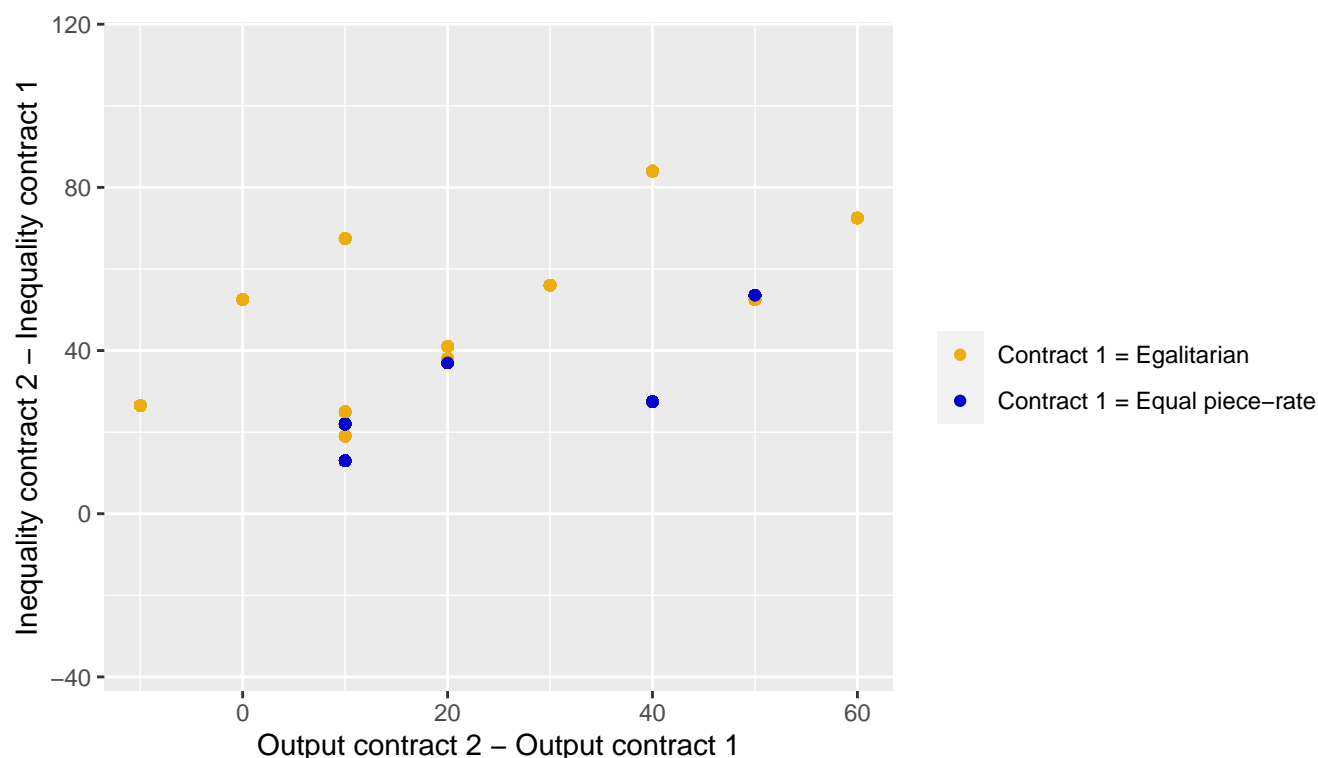
Before detailing the choices that principals face, we want to clarify the different types of contracts. Our experiment studies three types of contracts. First, we have egalitarian contracts,

where the low ability worker ( $Worker_L$ ) receives a *higher* piece-rate than the high ability worker ( $Worker_H$ ). If workers best respond to wages, the egalitarian piece rate will either perfectly equalize income levels (as in Contract 1 of Choices 1-2, 7-9, and 15) or significantly decrease inequality compared to its alternative (as in Contract 1 of Choices 3-7 and 16) because the low ability worker now receives more for each unit produced. This results in a situation where equality in *effort* yields equality in income. Second, we have equal piece-rate contracts. These contracts pay an equal piece-rate to each worker. They, thus, yield a situation where equality in production leads to equality in incomes. However, in this case, equal effort does not yield equal income, since the two workers differ in their productivity. Third, we have a class of choices, where  $Worker_H$  receives a higher piece-rate than  $Worker_L$ . This contract leads to high inequalities *ex-post* but also to higher levels of joint output compared to the alternative.

In the choices that principals make, two of these contracts are posited against each other. As shown in Table 2.2, we have three classes of choices. First, we have choices, where principals choose between an egalitarian ( $Worker_L$  receives a higher piece rate compared to  $Worker_H$ ) and an output-maximizing contract. Second, we have choices, where principals choose between an equal piece rate contract (both workers receive the same piece rate) and an output-maximizing contract. Third, we have choices where the egalitarian contract is posited against an equal piece-rate contract. Contract 2 yields higher output compared to Contract 1 in all decisions, except for Choice 3, where the egalitarian choice is also output maximizing. This Choice permits us to test for situations in which the egalitarian or equal piece rate contract is output-maximizing to avoid positing that equality is always desirable. Some people may consider that ability-induced inequality is fair. However, Contract 2 always leads to larger inequality when workers best respond. For the sake of simplicity, we abstract from this and retain “output-maximizing” label when referring to Contract 2.

Within each class, choices were calibrated so that both inequality and joint output vary across choices, but without a perfect positive correlation. Otherwise, it would have been impossible to disentangle their respective impacts on contract choices. Figure 2.1 shows how differences in inequality between Contract 2 and 1 (on the y-axis) and output (difference in output between Contracts 1 and 2, on the x-axis) vary across choices. Orange dots represent each case in which Contract 1 is egalitarian and blue dots show when Contract 1 is an equal piece rate (equal procedure) contract. Choices with an egalitarian contract are naturally located at the top of the graph since they lead to a more drastic compression of wages than equal piece rate contracts. The difference in inequality ranges from 13 to 84 ECU and difference in output ranges from -10 to 60 units produced. In ECU terms, the difference in output-based income is twice as small, since each unit of output is sold at 0.5 ECU. Therefore, in ECU terms, we can say that the inequality level varies more than the output level across choices. This calibration decision is based on pilot data showing that if output differences are too large across Contracts 1 and 2, principals eventually adopt a corner solution in which they maximize income. Consequently, if inequality and output varied on about the same scale, we would not be able to see that people also care about inequality to some extent: all principals would be mistakenly described as selfish income-maximizers. In this study, we focus on the window in which there is a trade-off between

Figure 2.1: Contract trade-offs assuming best responses



*Notes:* The Figure plots the theoretical trade-offs (assuming best responses), underlying the 16 contract choices that principals have to make. The y-axis shows the difference in inequality between both contracts, and the inequality of a contract is measured by the high-ability worker’s wage minus the low-ability worker’s wage. Hence, Contract 2 becomes increasingly unequal relative to Contract 1 as we move up the y-axis. The x-axis is the difference in output between contracts. Contract 2 becomes more efficient relative to Contract 1 as we move to the right-hand side of the plot. Yellow dots represent the trade-offs of equal piece rate contracts vs output-maximizing contracts, and the blue dots represent the trade-offs of egalitarian contracts vs output-maximizing contracts.

maximizing output and equality.

Figure A3 shows how we asked principals to make contract choices during the experiment. The top part of the screen shows the tables summarizing the information for Workers A and B<sup>17</sup>, the middle part asks principals to choose between both contracts, and the bottom part simulates the consequences of such a choice, both for the workers and for the principal. This simulation part helps to ease the cognitive burden and saves computation time. This simulation is based on the effort belief elicited beforehand. We remind them of the effort level they expect their workers to choose. We then inform them about the expected production associated with such effort levels and the variable income that each worker would receive under the selected contract. The table is updated when the principal selects a different contract. We instruct them to try out both simulations before making a choice.

Since this screen must be repeated 16 times for each of the choices, we randomize several features to avoid any anchoring biases. The 16 choices appear in a random order at the subject-level. Within a choice, the labeling of contracts as “Contract 1” or “Contract 2” is randomized.

<sup>17</sup>Note that the production and cost of each worker for each effort level are not shown, only their net variable income. We wanted to avoid overloading the decision table and therefore opted to omit this part from the representation. However, they are told about the composition of the worker’s wage in the instructions and comprehension test, and they can access this information by clicking on the description button on the top-right corner of the screen.



This implies that people cannot always choose Contract 2 to maximize their own income. The “Worker A” and “Worker B” labels are randomly assigned to the high-ability and low-ability workers and are thus independent of productivity differences.

**Treatments** Between subjects, we will implement two treatments: (1) the *spectator treatment* and (2) the *stakeholder treatment*. The treatment varies across sessions, meaning that principals in the same session faced the same treatment.

In the spectator treatment, the principal receives a fixed wage of 20 € that is completely independent of her workers’ output. The treatment enables us to identify how normative distributive preferences affect preferences over contracts without any personal and monetary cost for the principal herself. In each decision, the principal is asked to make a trade-off between the implementation of an egalitarian (or equal piece rate) and an output-maximizing contract, keeping her own income constant across all the decisions. The size of the trade-off is documented in column (9), if the principals believe the agents are best-responding. The treatment can be seen as analogous to a situation in which principals have no personal stake in the outcome of their organization (e.g. civil servants at the end of their career).

In the stakeholder treatment, the principal receives a fixed participation fee of 60 ECU (6 €) and a variable share from the sales of the output produced by the workers. For each unit produced, she receives 0.5 ECU. She now faces a trade-off between maximizing her own income and implementing an egalitarian (or equal piece rate) contract. By analyzing choice patterns, we can infer from this treatment the price the principals are willing to pay in order to implement an egalitarian or equal piece rate contract. The size of the trade-off depends largely on the principals’ beliefs regarding whether or not they expect workers to best respond to the piece rates. This highlights the importance of the belief-elicitation part of the experiment.

**Principals’ information set** Principals have full information about the environment they face. Principals are instructed that workers are not informed about the salary, nor the existence, of the second worker. They also know that the workers are not explicitly informed about how the principal herself is incentivized. Importantly, they know which of the workers is of low and high ability. Furthermore, they are also informed of the cost that workers have to endure for a given level of effort. This information is necessary for managers to infer the distributive consequences of their actions. Principals are specifically instructed about this in the comprehension test (see Appendix 2.F).

#### 4.4 Summary statistics

Table A3 shows the subjects’ sociodemographic characteristics by role. Approximately 50% of the subjects are female, the average age is around 25 years old and 60% are students. There are no systematic differences in observed characteristics between workers and principals. Table A4 reports the same statistics focusing on principals only. It shows how our two treatment groups, Spectators and Stakeholders, differ along observed characteristics. Differences are non-significant, except for gender. Despite randomization across treatment groups, Stakeholders

are more often female than Spectators. If anything, this bias in our sample should yield more conservative estimates of differences across treatment groups. Women are often found to be more inequality-averse in dictator games (Croson and Gneezy, 2009), which in our case, should lead to a smaller difference in contract choice between Spectators and Stakeholders. Nevertheless, we control for this variable in all our regressions.

## 5 Results

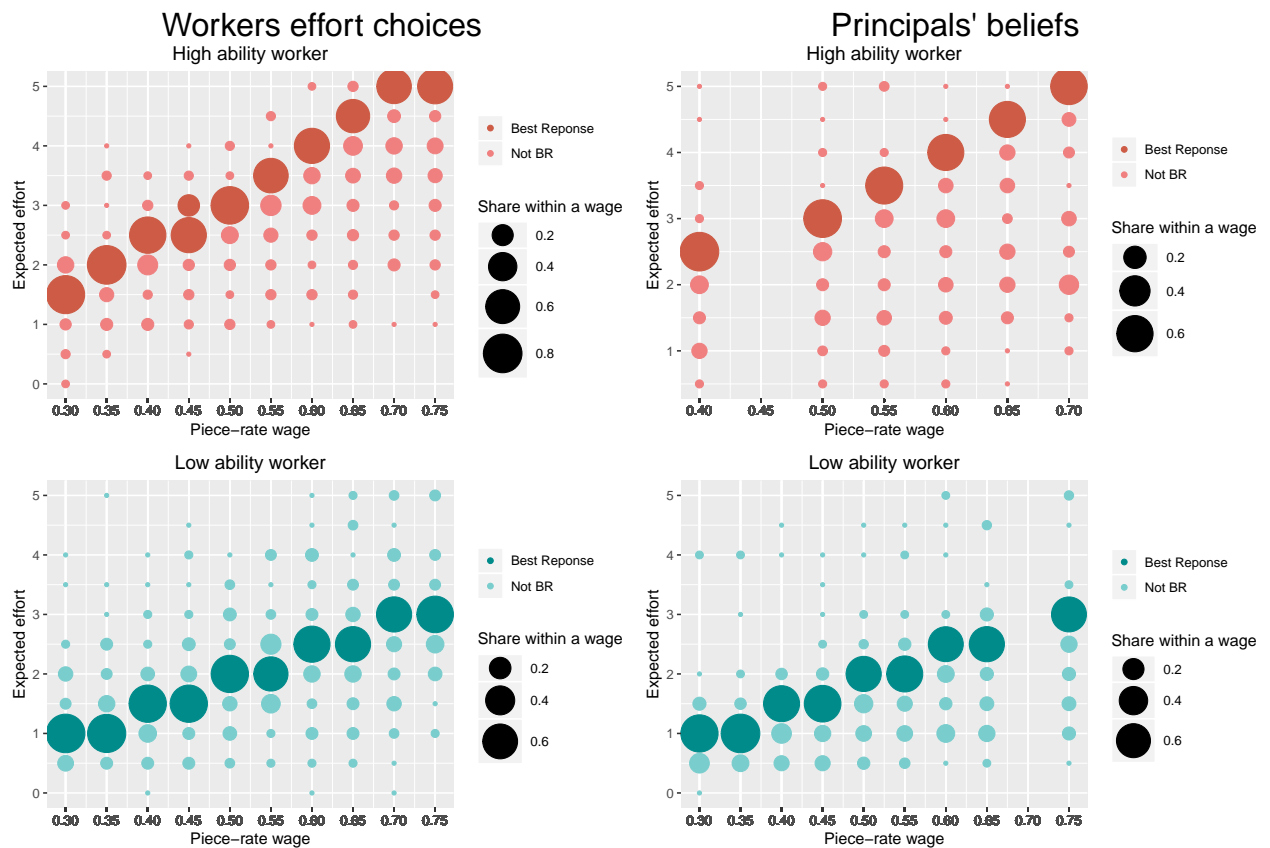
### 5.1 Effort choices and effort beliefs

We first describe, side-by-side, the effort levels chosen by workers for each piece rate wage and principals’ corresponding beliefs. Figure 2.2 plots workers’ effort choices by ability type (high-ability workers in red and low-ability workers in blue) on the left-hand side, and principals’ beliefs on the right-hand side. For each piece rate wage on the x-axis, we use mass points to display the share of subjects selecting each effort level. Theoretical best responses (effort levels that maximize worker’s wage) are reported with a darker color. For instance, we see that around 80% of the high ability workers choose an effort level equal to 1.5 when they are offered a piece rate wage of 0.30, which also happens to be the best response. We find a clear cluster of choices around best responses, both for high ability and low-ability workers. On average, 67% of low-ability workers and 63.5% of high-ability workers choose the best response effort level. These figures increase to 84% and 82% respectively when allowing for 0.5 deviations (+0.5 or -0.5 from the best response). Conversely, on the right-hand side of the graph, we see that principals often expect workers to best respond. They expect such behavior in 66% of the cases (87% when allowing for 0.5 effort deviation), with no significant differences in beliefs across treatment groups. Principals were also fairly accurate at predicting deviations from the best responses. They correctly anticipated that high-ability workers would deviate mostly downward. They expected this type of downward bias for low-ability workers too, but these workers deviated more uniformly either up or down.

### 5.2 Belief-based contract trade-offs

We now show how these beliefs translate into contract characteristics. The need to create pairs of contracts requiring principals to carry out a trade-off between output maximization and egalitarian concerns guided our contract calibration. Figure 2.3 shows how principals’ expectations regarding workers’ effort choices altered these theoretical trade-offs. We interpret the results based on theoretical trade-offs as reduced-form estimates: these trade-offs are exogenous to principals’ characteristics. Belief-based trade-offs show how contracts are perceived in reality by principals. This is valuable because we can rely on the true trade-offs principals believe they are facing when making their choices in order to reduce the noise in our estimations. However, these perceptions may be endogenous to principals’ characteristics. For instance, certain principals may imagine that low-ability workers will decide to sabotage the experiment and choose a zero-effort level. This particular belief may be correlated to some of the principals’ observed or

Figure 2.2: Workers’ stated effort and principals’ expected effort by piece rate wage



*Notes:* The figures on the left-hand side plot the workers’ choices of effort level for each piece rate (on the x-axis) by ability type. The figures on the right-hand side plot principals’ beliefs regarding the effort level chosen by workers for each piece rate. High-ability workers are in red and low-ability workers are in blue. Each dot on the figures on the left-hand side represents the share of workers choosing a particular effort level at a given piece rate wage. For example, we see that around 80% of the high-ability workers choose an effort level equal to 1.5 when they are offered a piece rate wage of 0.30. The size of the dots on the figures on the right-hand side represents the corresponding shares for principals. Hence, we see that around 60% of principals expect high-ability workers to choose an effort level of 2.5 when offered a piece rate of 0.40 ECU. Best responses for each piece rate are highlighted in darker colors. Data for several of the piece rates for principals’ beliefs is missing. We only elicited principals’ beliefs regarding the piece rates that have a chance of being implemented. For instance, the piece rate of 0.45 is never used for the high-ability worker in any of the contracts described in Table 2.2.

unobserved characteristics. In the regressions, we thus present results using both the theoretical and the belief-based trade-offs to account for these two aspects.

On the x-axis of Figure 2.3, we plot the difference in output between Contract 2 (the theoretically output maximizing contract) and Contract 1 (an egalitarian or an equal piece rate Contract). On the y-axis, we plot the difference in inequality between Contract 2 and Contract 1. We measure contract inequality as the difference in wages between the high-ability worker and the low-ability worker. Hence, the y-axis is a difference of a difference and a positive number means that Contract 2 yields more inequality than Contract 1. Similarly, positive numbers on the x-axis mean that Contract 2 yields a larger output, and therefore income, for the principal, relative to Contract 1. The small black dots represent the theoretical trade-offs, those assuming workers’ best respond to piece rate wages. The red and green dots correspond to the belief-based combination of output differences and inequality differences associated with the 16 contract choices facing each principal. We can interpret these dots as the actual trade-offs that principals perceive. The size of the dots represents the frequency of observations, implying the same trade-off. Figure 2.3 shows that many decisions are consistent with our theoretical trade-offs, as expected given the belief-elicitation results in Section 5.1.

Figure 2.3: Principals’ belief-based contract trade-offs



*Notes:* The figure plots the trade-off that principals believe must be made. The y-axis shows the difference in inequality between both contracts, and the inequality of a contract is measured by the high-ability worker’s wage minus the low-ability worker’s wage. Hence, Contract 2 becomes increasingly unequal relative to Contract 1 as we move up the y-axis. The x-axis is the difference in output between contracts. Contract 2 becomes more efficient relative to Contract 1 as we move to the right of the plot. The size of the dots represents the frequency of choices implying the same trade-off. Black dots identify the theoretical trade-offs assuming best responses and are identical to those shown on Figure 2.1. Green dots show beliefs when there is a trade-off between output and equality, and red dots show cases in which one contract is both output-maximizing and egalitarian given the principal’s beliefs (no trade-off).

We further classify trade-offs into two types. In green, we identify all the belief-based contract decisions that generate a trade-off between equality and output. In red, we plot decisions for which one of the contracts yields both a larger output and a lower inequality level. 32% of the

decisions fall in the red category and do not generate any particular trade-off for people who care about output and want to reduce inequality. However, we do not assume these cases to be irrelevant. For some subjects, it may be fair to over-compensate the high-ability worker. In this case, both inequality and output-maximization would be desirable outcomes and the red dots would represent a real trade-off for these subjects. The finite mixture model can be used to test whether such behavior is common in the data. For that reason, we retain the red decisions in our estimation.

That being said, certain observations remain problematic as the implied trade-offs are too large and constitute outliers. These extreme cases must be discarded in order to avoid distorting our estimates, especially with the finite mixture model. We discard observations for which the difference in output between both contracts is greater than 100 or smaller than -100 (58 out of 1808 observations are deleted). The descriptive results of Section 5.3 are barely sensitive to the inclusion or exclusion of these observations because we show mean contract choices by trade-off brackets. Extreme trade-offs only distort the mean of the far-left-hand and far-right-hand brackets, not the intermediate brackets. However, in the finite mixture model, trade-offs directly enter the objective function and the estimation is quite sensitive to these outliers, though the results remain qualitatively the same. We come back to the issue of outliers in detail in the relevant sections below.

### 5.3 Principals’ choices

We now describe the pattern of choices across treatment groups. The y-axis of Figure 2.4 shows the share of cases in which the most egalitarian contract of the pair is selected. We plot this share by the size of the trade-off: Contract 2 increases in relative to Contract 1 as we move to the right of the graph. Spectator’s choices are plotted with a solid blue line, while Stakeholders’ choices are shown with a dotted dark blue line.

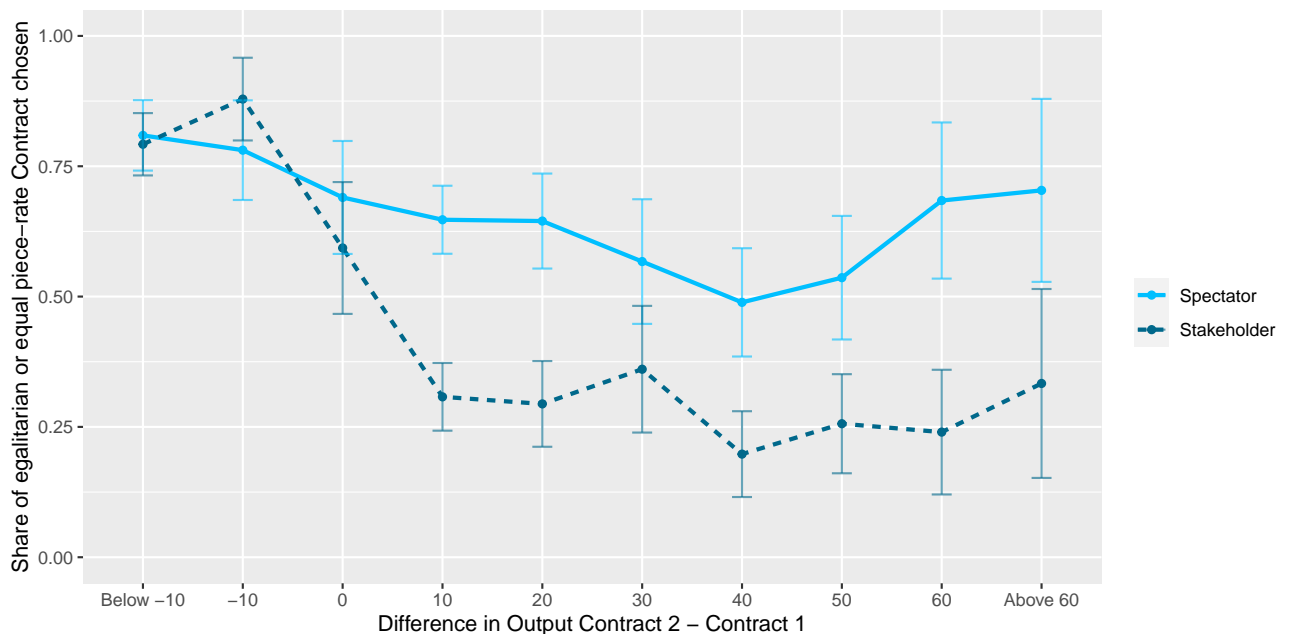
Overall, we find that, on average, both treatment groups compress wages to a certain extent, given that for all trade-offs, the share of the inequality-minimizing Contract 1 decisions is always significantly different from 0. This confirms our hypothesis that, generally speaking, principals demand a reduction in inequality across workers. Now turning to differences across treatment groups, we find that Spectators are more likely than Stakeholders to choose an egalitarian contract. Interestingly, when Stakeholders do not face any trade-offs (differences in output between both contracts is 0 or even negative), then the behaviors of the treatment groups become indistinguishable. This suggests that Stakeholders are sensitive to the size of the stakes. This is further confirmed when examining their choices at the intensive margin. Stakeholders are increasingly unlikely to choose an egalitarian Contract 1 as Contract 2 increases in output in relation to Contract 1. On the contrary, Spectators seem less sensitive to output differences.

Furthermore, the figure captures a slightly U-shaped pattern, indicating that when reaching a difference in output of about 40, the share of egalitarian contracts is not further declining. This can have two reasons: First, by design, contracts that feature a high difference in output are also characterized by high levels of inequality. This concern may lead to a rejection of contracts that have high output differentials, due to concerns for the large inequality they induce. In

the regressions discussed below, we find evidence that our principals are indeed attentive to inequality differentials after controlling for output differences across contracts. Second, some principals may prefer to redistribute income at all costs. Hence, a share of principals may always prefer the egalitarian contract and this can explain the flattening of the curve after the initial decrease. In Section 6, we show that a substantial share of principals, in particular spectators, are characterized by purely egalitarian choices.

Note that the outliers we described in Section 5.2 can only affect the first and end points of the graph (very low and very high expected difference in output). Plotting the same graph without the outliers barely affects the results. If anything, the share of Contract 1 is lower for a high difference in output.

Figure 2.4: Principals’ contract choices by treatment groups



Notes: the Figure shows the share of observations in which the most egalitarian contract of the pair is selected (either an egalitarian or equal piece rate contract). We calculate these shares by output trade-off, i.e. the difference in output between Contract 2 and Contract 1. The solid blue line represents the choices of the Spectator group and the dotted dark blue line shows the choices of the Stakeholder group. The measures are calculated using principals’ beliefs on workers’ behavior. The same figure using belief-based data is Figure A4 in the appendix. We show 95% confidence intervals for the shares.

We can also use a regression analysis to characterize principals’ choices. Table 2.3 regresses the choice of Contract 2 on a binary Stakeholder treatment variable (characterizing the extensive margin trade-off), a dummy “1 is equal piece rate” indicating that the alternative (Contract 1) is an equal piece rate contract (rather than an egalitarian contract), and the difference in output and inequality between both contracts. The last two variables characterize an intensive-margin trade-off between output and inequality. We interact these variables with the Stakeholder dummy to test whether the intensive margin treatment has a differential effect across Spectators and Stakeholders. Columns (2) and (4) additionally control for whether the decision is an egalitarian vs. equal piece-rate choice. We include these dummies because Choices 15 and 16 involve a *direct* choice between an equal piece rate and an egalitarian contract, and behavior in these decisions may not be captured by the difference in inequality or difference in output.

Beyond that, the equal piece rate contract is Contract 2 in this case, and is not picked up by the equal piece rate dummy.

The first two columns of Table 2.3 calculate these trade-offs assuming that workers best respond (theoretical trade-offs), which can be interpreted as reduced-form estimates and, importantly, they are robust to any endogeneity in beliefs. The drawback of these measures is that they may be less precise given that principals may expect deviations from the best responses, and therefore a quite different trade-off in reality. Columns (3) and (4) show the results using belief-based trade-offs. The fit is better for the regressions using the belief-based trade-off (the  $R^2$  rises from about 0.1 to 0.17). This indicates that beliefs capture meaningful variations and reduce measurement error in the trade-off principals really face.

The results in Table 2.3 show that principals are on average significantly more willing to choose a contract if it is expected to yield a larger output relative to its alternative. The decreasing slope in Figure 2.4 captures this significant effect of the output gap on the Choice probability. This applies to Stakeholders and Spectators alike, but Stakeholders are even more sensitive to this trade-off relative to Spectators (positive and significant interaction term at the 1% level for belief-based regressions). The significant and positive main effect of  $\frac{\Delta(\text{Output 2 and 1})}{10}$  indicates that even Spectators want to improve output, on average. Therefore, principals are intrinsically motivated to maximize output and they still respond to changes in the output gap, even after controlling for differences in inequality. We can interpret this result as a residual effect of identity: even if Spectators have no stakes in the production process, they are placed in a managerial position, which can lead them to care about output anyway. These results hold qualitatively for regressions using beliefs (Columns (3) and (4)), as well as those assuming that agents best-respond to incentives (Columns (1) and (2)).

The first row shows that stakeholders are, on average, 26 percentage points more likely to choose a high-inequality contract (coefficient positive and significant at the 5% level with theoretical trade-offs, and at the 1% level for belief-based regressions). Principals are more likely to accept inequality if they are explicitly incentivized, even after taking into account the expected cost of equality, which characterizes the shift in the intercept of the two curves in Figure 2.4. In Section 6 we will subsequently show that a significantly higher proportion of stakeholders always choose an output-maximizing contract but that no spectators do so. These individuals may characterize the extensive margin differences between the two groups.

Relative inequality between contracts is only a significant predictor if we consider regressions (1)–(3) (significant at the 5 percent level). In these instances, principals are less likely to choose a contract that involves greater inequality after controlling for the difference in output, and this further explains the convexity shown in the plots of Figure 2.4. The average effect becomes insignificant once we control explicitly for a decision being an equal piece-rate vs. egalitarian choice and use belief-based trade-offs, which indicates that this may pick up a peculiarity characterized by these two choices. The interaction term between difference in inequality and the Stakeholder dummy is not significant for both theoretical trade-offs and belief-based trade-offs.

Next, we ask whether equal piece-rate contracts are considered as more attractive than egalitarian contracts by the principals. Our data allows us to study this from two angles.

Table 2.3: Regressions that characterize Contract decisions

	Theoretical trade-offs		Belief-based trade-offs	
	(1)	(2)	(3)	(4)
<b>Dependent variable: Contract 2 (high inequality) was chosen</b>				
Stakeholder	0.177** (0.0837)	0.177** (0.0837)	0.266*** (0.0771)	0.265*** (0.0780)
$\frac{\Delta(\text{Output 2 and 1})}{10}$	0.0467*** (0.0129)	0.0429*** (0.0131)	0.0316*** (0.00667)	0.0302*** (0.00654)
$\frac{\Delta(\text{Output 2 and 1})}{10}$ * Stakeholder	0.0300* (0.0173)	0.0299* (0.0173)	0.0339*** (0.0113)	0.0339*** (0.0112)
$\frac{\Delta(\text{Inequality 2 and 1})}{10}$	-0.0483*** (0.0138)	-0.0358** (0.0142)	-0.0218** (0.00994)	-0.0132 (0.00994)
$\frac{\Delta(\text{Inequality 2 and 1})}{10}$ * Stakeholder	-0.00171 (0.0184)	-0.00165 (0.0184)	-0.0180 (0.0144)	-0.0177 (0.0145)
1 is equal piece rate	-0.0864* (0.0451)	-0.0444 (0.0446)	-0.0263 (0.0415)	0.00490 (0.0408)
1 is equal piece rate * Stakeholder	0.00669 (0.0637)	0.00660 (0.0637)	-0.00421 (0.0543)	-0.00447 (0.0545)
Egalitarian vs. Equal piece-rate		0.0977*** (0.0325)		0.0907** (0.0388)
Constant	0.476*** (0.0873)	0.406*** (0.0871)	0.428*** (0.0793)	0.375*** (0.0788)
Control variables	Yes	Yes	Yes	Yes
Observations	1750	1750	1750	1750
$R^2$	0.102	0.105	0.165	0.167

Standard errors clustered on the subject level in parentheses

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

The specification regresses a dummy indicating the choice of Contract 2 on other Choice characteristics using a linear probability model. This samples excludes observations where the difference in *expected* output is less than or equal to 100. In columns (1) and (2), explanatory variables include a Stakeholder treatment dummy variable, the theoretical difference in output between Contract 2 and 1, the theoretical difference in inequality between Contracts 2 and 1 (both assuming workers’ best responses), a dummy for whether Contract 1 constitutes an equal piece rate contract rather than an egalitarian contract, and the interactions of these variables with the Stakeholder dummy. In columns (3) and (4) principals’ beliefs are used to calculate the difference variables. Columns (2) and (4) add controls for whether the choice was an equal egalitarian vs. equal piece-rate choice. All the specifications include the following controls: female dummy, economics background dummy, whether the subject is currently a student and whether he is currently in a relationship.



First, we ask whether subjects are more willing to trade-off output for a reduction in *piece-rate* inequality compared to their willingness to trade-off output for a reduction in ex-post *income* inequality. We test for this by including a dummy that indicates that Contract 1 was an equal piece-rate contract rather than an egalitarian contract (*1 is equal piece-rate*). The coefficient of this variable indicates whether subjects are more or less likely to choose a contract with higher inequalities if the alternative is an equal piece-rate contract rather than an egalitarian contract after controlling for differences in output and inequality. We further interact this variable with the stakeholder dummy to test whether this sensitivity differs across treatment groups. Second, we ask whether subjects are more or less likely to embrace an equal piece-rate contract if they face a direct choice between an egalitarian and an equal piece-rate contract after controlling for differences in inequality reduction and output. This is the case for Choices 15 and 16, where subjects have the choice between an egalitarian and an equal piece-rate contract. We capture this through the “egalitarian vs. equal piece-rate” dummy.

The low-inequality alternative being an equal piece-rate contract (rather than an egalitarian contract) is not a significant predictor of the principal’s decision once we take into account the characteristics of the contract, such as expected inequality and expected output. This does not mean that principals never choose the equal piece-rate contract; it simply means that they are not more likely to choose an equal piece rate than an egalitarian contract after controlling for differences in output and inequality. This suggests that subjects are equally interested in implementing their preferred outcome rather than treating both agents identically.

This assessment changes for some subjects if we posit an equal piece-rate contract directly against an egalitarian contract. In this case, subjects are indeed significantly more likely to choose the equal piece-rate contract, as suggested by the positive and significant effect of facing an egalitarian vs. an equal piece-rate contract. On average, subjects are 10 percentage points more likely to choose an output maximizing contract if this contract also provides equality in piece-rates compared.

Putting both results together, we can conclude that equal piece-rates are indeed attractive for some principals from a fairness perspective if promoted as a direct alternative to an egalitarian contract, but their willingness to implement an equal piece-rate contract is not different from their willingness to implement an egalitarian contract, as suggested by the insignificant equal piece-rate dummy presented above. Note also that the egalitarian contract remains attractive for around half of the subjects in either case.<sup>18</sup>

To sum up, the pooled results show that principals are increasingly willing to accept inequality as the cost of the egalitarian contract rises. Average sensitivity to difference in output is relatively higher for stakeholders than spectators. Furthermore, Stakeholders are significantly more likely to choose a high inequality – high output contract at any given level, suggesting a strong extensive margin effect of incentives on inequality acceptance. Although making Contract 1 an equal piece rate contract does not seem to affect how principals evaluate these contracts, they are significantly more likely to choose an equal piece rate contract if it is posited against an egalitarian contract.

---

<sup>18</sup>The individual fixed effects regressions in Table A5 suggest that this effect is mainly driven by stakeholders.

Table A5 shows the results for belief-based trade-offs that control for individual fixed effects. This is an additional way to account for individual-specific heterogeneity in beliefs. The results are more or less the same.<sup>19</sup> Table A6 replicates Table 2.3 but includes belief-outliers, i.e. observations where the absolute difference in output is higher than 100, which constitute 3% of the total sample. The results are qualitatively very similar but the interaction term of difference in output and being a stakeholder becomes insignificant and the magnitude of the main effect is attenuated. Given the drop in the  $R^2$  it can be assumed that these differences are largely driven by measurement error in outlier-beliefs and do not reflect systematic variations in behavior.

## 6 Structural Characterization of Distributive Preferences

In this section, we will estimate the distributive preferences of principals and characterize the heterogeneity in these preferences. The goal of this exercise is to perform a counterfactual analysis that will allow us to assess *when* these preferences lead to frictions and inefficiencies. To this end, we posit a simple social preference utility function that captures several motives.

### 6.1 Making distributive decisions ex-ante

Before specifying the actual utility function that we want to estimate, it is worth re-emphasizing the context in which managers make decisions. While most studies on distributional preferences take the ex post perspective – dictators make distributive decisions *after* agents have worked, as in most dictator games with a preceding production stage, e.g. Cappelen et al. (2007), we are taking account of the fact that principals typically make incentivization – and hence distributive – decisions in an uncertain environment, before agents have exerted any effort.<sup>20</sup> It also enables testing for the importance of treating *unequal* agents equally, which has not been explored previously.

We assume that principals make decisions that maximize their expected utility  $E(U(y_p, \pi(e_h(w_h), e_l(w_l)), y_h(e_h(w_h)), y_l(e_l(w_l)), w_h, w_l))$ , where the principal’s income is denoted by  $y_p$  and  $\pi$  is the agents’ joint output, which is a function of  $w_h, w_l$ , the workers’ piece rates, and  $e_h, e_l$  their effort levels (for the low- and high-ability agent respectively). The agents’ ex post income is denoted by  $y_l, y_h$ , and also depends on the piece rates and effort level chosen. This specification enables principals to care about the distribution of income *after* workers have made their effort decisions, i.e. ex post income as a function of expected effort. It also enables principals to care about equality of *procedure*: in this case, principals dislike differences in piece rates. Note that our notion of equal procedure is somewhat different from that considered in previous work studying social preferences in a risky environment (e.g. Brock et al., 2013; Krawczyk, 2010), because agents are not identical to begin with (different ability levels leading to different

<sup>19</sup>Note that there is no need to control for individual fixed effects with theoretical trade-offs since there is no individual-level variation in trade-offs in that case. Theoretical trade-offs are completely exogenous to individual characteristics.

<sup>20</sup>Whether this feature matters is naturally dependent on the nature of the research question. The ex ante perspective is, for example, more relevant in our case where the principal has to decide before the workers have made their effort decisions, than in research where the question asks whether citizens perceive a realized distribution as fair.

effort levels), which implies ex post inequality even when both agents are treated equally with equal piece rates.

## 6.2 Utility function specification

The utility function characterizes principals’ concern about their own income  $y_p$ , total output  $\pi = \pi_h(e_l(w_l)) + \pi_l(e_h(w_h))$ , and the distribution of income ex post between both workers  $y_l(e_l(w_l)), y_h(e_l(w_h))$ . The low-ability worker’s income is denoted as  $y_l(e_l(w_l))$  and the high-ability worker’s income is denoted as  $y_h(e_l(w_h))$ . In the following explanation, we will refer to these incomes as  $y_l$  and  $y_h$  with the dependence on the piece rates dropped for expositional purposes, but the reader should bear in mind that workers’ income is always a function of their piece rate and their subsequent effort decision.

To capture other-regarding preferences in a flexible manner that fits our framework, we assume these preferences to be non-linear for piece rates. The importance principals attach to the high- or low-ability worker’s income depends on which worker is receiving a higher *piece rate*  $(w_h, w_l)$ . This is captured by the indicator variables.

$$E(U) = E\left(y_p + \gamma\pi + [(\alpha * \mathbb{1}(w_h \leq w_l) + \beta * \mathbb{1}(w_h > w_l))y_l - (\alpha * \mathbb{1}(w_h \leq w_l) + \beta * \mathbb{1}(w_h > w_l))y_h]\right) \quad (2.1)$$

Parameters of interest are  $\alpha$  and  $\beta$ , and  $\gamma$ . We measure the extent to which the principal values output on top of profit maximization by  $\gamma$ . This proxies for an intrinsic motivation to maximize profits. The  $\alpha$  and  $\beta$  parameters characterize distributive preferences flexibly, by considering two cases.

- $\alpha$  quantifies the extent to which the principal cares about the low-ability worker relative to the high-ability worker if the latter receives a *lower* piece rate than the former.
- $\beta$  quantifies the opposite scenario, i.e. how much the principal cares about the low-ability worker relative to the high-ability worker if the latter receives a *higher* piece rate than the former.

Note, that we also allow for  $\alpha = \beta$ . This then boils down to a more standard model of inequality aversion. We allow for this discontinuity in order to capture a distinct preference for equal procedure or the acceptance of moderate inequality.<sup>21</sup>

We can identify several cases:

1. **Output oriented**  $\alpha = 0, \beta = 0$ : This principal only cares about the maximization of output. The way income is distributed among workers is irrelevant.

---

<sup>21</sup>Ideally, we should also capture altruistic motives where the manager’s utility increases when the sum of his agents’ payoffs also increases. This would, however, be too difficult to identify along with the other motives used in our data. We have therefore decided to focus on key elements of our design, which are a preference for equal procedure and a preference for equality among agents.

2. **Equal procedure**  $\alpha < 0, \beta > 0$  : This principal attaches positive importance to the high ability worker’s income, when his piece rate is lower than that of the low ability worker, and the principal attaches positive importance to the low-ability worker’s income in the opposite case. Therefore, this principal is averse to inequality in piece rate wages and prefers to treat both agents identically.
3. **Redistributive**  $\alpha > 0, \beta > 0$  : This principal attaches positive importance to the low-ability worker under all circumstances facing principals in our experiment. In our setting, this implies that principals have strong preferences for redistribution from the high- to low-ability worker, and achieve equality ex post.<sup>22</sup>
4. **Inequality-targeting**  $\alpha = 0, \beta > 0$  : This principal is focused on maximizing output if  $w_h \leq w_l$  but she is willing to redistribute as soon as  $w_h > w_l$ . In our experiment, contracts in which  $w_h \leq w_l$  are characterized by relatively low inequality, while it is relatively high for contracts in which  $w_h > w_l$ . Principals with such preferences can therefore be labeled as averse to high inequality but less averse to low inequality.
5. **Rewarding**  $\alpha < 0, \beta < 0$  : This principal attaches positive importance to the high-ability worker under all circumstances facing principals in our experiment. In our setting, this implies that principals strongly prefer giving a higher income to the high-ability agent, even when she is already paid a higher piece rate, and even if this comes at the cost of lowering total output.

**How do these preferences translate into choices within our experiment?** In our experimental design, principals are asked to make multiple decisions between two piece rate contracts. These contracts vary according to whether they come under the  $w_h \leq w_l$  or  $w_h > w_l$  domain. The choice of contract also affects the agents’ (expected) income because they will subsequently work under the chosen contract. Given the evidence presented in Section 5.3, we assume that preferences are defined over distributional outcomes, i.e. the workers’ expected income.

More specifically, we can run through the predicted choice patterns of each case listed above. An output-oriented principal (case 1) will always choose the contract that gives her the highest output. A principal who is interested in equal procedure (case 2) will favor a contract that helps the high- (low-) ability agent in the case of both options being characterized by  $w_h \leq w_l$  ( $w_h > w_l$ ). In the case of one option being in the  $w_h \leq w_l$  domain and the other option being  $w_h > w_l$ , it depends on the relative strength of  $\alpha$  and  $\beta$ , as well as the cost in terms of forgone output. Principals that are characterized by strong redistributive preferences (case 3) prefer the contract that minimizes ex post inequality between both workers. The willingness to forgo output for the sake of redistribution can vary according to who receives the higher piece rate, and is characterized by the magnitude of  $\alpha$  and  $\beta$ . Principals who only care about the relative income of agents if  $w_h > w_l$  (case 4) will choose the output-maximizing contract for all

---

<sup>22</sup>Our experiment does not include cases where the low-ability worker receives *higher* ex post earnings than the high-ability worker.

cases where  $w_h \leq w_l$  in both contract options. They only take distributive consequences into consideration if a contract gives the high-ability agent a higher piece rate. In this case, they will reject contracts if the difference in payoffs becomes too great under a contract in which  $w_h > w_l$ . Finally, “rewarding” principals have a preference for maximizing the income of the high-ability agent relative to that of the low-ability agent. Consequently, they will always choose a contract that gives the high-ability agent a higher piece rate. In our experiment, there are two situations in which such a contract is *not* the output-maximizing contract, under the assumption that workers best-respond.

**Identification of  $\gamma$**  We use our treatment variation to identify  $\gamma$ . Spectators’ own income was kept constant but agents’ joint output varied, while both dimensions were varied for stakeholders. The parameter  $\gamma$  informs us how much less (if  $\gamma < 1$ ), or more (if  $\gamma > 1$ ) Spectators care about output relative to Stakeholders, keeping the other-regarding part of the function constant. This informs us about the relative importance of output once we take away the principals’ extrinsic motives to maximize output. Intuitively, this parameter captures the intrinsic motivation to maximize output. Principals may believe that maximizing output is the managers’ job, as some kind of social norm. Even Spectators may care about output for this reason, even if they have no extrinsic (monetary) incentives to do so. This may be a consequence of the framing of the study, or an identity effect.

We can also characterize differences across treatment groups by estimating a more reduced-form model in which we do not differentiate between intrinsic and extrinsic motives to maximize agents’ joint output.

$$E(U) = E\left(\pi + [(\alpha * \mathbb{1}(w_h \leq w_l) + \beta * \mathbb{1}(w_h > w_l))y_l - (\alpha * \mathbb{1}(w_h \leq w_l) + \beta * \mathbb{1}(w_h > w_l))y_h]\right) \quad (2.2)$$

In this case, the joint output  $\pi$  has different meanings for Stakeholders and Spectators. For the former, it encompasses both intrinsic and extrinsic incentives, while it can only represent intrinsic incentives for the latter. Hence, in this specification, cross-treatment differences can only be evaluated in  $\alpha$  and  $\beta$ .

### 6.3 Pooled results

Table 2.4 focuses on average results for the entire population. Column (1) presents results from a conditional logit model that fits equation (2.1). The results mirror the results from Table 2.3, where we ran similar regressions but without assuming any underlying utility function.  $\gamma$  is significantly different from 0 ( $p < 0.01$ ) but also significantly smaller than 1 (t-test,  $p < 0.01$ ). The fact that  $\gamma$  is smaller than 1 implies that monetary incentives for Stakeholders reduce their intrinsic motivation to increase output. In fact, it reduces the importance they attach to output by more than two thirds. However, the fact that  $\gamma$  is above 0 shows that Spectators still care about output for intrinsic motives.

Table 2.4: Results from a pooled specification

	Model based on equation (2.1)	Model based on equation (2.2)		
	(1) Interaction	(2) All	(3) Stakeholder	(4) Spectator
$\gamma$	0.311** (0.103)	–	–	–
$\alpha$	0.07 (0.055)	0.095 (0.04)	0.05 (0.09)	0.21 (.18)
$\beta$	0.17*** (0.035)	0.26*** (0.04)	0.1* (0.05)	0.55*** (0.13)
$\sigma$	0.08*** (0.013)	0.05*** (0.005)	0.06*** (0.012)	0.033*** (0.008)
Observations	1750	1750	898	852

The parameters are estimated using a conditional logit model. Standard errors are clustered at the subject level using the sandwich formula. Column (1) reports parameters from equation (2.1); column (2) reports parameters from equation (2.2). Columns (3) and (4) use the model based on equation (2.2) for the Stakeholder and Spectator sample separately. Observations are on the subject-choice level. \*\*\* denotes statistical significance at the 1 percent level, \*\* at the 5 percent level, and \* at the 10 percent level.

We can also see that, on average,  $\alpha$  is not significantly greater than zero but  $\beta$  is ( $p < 0.01$ ). This corresponds to the behavior outlined in case 4 (inequality-targeting principals). The non-significant  $\alpha$  suggests that principals are only willing to sacrifice output up to the point at which both agents receive the same piece rate.

Columns (2) to (4) fit a conditional logit model assuming equation (2.2) to be the underlying utility function. Column (2) fits the model by examining the entire sample, and columns (3) and (4) presents results that are based on the Stakeholder and Spectator sample. Comparing  $\beta$  across columns (2) and (3), we can observe that Stakeholders are significantly less concerned about inequality if the high-ability agent is paid a higher piece rate wage, capturing the crowding-out effect.  $\alpha$  is non-significant for both samples, but the point estimate is larger for spectators. The point estimate is estimated relatively imprecisely.

#### 6.4 Characterizing heterogeneity in preferences

The characterization of heterogeneity in preferences within our sample identifies which types of principals are prevalent. We can then make counterfactual analyses to determine how inequality and output vary across types when making changes to the work environment. The idea is that being a redistributive principal leads to significant inefficiencies in our setting, but this may not be so true in a different context. We focus more particularly on the case in which workers stop being neutral and start comparing their own piece rate with a co-workers’ piece rate. We can do this by assuming social preferences in the agents’ utility function. We then simulate the optimality of each type of principal type under this new context.

To identify principals’ distributive types, we fit a finite mixture model on contract choices assuming equation (2.2) to be our underlying utility function. We then observe how principals are sorted into different preference classes as a function of being either a Stakeholder or a Spectator. This approach has the advantage of characterizing heterogeneity in a more comprehensive manner. Finite mixture models (FMM) can be used to characterize heterogeneity in social preferences by grouping subjects into different types. This approach has become increasingly popular in the social preference literature (e.g Cappelen et al., 2007; Bruhin et al., 2018; Sutter et al., 2018) since it is a powerful tool for summarizing the distribution of preferences and relaxing homogeneity assumptions. FMMs are less demanding in terms of data than individual-level estimations of preference parameters, and their predictive properties have been shown to be similar to those of individual estimates (Bruhin et al., 2018). Unfortunately, we cannot specify a finite mixture model assuming equation (2.1) to be the underlying function because we would need within-principal variations in incentives.<sup>23</sup> The framework then allows us to measure how the propensity to care about the well-being of the two agents relative to output changes across treatment groups.

To estimate the parameters of the utility function posited above, we use the random utility model framework for discrete choices introduced by McFadden (1973) but assume that the population is composed of a discrete number of types. In Appendix 2.C, we detail the derivation of the type-specific conditional density  $f_k(\theta, \sigma | X_1, X_2, Choice)$  following McFadden (1973).  $\theta$  is the vector of parameters in the utility function (2.1),  $\sigma$  is a choice-sensitivity parameter,  $X_1$  ( $X_2$ ) is a vector of contract characteristics associated with contract 1 (2), and  $Choice$  is a dummy indicating the decision made by the agent.

The finite mixture model assumes heterogeneity in  $\theta$  and  $\sigma$ . It posits that the population can be categorized into  $K$  preference types, where each type has a distinct parameter vector  $(\theta_k, \sigma_k)$ . Note that the true type membership is not observable. Hence, the model assumes that every subject belongs to type  $k$  with probability  $p_k$  *ex ante*. The individual contribution to the likelihood is a weighted sum over type-specific conditional densities

$$l_i(p_2, \dots, p_K, \theta, \sigma | X_1, X_2, Choice) = \sum_{k=1}^K p_k f_i(\theta_k, \sigma_k | X_1, X_2, Choice)$$

whereby  $p_1 = 1 - \sum_{k=2}^K p_k$ .

The overall log-likelihood function takes the logarithm over  $l_i$  and sums across all  $N$  individuals.

$$ll(p_2, \dots, p_K, \theta, \sigma | X_1, X_2, Choice) = \sum_{i=1}^N \log \left( \sum_{k=1}^K p_k f_i(\theta_k, \sigma_k | X_1, X_2, Choice) \right) \quad (2.3)$$

In our estimation of type-specific parameters of the utility function (2.2), we are interested

---

<sup>23</sup>Without this variation we have to make very strong sorting assumptions. Some Stakeholder principals always choose the contract that maximizes output and therefore their own income. Without within-subject treatment variations in individual incentives, we do not know how behavior within this class changes, i.e. we do not know whether selfish agents are also more likely to care about output for intrinsic reasons.

in documenting how the classes are divided across treatment groups. In other words, we want to characterize the *ex ante* class probability as a function of the treatment group  $T_i$ . This shows how treatment groups are sorted differently into types. To do this, we specify the probability of being a member of class  $k > 1$  using a logit specification where  $\alpha_{i,k}$  determines how much more (or less) likely a subject in the Spectator sample is to be in class  $k$ , relative to being in the Stakeholder sample.

$$p_k = \frac{\exp(\alpha_{0,k} + \alpha_{1,k}T_i)}{1 + \sum_{k=2}^K \exp(\alpha_{0,k} + \alpha_{1,k}T_i)}$$

The number of types must be determined by the researcher and should accurately describe the heterogeneity of the data, without over-specifying the model. We follow Bruhin et al. (2018) in using the normalized entropy criterion (NEC) to determine the optimal number of types.<sup>24</sup>

The NEC measures ambiguity in the *ex post* assignment of individuals to types. We can use Bayes’ rule to estimate the *ex post* probability  $\tau_{i,k}$  that subject  $i$  is in class  $k$ .

$$\tau_{i,k} = \frac{\hat{p}_k f_i(\hat{\theta}_k, \hat{\sigma}_k | X_1, X_2, C)}{\sum_{m=1}^K \hat{p}_m f_i(\hat{\theta}_m, \hat{\sigma}_m | X_1, X_2, C)}$$

Ideally, the aim is to obtain an unambiguous mapping of subjects into types. This implies that  $\tau_{i,k}$  should be either close to 0 or close to 1. The NEC normalizes entropy,  $E(k)$  is close to 0 if all  $\tau_{i,k}$ ’s are close to 0 or 1. If the number of classes leads to an ambiguous mapping of subjects into types,  $\tau_{i,k}$ ’s are closer to 0.5 and  $E(K)$  increases.

$$E(K) = - \sum_{k=1}^K \sum_{i=1}^N \tau_{i,k} \log \tau_{i,k}$$

$$NEC(K) = \frac{E(K)}{\mathcal{H}(K) - \mathcal{H}(1)}$$

To determine the optimal number of types, we compare the NEC for different  $K$  values ( $K > 1$ ) and select the model with the lowest NEC. Note that this cannot exclude the possibility of a model with only one class performing better. Since the NEC cannot be calculated for  $K = 1$  we will fit a model with only one class and then examine whether there is *clear* evidence that a non-negligible proportion of subjects follow a decision rule that is inconsistent with the model implied by the parameters (e.g. selfish behavior, although the model implies strong inequality aversion).

To estimate the finite mixture model, we use the expectation-maximizing (EM) algorithm. The EM-algorithm is a numerical method used to maximize the likelihood function but does not yield standard errors (see McLachlan and Peel, 2000, chapter 2 for a detailed description of how to use the EM algorithm to fit finite mixture models). It is frequently used in the estimation of finite mixture models because gradient-based algorithms tend to suffer from convergence prob-

---

<sup>24</sup>We refer to the discussion and summary of the econometric literature by (Bruhin et al., 2018, p.16) on which criterion is best to determine the optimal  $K$  in a very similar setting.



lems due to the non-linearity of the likelihood function. We follow McLachlan and Peel (2000), p.64, in their procedure for calculating standard errors by bootstrapping them parametrically using 1000 iterations and clustering at the individual level.

## 6.5 Results from the finite mixture model

The FMM characterizes heterogeneity using the value function specified in equation (2.2). It should be interpreted as reduced-form because we bundle intrinsic and extrinsic motivation to maximize output. This approach has the advantage that we can characterize the crowding out of inequality concerns by incentivizing principals based on sorting into classes, conditional on their treatment.

As mentioned above, we use the NEC to select the optimal number of types. The number of classes that yield the lowest NEC is 3. The NEC for the model with two classes is 0.03; it is 0.02 for the model with three classes and 0.07 for the model with four classes. The specification with  $K = 4$  performs clearly worse than the other two specifications, and the specification with  $K = 3$  performs better than the specification with  $K = 2$ .<sup>25</sup> Figure A5 shows that nearly all subjects can be unambiguously assigned to one of the classes based on their behavior, confirming that class-assignment is relatively straightforward under this specification. The FMM results are shown in Table 2.5.

The model yields three classes that can be easily interpreted. The first class attaches no importance to agents’ well-being, irrespective of whether one agent is better or worse off. This class makes up 21% of the overall sample but is exclusively composed of Stakeholders. These principals are *not* willing to pay for a reduction in inequality; they only care about maximizing output and – given that this group is completely composed of Stakeholders – their own income.<sup>26</sup>

The second class of subjects (Intermediate type) has a positive and significant  $\beta$ . This means that they are willing to increase the income of the low ability worker when she receives a lower piece rate than the high-ability worker. The point estimate is significantly lower than that of group (3), therefore their willingness to redistribute in these situations is limited.  $\alpha$  is indistinguishable from 0 but it is estimated relatively imprecisely. However, what we can conclude from this group is (i) that they do care about the distributive consequences of their decisions and (ii) that they are concerned about situations with a very high degree of inequality – situations in which the low-ability agent is strongly disadvantaged relative to the high-ability agent. This group of principals constitutes around 65% of the overall sample. Most of the Spectators (79%) can be classified as Intermediate types and around half of the Stakeholders fall into this category.

Finally, the third class (Strong redistributors) attaches considerable importance to the income of the low-ability worker when her piece rate is higher than that of the high-ability agent, and similarly for cases in which the low-ability agent receives a higher piece rate than the high-ability agent. This group always seeks to increase the low-ability agent’s income. In our

<sup>25</sup>Results for the specifications with  $K=2$  or  $K=4$  are available on request.

<sup>26</sup>We can make this statement because none of the Spectators are sorted into this group.

Table 2.5: Results from the finite mixture model with three classes

	Output maximizers (1)	Intermediate (2)	Strong redistributors (3)
<b>Parameters</b>			
$\alpha$	-0.01 [-0.1, 0.06]	0.04 [-0.11, 0.33]	0.49 [0.37, 0.71]
$\beta$	0.00 [-0.02, 0.03]	0.27 [0.2, 0.33]	0.63 [0.58, 0.77]
$\sigma$	0.47 [0.36, 0.78]	0.03 [0.026, 0.04]	0.27 [0.19, 0.45]
<b>Shares</b>			
Full sample	0.21	0.64	0.15
if Stakeholder	0.42 [0.35, 0.43]	0.49 [0.465, 0.56]	0.09 [0.065, 0.125]
if Spectator	0.00 [0.0, 0.0]	0.79 [0.77, 0.83]	0.21 [0.17, 0.23]

Bootstrapped 95% confidence intervals in squared brackets clustered at the individual level using 1000 iteration (McLachlan and Peel, 2000, p.64). One observation is at the subject-choice level ( $N = 1750$ ). The NEC is 0.02 for a mixture model with three classes.

This table presents results from a finite mixture model outlined in section 6.2. The model uses three discrete classes. The columns separate preferences across the three classes. The first panel displays the parameter across classes and the second panel displays class shares. We only use observations where the difference in output based on elicited beliefs is lower than 100. Table B1 replicates this table using the full sample.

framework, this boils down to a model in which the principal has strong redistributive concerns and wants to minimize inequality as far as possible.<sup>27</sup>

What are the consequences of these preferences on contract choice? To answer this question, we present results from a simulation exercise in Appendix 2.D. There, we simulate the piece-rates chosen by each of the three types in a richer environment. Importantly, we also assume that workers hold social preferences and may reject unequal contracts. Principals are either informed about the social preferences of the workers or they are naive, in the sense that they falsely believe that workers are selfish income maximizers. The simulations yields interesting results and show that principals with egalitarian preferences are not necessarily those that have bad firm outcomes. For example, if agents are selfish, strong redistributors implement contracts that lead to lower profits compared to the contracts implemented by output maximizers or inequality targeters. If agents become averse to inequality that hurts the high ability worker or they care about equality in procedure, inequality targeters behave no different (better) than output maximizing principals that are (not) informed about the preferences of the workers. This is the case, because they now face no conflict between the profit maximizing and the distributive

<sup>27</sup>Note that redistributive contracts do not allow for situations in which the low-ability agent is better off ex post than the high-ability agent. Consequently, we can readily interpret these decisions as redistributive.

motive. A similar argument can be made for redistributive principals who perform as well (better) than (naive) output maximizers. Output maximizers perform more poorly than all other types if they falsely think that agents are selfish. This also shows that the principals who hold distributive preferences use incentivization strategies that are more robust to misperception of workers’ attitudes towards inequality.

Comparing the class shares across treatment groups, we observe that the results show clear crowding out. While virtually none of the Spectators are characterized by the output-maximizing class, we find that 42% of Stakeholders are sorted into this group. We thus show that monetary incentives completely crowd out other-regarding behavior for 42% of principals.

This is on the a stark result. On the one hand side, we show that no principal in our sample only cares about output maximization for purely normative concerns because all spectators care about the distribution of incentives above and beyond their effect on output. On the other hand, we find very strong crowding out as nearly half of all principals do not take into account at all the distributive concerns, once they hold a stake in the outcome of the firm. Nonetheless, more than half of the principals are still willing to sacrifice part of their own income in order to avoid very high inequalities. This indicates that distributive preferences remain strong and relevant on average, mirroring the result from the pooled regression in Section 5.

## 7 Conclusion

Our results suggest that we should rethink how social preferences affect labor market interactions by modeling them under the assumption that other-regarding preferences are important not only to agents, but also to principals. Managers are the decision-makers for wage-allocation schemes and should therefore be a more frequent focus of research, in order to develop a better understanding of the determinants of wage inequality. Our survey evidence shows that even after controlling for a wide array of firm and manager-level characteristics, a significant correlation remains between the implementation of performance pay within firms and managers’ fairness beliefs. Although the existence of other-regarding preferences is well-established in the behavioral economics literature, we show that its realm extends even to situations where output-maximization should be key to survival in a competitive economy.

Our experiment, in a controlled setting, establishes that such a relationship is causal, at least in the context of our experiment, and that principals hold normative distributive preferences that are partially crowded-out by incentive concerns. Extensive margins (irrespective of whether the principal has a monetary stake in the production process) are crucial to understanding wage contract choices. Intensive margins (the size of the trade-off between output and equality) also matter, but to a lesser extent.

External validity is an obvious concern in such kind of experiments. We can worry about the fact that, in real situations, individuals are partly self-selected into managerial positions and their distributive preferences may be one of the factors determining their access to such positions. In our experiment, individuals are randomly selected into the manager position. Our particular problem amounts to the larger issue in the experimental economics literature about

whether experienced professionals behave similarly to traditional lab samples (mostly students), in firm-like experimental games. Fréchette (2015, 2016) review this literature and finds that overall, those two types of populations don’t behave too differently in experimental games such as bargaining games, signaling games and other-regarding games. Cooper et al. (1999) find no differences in the long-run for repeated signaling games across real managers and students in China. Fehr and List (2004) find that CEOs are more trusting and more trustworthy than students in a trust game, but they react in a similar way to the features of the experimental design. This suggests that if the magnitude of the treatment effects may differ across students and managers, the direction of the treatment is probably the same. Future research should generate experimental evidence from the field by eliciting managers’ other-regarding preferences and their beliefs in an incentivized manner, and link them to firm outcomes. Furthermore, it would be of great interest to document how managers sort into different sectors or firms based on their other-regarding preferences.

## 8 References

- Abeler, Johannes, Steffen Altmann, Sebastian Kube, and Matthia Wibrals. Gift exchange and workers' fairness concerns: When equality is unfair. *Journal of the European Economic Association*, 8(6):1299–1324, 2010.
- Alekseev, Aleksandr, Gary Charness, and Uri Gneezy. Experimental methods: When and why contextual instructions are important. *Journal of Economic Behavior and Organization*, 134: 48–59, 2017.
- Almås, Ingvild, Alexander W. Cappelen, and Bertil Tungodden. Cutthroat Capitalism versus Cuddly Socialism: Are Americans More Meritocratic and Efficiency-Seeking than Scandinavians? *Journal of Political Economy*, 128(5):1753–1788, 2020.
- Ashraf, Nava and Oriana Bandiera. Social Incentives in Organizations. *The Annual Review of Economics*, 10:439–463, 2018.
- Balafoutas, Loukas, Martin G. Kocher, Louis Putterman, and Matthias Sutter. Equality, equity and incentives: An experiment. *European Economic Review*, 60:32–51, 2013.
- Bandiera, Oriana, Iwan Barankay, and Imran Rasul. Social preferences and the response to incentives: Evidence from personnel data. *The Quarterly Journal of Economics*, 120(3):917–962, 2005.
- Bandiera, Oriana., Iwan. Barankay, and Imran. Rasul. Incentives for Managers and Inequality among Workers: Evidence from a Firm-Level Experiment. *Quarterly Journal of Economics*, 122(2):729–773, 2007.
- Bartling, Björn and Ferdinand A Von Siemens. The intensity of incentives in firms and markets: Moral hazard with envious agents. *Labour Economics*, 17(3):598–607, 2010.
- Bastos, Paulo and Natália P. Monteiro. Managers and wage policies. *Journal of Economics and Management Strategy*, 20(4):957–984, 2011.
- Bellemare, Charles and Bruce Shearer. Gift giving and worker productivity: Evidence from a firm-level experiment. *Games and Economic Behavior*, 67(1):233–244, 2009.
- Bertrand, Marianne and Antoinette Schoar. Managing with style: The effect of managers on firm policies. *Quarterly Journal of Economics*, 118(4):1169–1208, 2003.
- Bloom, Nicholas, Renata Lemos, Raffaella Sadun, Daniela Scur, and John Van Reenen. Jeeafbbva Lecture 2013: The New Empirical Economics of Management. *Journal of the European Economic Association*, 12(4):835–876, 2014.
- Bolton, Gary E and Peter Werner. The influence of potential on wages and effort. *Experimental Economics*, 19:535–561, 2016.

- Brandts, Jordi, José M. Ortiz, and Carles Solà Belda. Distributional Concerns in Managers' Compensation Schemes for Heterogeneous Workers: Experimental Evidence. *Review of Behavioral Economics*, 6(3):193–218, 2019.
- Breza, Emily, Supreet Kaur, and Yogita Shamdasani. The morale effects of pay inequality. *The Quarterly Journal of Economics*, 133(2):611–663, 2018.
- Brock, J. Michelle, Andreas Lange, and Erkut Y Ozbay. Dictating the Risk: Experimental Evidence on Giving in Risky Environments. *American Economic Review*, 103(1):415–437, 2013.
- Bruhin, Adrian, Ernst Fehr, and Daniel Schunk. The many faces of human sociality: Uncovering the distribution and stability of social preferences. *Journal of the European Economic Association*, 17(4):1025–1069, 2018.
- Cabrales, Antonio, Raffaele Miniaci, Marco Piovesan, and Giovanni Ponti. Social preferences and strategic uncertainty: An experiment on markets and contracts. *American Economic Review*, 100(5):2261–2278, 2010.
- Cappelen, Alexander W, Astri Drange Hole, Erik Ø Sørensen, and Bertil Tungodden. The pluralism of fairness ideals: An experimental approach. *The American Economic Review*, 97(3):818–827, 2007.
- Card, David, Alexandre Mas, Enrico Moretti, and Emmanuel Saez. Inequality at work: The effect of peer salaries on job satisfaction. *The American Economic Review*, 102(6):2981–3003, 2012.
- Charness, Gary and Peter Kuhn. Does Pay Inequality Affect Worker Effort? Experimental Evidence. *Journal of Labor Economics*, 25(4):693–723, 2007.
- Charness, Gary, David Masclet, and Marie Claire Villeval. The dark side of competition for status. *Management Science*, 60(1):38–55, 2013.
- Clark, Andrew E, David Masclet, and Marie Claire Villeval. Effort and comparison income: Experimental and survey evidence. *Industrial and Labor Relations Review*, 63(3):407–426, 2010.
- Cohn, Alain, Ernst Fehr, and Lorenz Goette. Fair wages and effort provision: Combining evidence from a choice experiment and a field experiment. *Management Science*, 61(8):1777–1794, 2014.
- Cooper, David J, John H Kagel, Wei Lo, and Qing Liang Gu. Gaming Against Managers in Incentive Systems: Experimental Results with Chinese Students and Chinese Managers. *American Economic Review*, 89(4):781–804, 1999.
- Cronqvist, Henrik and Frank Yu. Shaped by their daughters: Executives, female socialization, and corporate social responsibility. *Journal of Financial Economics*, 126(3):543–562, 2017.

- Croson, Rachel and Uri Gneezy. Gender differences in preferences. *Journal of Economic literature*, 47(2):448–474, 2009.
- DellaVigna, Stefano, John A List, Ulrike Malmendier, and Gautam Rao. Estimating social preferences and gift exchange at work. *Mimeo*, 2016.
- Eisenkopf, Gerald, Urs Fischbacher, and Franziska Föllmi-Heusi. Unequal opportunities and distributive justice. *Journal of Economic Behavior and Organization*, 93, 2013.
- Engelmann, Dirk and Martin Strobel. Inequality aversion, efficiency, and maximin preferences in simple distribution experiments. *The American Economic Review*, 94(4):857–869, 2004.
- Englmaier, Florian and Achim Wambach. Optimal incentive contracts under inequity aversion. *Games and Economic Behavior*, 69(2):312–328, 2010.
- Fehr, E., G. Kirchsteiger, and A. Riedl. Does Fairness Prevent Market Clearing? An Experimental Investigation. *The Quarterly Journal of Economics*, 108(2):437–459, 1993.
- Fehr, Ernst and John A. List. The Hidden Costs and Returns of Incentives—Trust and Trustworthiness among Ceos. *Journal of the European Economic Association*, 2(5):743–771, 2004.
- Fehr, Ernst and Klaus M. Schmidt. Fairness and Incentives in a Multi-Task Principal-Agent Model. *The Scandinavian Journal of Economics*, 106(3):453–474, 2004.
- Fehr, Ernst, Alexander Klein, and Klaus M Schmidt. Fairness and contract design. *Econometrica*, 75(1):121–154, 2007.
- Fischbacher, Urs. Z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 2007.
- Fisman, Raymond, Shachar Kariv, and Daniel Markovits. Individual preferences for giving. *American Economic Review*, 97(5):1858–1876, 2007.
- Fréchette, Guillaume R. Laboratory Experiments: Professionals Versus Students. In Fréchette, Guillaume R. and Andrew Schotter, editors, *Handbook of Experimental Economic Methodology*, pages 360–390. Oxford University Press, 2015.
- Fréchette, Guillaume R. Experimental Economics across Subject Populations. In Kagel, John H. and Alvin E. Roth, editors, *The Handbook of Experimental Economics, Volume Two*. Princeton University Press, Princeton, 2016.
- Gagnon, Nickolas, Kristof Bosmans, and Arno Riedl. The Effect of Unfair Chances and Gender Discrimination on Labor Supply. *Working Paper*, 2020.
- Greiner, Ben. Subject Pool Recruitment Procedures: Organizing Experiments with ORSEE. *Journal of the Economic Science Association*, 1(1):114–125, 2015.

- Gross, Till, Christopher Guo, and Gary Charness. Merit pay and wage compression with productivity differences and uncertainty. *Journal of Economic Behavior and Organization*, 117: 233–247, 2015.
- Kocher, Martin G., Ganna Pogrebna, and Matthias Sutter. Other-regarding preferences and management styles. *Journal of Economic Behavior and Organization*, 88:109–132, 2013.
- Konow, James. Fair Shares: Accountability and Cognitive Dissonance in Allocation Decisions. *American Economic Review*, 90(4):1072–1091, 2000.
- Krawczyk, Michal. A glimpse through the veil of ignorance: Equality of opportunity and support for redistribution. *Journal of Public Economics*, 94(1):131–141, 2010.
- McFadden, Daniel. Conditional Logit Analysis of Qualitative Choice Behavior. In Zarembka, P., editor, *Frontiers in Econometrics*, volume 1, pages 105–142. Academic Press New York, 1973.
- McLachlan, Geoffrey J. and David Peel. *Finite Mixture Models*. Wiley Series in Probability and Statistics. Applied Probability and Statistics Section. Wiley, New York, 2000.
- Sutter, Matthias, Francesco Feri, Daniela Glätzle-Rützler, Martin G. Kocher, Peter Martinsson, and Katarina Nordblom. Social preferences in childhood and adolescence. A large-scale experiment to estimate primary and secondary motivations. *Journal of Economic Behavior and Organization*, 146:16–30, 2018.
- von Siemens, Ferdinand A. Heterogeneous social preferences, screening, and employment contracts. *Oxford Economic Papers*, 63(3):499–522, 2011.



# Appendices

## 2.A Tables and Figures

Table A1: Description of the main variables used in the REPOSE survey

Variable name	Original question in the survey	Scale used in the analysis
<b>Dependent variables</b>		
White-collar individualized wage raise	<i>Did white-collar workers benefited from individualized wage raises</i>	0 = No; 1 = Yes
White-collar performance-based bonus	<i>Did white-collar workers benefited from bonuses related to individual performance?</i>	0 = No; 1 = Yes
Blue-collar individualized wage raise	<i>Did non-white collar workers benefited from individualized wage raises</i>	0 = No; 1 = Yes
Blue-collar performance-based bonus	<i>Did non-white collar workers benefited from bonuses related to individual performance?</i>	0 = No; 1 = Yes
<b>Main explanatory variables</b>	<i>With regard to individualized wages (regardless of whether it is implemented in your firm), what do you think about the following assertions? Individualized wages designate a one-off wage-raise or bonus policy that is differentiated across employees and depends on individual assessments</i>	
Individualized wage raises are unfair.	<i>They are fairer than undifferentiated increases. 1= Completely agree; 2=Somewhat agree; 3=Somewhat disagree; 4=Completely disagree</i>	1=Disagree ; 0 = Agree
Individualized wage raises create tension	<i>They create tensions that could undermine collective functioning. 1= Completely agree; 2=Somewhat agree; 3=Somewhat disagree; 4=Completely disagree</i>	1=Disagree ; 0 = Agree
Individualized wage raises are subjective	<i>They cannot be based on objective criteria. 1= Completely agree; 2=Somewhat agree; 3=Somewhat disagree; 4=Completely disagree</i>	1=Disagree ; 0 = Agree
Individualized wage raises motivate	<i>They motivate employees. 1= Completely agree; 2=Somewhat agree; 3=Somewhat disagree; 4=Completely disagree</i>	1=Disagree ; 0 = Agree

The control variables we use can be classified in two types:

- **Individual controls:** they correspond to the individual-level characteristics of the managers who answered the survey. We control for gender, two education dummies (whether

the respondent has at least a high-school diploma, and whether the respondent has partially or totally completed undergraduate studies), the position held by the manager within the firm (executive manager, local manager or human resources manager).<sup>28</sup>

- **Firm controls:** five dummies for the size of the plant (below 30 employees, 20-49, 50-99, 100-199, 200-499), four dummies for the age of the plant (under 5 years old, 5-9, 10-19, 20-49), four dummies for the main type of employee working in the firm (blue-collar worker, employee, technicians, sales, white-collar is omitted), the proportion of people on short-term contracts, whether the firm uses interim contracts, whether the firm follows a 35-hour-per-week system, whether it has an independent status (i.e. not belonging to a larger firm), four dummies for the share of unionized people in the firm (0%, 1 to 5%, 5 to 10%, 11% to 20%).<sup>29</sup>

Table A2: Performance-based bonuses and managers' distributive preferences

	White-collar workers			Blue-collar workers		
	(1)	(2)	(3)	(4)	(5)	(6)
<b>Dep var = Did white/blue-collar workers benefited from bonuses based on individual performance?</b>						
Individualized wage raises are unfair	-0.181*** (0.0135)	-0.111*** (0.0149)	-0.0834*** (0.0160)	-0.131*** (0.0150)	-0.0810*** (0.0163)	-0.0606*** (0.0182)
Individualized wage raises create tension		-0.110*** (0.0110)	-0.0613*** (0.0119)		-0.0624*** (0.0117)	-0.0445*** (0.0133)
Individualized wage raises motivate		0.0382** (0.0181)	0.0104 (0.0190)		0.0421** (0.0194)	0.0298 (0.0213)
Individualized wage raises are subjective		-0.0940*** (0.0128)	-0.0489*** (0.0136)		-0.0712*** (0.0139)	-0.0775*** (0.0155)
Wave dummy	Yes	Yes	Yes	Yes	Yes	Yes
Individual controls	No	No	Yes	No	No	Yes
Firm controls	No	No	Yes	No	No	Yes
Observations	7689	7587	5785	8152	8046	6162
Pseudo $R^2$	0.020	0.042	0.140	0.009	0.016	0.040

Robust standard errors in parentheses

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

The Table shows marginal effects from logit specifications. We regress a binary variable for whether white-collar workers benefited from bonuses based on individual performance in columns (1) to (3) (blue-collar workers in columns (4) to (6)) on binary variables for whether the manager answering the survey thinks that individualized wage raises are unfair, whether they create tensions, motivate, or are subjective. All regressions include a 2017 wave dummy. We additionally control for individual and firms controls in columns (3) and (6). See Appendix Table A1 for a description of all the variables.

<sup>28</sup>As age information is missing for the 2011 wave, we do not control for it. Our results hold true for the 2017 wave with age dummy controls.

<sup>29</sup>We use the data reported by the manager answering the survey. This information is sometimes missing, hence the drop in observations when we add firm controls. Our results hold true when we remove these union dummies.

Table A3: Summary statistics Agents vs principal

Variable	(1) Workers	(2) principals	(3) Diff.	(4) Obs.
Female	0.500 (0.501)	0.434 (0.498)	-0.066 (0.058)	339
Age	25.468 (5.356)	25.514 (4.154)	0.046 (0.593)	325
In a relationship	0.346 (0.477)	0.343 (0.477)	-0.003 (0.056)	325
Student	0.615 (0.488)	0.596 (0.493)	-0.018 (0.057)	327
Econ student	0.314 (0.465)	0.310 (0.464)	-0.004 (0.054)	339
Master or PhD education level	0.438 (0.497)	0.434 (0.498)	-0.004 (0.057)	339
Observations	226	113		339

Table A4: Summary statistics principals Spectator vs Stakeholder

Variable	(1) Spectators	(2) Stakeholders	(3) Diff.	(4) Obs.
Female	0.345 (0.480)	0.517 (0.504)	0.172 (0.093)*	113
Age	25.420 (3.923)	25.600 (4.387)	0.180 (0.815)	105
In a relationship	0.377 (0.489)	0.309 (0.466)	-0.068 (0.092)	108
Student	0.667 (0.476)	0.527 (0.504)	-0.139 (0.094)	109
Econ student	0.364 (0.485)	0.259 (0.442)	-0.105 (0.087)	113
Master or PhD education level	0.436 (0.501)	0.431 (0.500)	-0.005 (0.094)	113
Observations	55	58		113

Table A5: Regressions that characterize contract decisions using belief-based trade-offs and individual fixed effects

	Stakeholders		Spectators	
	(1)	(2)	(3)	(4)
<b>Dependent variable: Contract 2 (high inequality) was chosen</b>				
A is equal piece rate	-0.0245 (0.0343)	0.0208 (0.0363)	-0.0310 (0.0451)	0.000998 (0.0445)
$\frac{\Delta(\text{Expected Output 2 and 1})}{10}$	0.0700*** (0.00854)	0.0674*** (0.00834)	0.0299*** (0.00794)	0.0284*** (0.00777)
$\frac{\Delta(\text{Expected Inequality 2 and 1})}{10}$	-0.0334*** (0.00887)	-0.0193* (0.00988)	-0.0241** (0.0112)	-0.0147 (0.0114)
Egalitarian vs. equal piece-rate		0.123*** (0.0437)		0.0875* (0.0503)
Constant	0.609*** (0.0413)	0.527*** (0.0493)	0.410*** (0.0508)	0.354*** (0.0538)
Fixed effects	Yes	Yes	Yes	Yes
N	898	898	852	852

Standard errors clustered on the subject level in parentheses

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

The specification regresses a dummy indicating the choice of Contract 2 on other Choice characteristics using a linear probability model. Explanatory variables include the expected difference in output between Contract 2 and 1, the expected difference in inequality between Contracts 2 and 1 (both based on principals' beliefs) and a dummy for whether Contract 1 constitutes an equal piece rate contract rather than an egalitarian contract. Columns (2) and (4) further add a control dummy for whether the current observation is an egalitarian vs. equal piece-rate choice. All columns include individual fixed effects.

Table A6: Regressions that characterize contract decisions including outliers

	Theoretical trade-offs		Belief-based trade-offs	
	(1)	(2)	(3)	(4)
<b>Dependent variable: Contract 2 (high inequality) was chosen</b>				
Stakeholder	0.190** (0.0830)	0.190** (0.0831)	0.209*** (0.0793)	0.208** (0.0800)
$\frac{\Delta(\text{Output 2 and 1})}{10}$	0.0452*** (0.0127)	0.0415*** (0.0127)	0.0171** (0.00668)	0.0154** (0.00659)
$\frac{\Delta(\text{Output 2 and 1})}{10}$ * Stakeholder	0.0317* (0.0167)	0.0317* (0.0167)	0.0183 (0.0124)	0.0190 (0.0121)
$\frac{\Delta(\text{Inequality 2 and 1})}{10}$	-0.0468*** (0.0134)	-0.0349** (0.0136)	-0.0201** (0.00952)	-0.00972 (0.00957)
$\frac{\Delta(\text{Inequality 2 and 1})}{10}$ * Stakeholder	-0.00727 (0.0180)	-0.00727 (0.0180)	0.000771 (0.0150)	0.000661 (0.0150)
A is equal piece rate	-0.0876** (0.0437)	-0.0475 (0.0435)	-0.0233 (0.0424)	0.0145 (0.0423)
A is equal piece rate * Stakeholder	0.0105 (0.0621)	0.0105 (0.0621)	0.0333 (0.0564)	0.0326 (0.0565)
Egalitarian vs. equal piece-rate		0.0941*** (0.0321)		0.112*** (0.0403)
Constant	0.483*** (0.0856)	0.416*** (0.0856)	0.441*** (0.0824)	0.378*** (0.0844)
Control variables	Yes	Yes	Yes	Yes
Observations	1808	1808	1808	1808
$R^2$	0.100	0.103	0.123	0.127

Standard errors clustered on the subject level in parentheses

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

The specification regresses a dummy indicating the choice of Contract 2 on other Choice characteristics using a linear probability model. The sample excludes observations where the difference in output is lower than or equal to zero or higher than a 100. In columns (1) and (2), explanatory variables include a Stakeholder treatment dummy variable, the theoretical difference in output between Contract 2 and 1, the theoretical difference in inequality between Contracts 2 and 1 (both assuming workers' best responses), a dummy for whether Contract 1 constitutes an equal piece rate contract rather than an egalitarian contract and the interactions of these variables with the Stakeholder dummy. In columns (3) and (4), principals' beliefs are used to calculate the difference variables. Columns (2) and (4) add controls for whether the current observation is an egalitarian vs. equal piece-rate choice. All the specifications include the following controls: female dummy, economics background dummy, whether the subject is currently a student and whether he is currently in a relationship.

Figure A1: The production and cost function per effort level and agent.

Worker A											
Effort level	0	0.5	1	1.5	2	2.5	3	3.5	4	4.5	5
Production	0	30	60	90	120	150	180	210	240	270	300
Effort cost	0	1	6	14	23	35	48	64	81	100	120

Worker B											
Effort level	0	0.5	1	1.5	2	2.5	3	3.5	4	4.5	5
Production	0	20	40	60	80	100	120	140	160	180	200
Effort cost	0	1	6	14	23	35	48	64	81	100	120

Figure A2: Screenshot of a decision made by agent B.

Part 4: real choices Real choice number 5

**Choice of effort level with a piece rate of 0.5 EU**

Effort level	0	0.5	1	1.5	2	2.5	3	3.5	4	4.5	5
Production	0	30	60	90	120	150	180	210	240	270	300
Effort cost	0	1	6	13.5	23	34.5	48	64	81	100	120
Variable income of the worker (net of effort cost) with a piece rate of 0.5 ECU	0	14	24	31.5	37	40.5	42	41	39	35	30

Which effort level do you choose with a piece rate of 0.5 ECU?

Make an effort choice:

0  
  0.5  
  1  
  1.5  
  2  
  2.5  
  3  
  3.5  
  4  
  4.5  
  5

**Your decision:**

Effort level: 3  
 Production with this decision: 180  
 Piece rate: 0.5  
 Cost with this decision: 48  
 Net income with this decision: 42

*Notes:* this is a translated version of the experiment. Original screenshots are available upon request. We recreated the exact same display as the French version.

Figure A3: Screenshot of a decision made by the principal.

Part 4: real choices
**Real choice number 1**
Description of the table

You have been matched to the following employees. Here are the tables summarizing their characteristics.  
Which piece rates do you choose?

Calculator

*Contract 1 (0.4 for employee A and 0.6 for employee B) or Contract 2 (0.5 for employee A and 0.5 for employee B)*

Worker A											
Effort level	0	0,5	1	1,5	2	2,5	3	3,5	4	4,5	5
Variable income of Worker A (net of effort cost) with a piece rate of 0.4 ECU	0	11	18	23	25	26	24	20	15	8	0
Variable income of Worker A (net of effort cost) with a piece rate of 0.5 ECU	0	14	24	32	37	41	42	41	39	35	30
Your income	0	15	30	45	60	75	90	105	120	135	150

Worker B											
Effort level	0	0,5	1	1,5	2	2,5	3	3,5	4	4,5	5
Variable income of Worker A (net of effort cost) with a piece rate of 0.6 ECU	0	11	18	23	25	26	24	20	15	8	0
Variable income of Worker A (net of effort cost) with a piece rate of 0.5 ECU	0	9	14	17	17	16	12	6	-1	-10	-20
Your income	0	10	20	30	40	50	60	70	80	90	100

Remember that employees receive in addition a fixed income of 90 ECU for their participation

Make a choice between both contracts (click on each of the contracts to see a simulation of the consequences of your choice)

Contract 1  0.40

Contract 2  0.50

Worker A    Worker B

0.40        0.60

0.50        0.50

**Simulation of the consequences of Contract 1, based on your anticipation of the behavior of both workers**

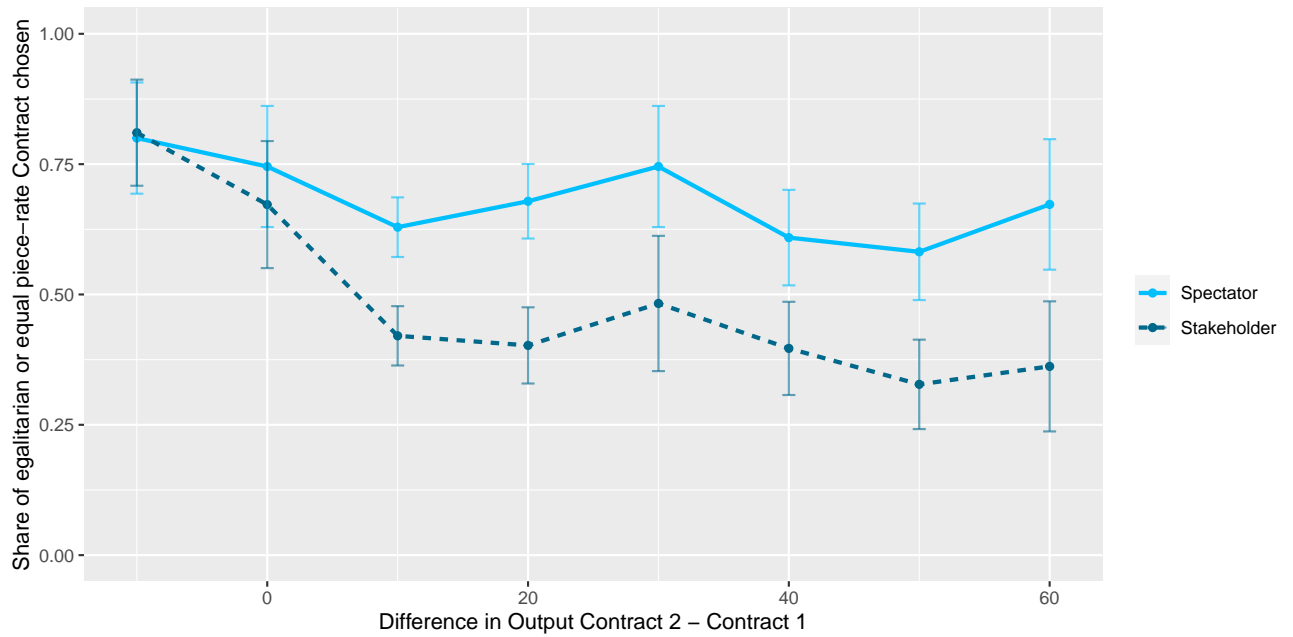
Consequences for both employees		Consequences for yourself	
	Worker A	Worker B	
Effort choice (according to your anticipations)	2.5	2.5	Your income with Contract 1 would be equal to 125 ECU according to the effort level you anticipate.
Production (computed based on your effort anticipations)	150	100	
Variable income of the worker (net of effort cost) (computed based on your effort anticipations)	25.5	25.5	

Click here to confirm your choice after taking note of both simulations OK

Notes: this is a translated version of the experiment. Original screenshots are available upon request. We recreated the exact same display as the French version.

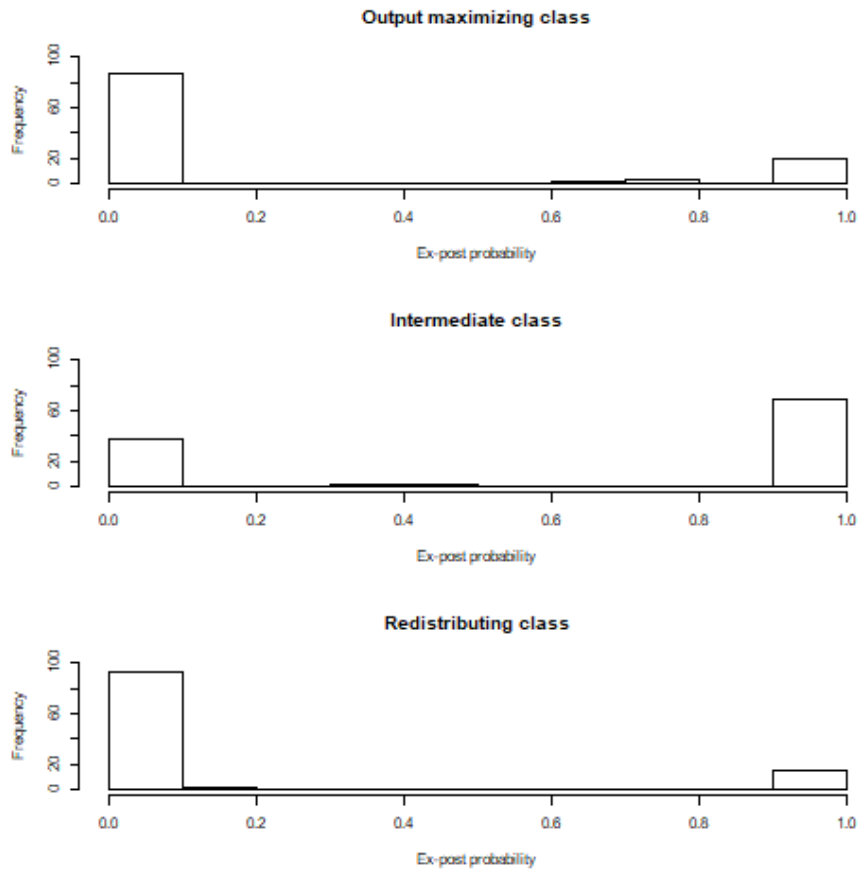


Figure A4: Principals' contract choices by treatment groups



*Notes:* the Figure shows the share of observations in which the most egalitarian contract of the pair is selected (either an egalitarian or equal piece rate contract). This corresponds to Contract 1 in all cases. We calculate these shares by output trade-off, i.e. the difference in output between Contract 2 and Contract 1. The solid blue line represents the choices of the Spectator group and the dotted dark blue line shows the choices of the Stakeholder group. The measures are calculated assuming that agents best respond. The same figure using belief-based data is Figure 2.4 in the main text. We show 95% confidence intervals for the shares.

Figure A5: Distribution of individual ex post probabilities to be part of a given class



*Notes:* These histograms characterize the distribution of ex post class probabilities on the individual level. The x-axis characterizes the ex post probability and the y axis reports the number of subjects within a bin. The first histogram shows the distribution of the ex post probability to be in the output maximizing class, the second to be in the redistributing class, and the third to be in the intermediate class.

## 2.B Robustness Checks

Table B1: Results from the finite mixture model with three classes including outlier beliefs

	Output maximizers (1)	Intermediate (2)	Strong redistributors (3)
<b>Parameters</b>			
$\beta$	0.00 [0.07]	0.36 [0.06]	10.38 [2.07]
$\sigma$	0.28 [0.07]	0.03 [0.012]	0.003 [0.04]
<b>Shares</b>			
Full sample	0.22	0.63	0.145
if Stakeholder	0.36 [0.02]	0.44 [0.04]	0.19 [0.03]
if Spectator	0.00 [0.01]	0.47 [0.07]	0.53 [0.07]

Standard errors from gradient based estimation in parentheses.

Bootstrapped standard errors (1000 iterations) in squared brackets following (McLachlan and Peel, 2000, p.64) .

This table presents results from a finite mixture model outlined in section 6.2. The model uses three discrete classes. Columns separate preferences across the three classes. The first panel displays the parameter across classes and the second panel displays class shares. Includes belief outliers.

Table B1 replicates Table 2.5 using all 1808 subject-choice observations. Compared to the table posted previously, we have a strong redistributor class that became extremely noisy given a low  $\sigma$ . Furthermore, the *NEC* became significantly higher (0.09), indicating a worse fit by the data.

What explains these differences given that we have only deleted 58 observations, i.e. 3% of the overall sample? As mentioned above, trade-offs directly enter the objective function in the framework of a conditional logit model and the estimation is moderately sensitive to these outliers because they may imply a relatively high willingness to pay for the reduction of inequalities. Given that we are certainly measuring these beliefs with noise, the subjects may not have always behaved in accordance with the model because we do not observe the “true” incentives these subjects faced. The likelihood of this being the case is higher for large deviations from the best-response benchmark and will generally imply a very high or low willingness to pay for equality that may not always be in accordance with the behavior in the other choices, thus generating a noisy strong redistributor group.

## 2.C Deriving the density of the likelihood on the individual level

The model assumes that utility has a deterministic ( $u$ ) component characterized by equation (2.2) and a random component ( $\varepsilon$ ) that is stochastic.

$$U^c(X_c, \theta, \sigma) = u(x_c, \theta) + \varepsilon_{X_c} \quad (2.4)$$

Let  $c \in \{A, B\}$  be the contract chosen by the principal, let  $X_c = (y_c, \pi_{1,c}, \pi_{2,c}, x_{1,c}, x_{1,c})$  be a vector of the contract's characteristics (own income, worker's production, worker's income); let  $\theta$  be a vector of parameters e.g.  $(\beta, \sigma)$  if we fit the model posited in equation (2.2);  $\varepsilon$  is an idiosyncratic error in the valuation of  $u$  that is assumed to follow a type-I extreme value distribution with a scale parameter  $\frac{1}{\sigma}$ .

The random component allows us to identify the probability that a principal chooses a given contract within his choice set (Contract 1 or 2). We assume that any principal will choose Contract 1 over 2 if  $U^1(X_c, \theta, \sigma) \geq U^2(X_c, \theta, \sigma)$ . This can be re-expressed as a probability and yields:

$$\begin{aligned} Pr(Choice_t = 1) &= Pr(u^1(X_1, \theta) - u^2(X_c, \theta) \geq \varepsilon_2 - \varepsilon_1) \\ &= \frac{\exp(\sigma u^1(X_1, \theta))}{\exp(\sigma u^1(X_1, \theta)) + \exp(\sigma u^2(X_2, \theta))} \end{aligned}$$

If  $\sigma$  is equal to zero, the probability that we choose any contract is equal to 0.5, and the deterministic part of the utility function does not affect her decision and the parameters are uninformative.

The subject's contribution to the conditional density at the choice level will therefore be

$$f_{i,t}(\theta, \sigma | X_1, X_2, Choice) = Pr(Choice_{i,t} = 1)^{\mathbb{1}(Choice_{i,t}=1)} Pr(Choice_{i,t} = 2)^{\mathbb{1}(Choice_{i,t}=2)}$$

where  $t$  denotes one of the  $T = 16$  individual decisions between two contracts. Taking the product over all the decisions the subject makes, we have the subject's overall contribution to the density.

$$f_i(\theta, \sigma | X_1, X_2, Choice) = \prod_{t=1}^T f_{i,t}(\theta, \sigma | X_1, X_2, Choice)$$

If we assume that heterogeneity is constant within a type, we can rewrite this density function as a type-specific contribution to the density. Therefore, this represents the contribution of an individual of type  $k$  to the density:

$$f_k(\theta_k, \sigma_k | X_1, X_2, Choice) = \prod_{t=1}^T f_{i,t}(\theta_k, \sigma_k | X_1, X_2, Choice)$$

## 2.D Counterfactual analyses

In this section we will use the estimates from Section 6 and simulate the contracts chosen by each principal and the corresponding profits. The counterfactual analysis is useful to get a better understanding under which situation we would expect differences in contracts across the different types of principals.

So far, we have assumed that *workers* do not have social preferences and are thus neutral to piece rate differences relative to their co-worker. This is a mechanical feature of our design since we did not inform workers that they were forming pairs. We wanted to isolate the principals’ normative preferences, abstracting from strategic concerns arising when workers compare themselves. Our structural estimation enables us to simulate what would have happened in situations where workers dislike inequality (with varying definitions of inequality). In these situations, we show that egalitarian principals’ choices become more optimal from an output-maximization perspective. The intuition is that egalitarian principals tend to treat workers more equally and are thus able to avoid sabotage situations that may arise due to undesired inequality.

The simulations are based on a simple principal-agent model in which principals maximize expected utility and their income is the profit made by the firm.<sup>30</sup> Agents hold a power-cost function (see e.g. DellaVigna et al., 2016)  $c(e) = \frac{ke^{1+s}}{1+s}$ , where we vary the curvature of the effort function,  $s$ , across high- and low-ability agents such that  $s_h < s_l$ . This characterizes the idea that high-ability agents tire less quickly as they increase their effort level.<sup>31</sup>

The extent to which principals are able to fully anticipate their workers’ social preferences is unclear. Even though the majority of principals believed that agents best respond to incentives in our experiment, this was not the case for all principals. We will study two natural benchmarks for *all* three distributive preference types: (1) sophisticated principals who correctly anticipate agents’ other-regarding concerns; (2) naive principals who falsely believe that agents are not other-regarding and so do not adapt their contracts’ choices as agents’ other-regarding concerns grow stronger. The two benchmarks show how profits change across the three distributive preference types as agents become more other-regarding, with principals’ expectations remaining constant.

### 2.D.1 Including social comparisons by agents

Social comparisons among agents matter in the field (Breza et al., 2018; Card et al., 2012) and also, but to a lesser extent, in the lab (e.g. Gagnon et al., 2020; Gross et al., 2015; Charness

---

<sup>30</sup>We depart from the design in this case because we need to introduce a budget constraint, and therefore letting the principals bear wage costs introduces a budget constraint. Otherwise, if principals only maximized output without any budget constraint, they would choose  $w_h = w_l \rightarrow \infty$  which minimizes inequality while maximizing output. In our experiment, since choices are binary, it is not necessary to introduce a constraint, and this would have overly complicated the design.

<sup>31</sup>This is another departure from the approach we adopted in the experiment, where agents are heterogeneous in terms of their marginal productivity. We diverge from this approach because the ability term is canceled out in the principal’s maximization problem and *always* yields equal piece rate contracts in equilibrium. Note that we can generate similar results by assuming that agents differ linearly in their productivity as in the experiment, but that high-ability agents have higher bargaining power due to their higher ability. However, we prefer the above approach as it does not require the modeling of the labor market.

and Kuhn, 2007). The standout finding from these studies is that agents generally accept inequality that makes them better off or that reflects differences in productivity. Only one study studies horizontal comparisons under differences in piece rates (Gagnon et al., 2020). One of their findings is that agents are averse to being treated differently. Nonetheless, we will assign several utility functions to agents and compare them across the following four different scenarios, covering a broad spectrum of social preferences: (1) caring about differences in piece rates; (2) caring about receiving higher piece rates; (3) caring about differences in potential income; and (4) caring about being better off in terms of potential income.<sup>32</sup>

We follow the general framework laid out by Breza et al. (2018). Workers do not only care about their own wage but also about the reference wage. Reference wage is hereby assumed to be her colleague’s wage. We posit that workers’ payoffs are denoted as

$$V(y_i, y_R, w_i, w_r, e_i) = y_i(e_i, w_i) + M(w_i, w_R, y_i(e_i, w_i), y_R(e_R, w_R))e_i \quad (2.5)$$

where  $y_i(e_i, w_i)$  is the ex post income of agent  $i$ , which depends on the effort level  $e_i$  she exerts and the piece rate she receives.  $M(\cdot)$  is the social preferences function, which depends on the agent’s piece rate  $w_i$ , the reference piece rate  $w_R$  and the reference ex post income  $y_R$ . We thus assume that workers may not only care about other workers’ ex post income but *also* about their colleagues’ piece rates. We will vary the precise structure of  $M(\cdot)$  across the scenarios.

**(1) Agents care about differences in piece rates** If agents have a distinct preference for equal treatment, we can model the agent as being averse to differences in piece rates:

$$M(w_i, w_R) = -\alpha_a(w_R - w_i | w_i < w_R) - \beta_a(w_i - w_R | w_R \leq w_i) \quad (2.6)$$

For a given level of  $\alpha_a$  and  $\beta_a$  they will reduce their effort level if the dispersion of piece rates becomes too high.<sup>33</sup> Indeed their optimal effort decision becomes:

$$e = \left[ \frac{w_i - \alpha_a(w_R - w_i | w_i < w_R) - \beta_a(w_i - w_R | w_R \leq w_i)}{k} \right]^{1/s_i} \quad (2.7)$$

We study two distinct cases: (1a) agents caring about differences in piece rates as such ( $\alpha_a = \beta_a \geq 0$ ) and (1b) agents caring about receiving a slightly *higher* piece rate, i.e. they are only upset when they get a lower piece rate relative to the other agent ( $\alpha_a \geq 0$  and  $\beta_a = 0$ ).

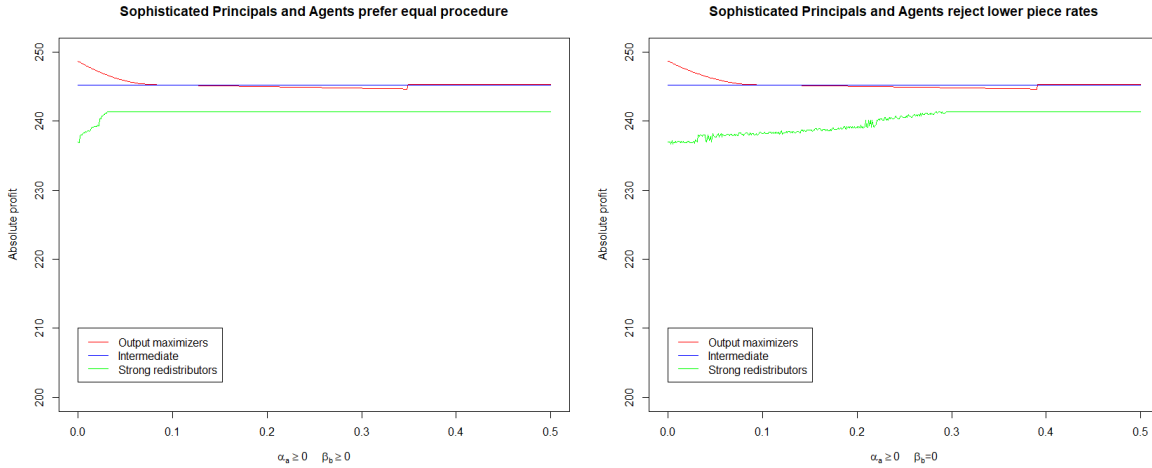
Note that principals are now continuously choosing between all possible piece rate contracts according to their expectations of agents’ responses. For each level of  $\alpha_a$  and  $\beta_a$  we have one

<sup>32</sup>We are not claiming that agents necessarily hold these exact preferences, we are merely generating hypothetical situations that give us an idea of what would have happened if agents held these preferences.

<sup>33</sup>This result is illustrated by the derivative of the agent’s utility with respect to effort:  $V'_{e_i}(\cdot) = y'_i(e_i) - (\alpha_a(w_i - w_R | w_i < w_R) + \beta_a(w_R - w_i | w_R \leq w_i))$ . At high levels of piece rate inequality or strong other-regarding motives, a marginal increase in effort will reduce utility even if  $y'(e_i) > 0$ . Note that  $y(e_i) = w_i * e_i$ .

pair  $(w_H, w_L)$  that is payoff-maximizing for principals. There is no closed-form solution for the principal’s problem. We can therefore present the results of the numerical simulations.

Figure B1: Firm’s profit if agents care about equal procedure and principals anticipate it correctly



(a) Agents reject inequality in piece rates

(b) Agents dislike having a lower piece rate

*Notes:* The graphs display simulated profit. Principals choose piece rates using the three preference types identified in the previous section. Agents hold preferences characterized by equation (2.6). The y-axis displays absolute profit. The x-axis displays variation in  $\alpha_a$  and  $\beta_a$ . Figure B1a simulates agents with  $\alpha_a = \beta_a \geq 0$ , while Figure B1b simulates agents with  $\alpha_a \geq 0$  and  $\beta_a = 0$ . Principals correctly anticipate their agents’ behavior.

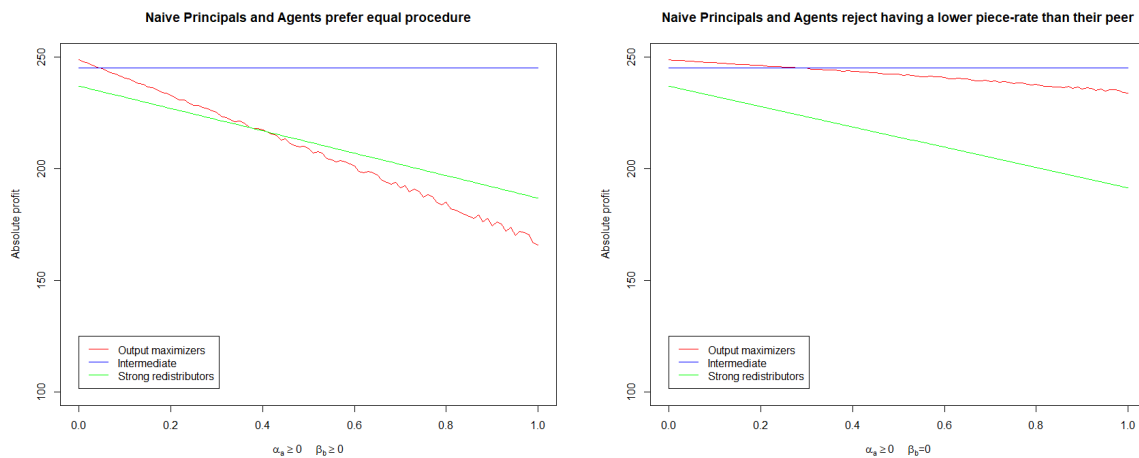
**Sophisticated Principals** Figure B1 plots the profits associated with each distributive type for different values of  $\alpha_a$  and  $\beta_b$ . Here, we make the assumption that principals correctly anticipate the agents’ social preferences. Figure B1a considers a case where  $\alpha_a = \beta_b$  and agents care about differences in piece rates symmetrically. While profits change for output-maximizers and egalitarian principals, they are constant for intermediate principals. This stems from the fact that they already implement an equal piece rate contract if agents do not hold any social preferences because they are averse to inequality once the high ability agent is paid at a higher piece rate. In addition, we see that as agents become more other-regarding, the output-maximizing principals’ profits decrease because they now face retaliation if there is a difference in piece rates. For low levels of  $\alpha_a$  and  $\beta_a$  the gap in profits between intermediates and output maximizers shrinks as  $\alpha_a$  and  $\beta_a$  increase. For high levels of  $\alpha_a$  and  $\beta_a$ , the gap eventually closes, and the two types prefer the same contract, which gives the same piece rate to both agents. Turning to the behavior of egalitarian principals, we can observe that profits rise as  $\alpha_a$  and  $\beta_a$  increase. This comes from the fact that it is now even more costly to implement redistributive contracts because they misallocate incentives, and they are disliked by agents because they do not pay equal piece rates. Egalitarian principals react to this pressure by issuing contracts that become more equal in piece rates and less distorting (yielding higher profits). For high levels of  $\alpha_a$  and  $\beta_a$ , egalitarian principals assign the same piece rates to both agents. However, this contract has a lower piece rate *level* than that preferred by the intermediate and output-maximizing types. This piece rate will indeed generate less inequality than the piece rate proposed by intermediate

principals, but is Pareto inferior for the agents as they could *both* be better off under the other equal piece rate contract.

Figure B1b examines the case in which agents dislike receiving a lower piece rate than their co-worker but do not mind receiving a higher piece rate. In this case, convergence is slower. This stems from the fact that the *high-ability* agent does not reject this contract whereas the low-ability agent does. This makes equality much more costly for the redistributive principal because he cannot reduce the high-ability agent’s piece rate without seeing a drop in his effort.

**Naive Principals** Figure B2 plots simulated profits for the three types of principals if they are naive about agents’ social preferences. A first look at the graphs reveals stark differences relative to Figure B1. Figure B2a considers the case in which agents dislike any difference in piece rates. Strikingly, intermediate principals do not incur any losses by wrongly anticipating that agents are averse to differences in piece rates because they already implement an equal piece rate contract if agents do not hold any social preferences. On the contrary, strong redistributors incur substantial losses because they implement a contract in which the high-ability agent receives a lower piece rate than the low-ability agent. Naive output maximizers also incur large losses as agents become more other-regarding. At some point, they are even less efficient than egalitarian principals because their preferred piece rate spread is too high.

Figure B2: Firm’s profit if agents care about equal procedure and principal is naive



(a) Agents reject inequality in piece rates

(b) Agents dislike having a lower piece rate

*Notes:* The graphs display simulated profit. Principals choose piece rates according to the three preference types identified in the previous section. Agents hold preferences characterized by equation (2.6). The y-axis displays absolute profit. The x-axis displays the variation in  $\alpha_a$  and  $\beta_a$ . Figure B2a simulates agents with  $\alpha_a = \beta_a \geq 0$ , while Figure B2b simulates agents with  $\alpha_a \geq 0$  and  $\beta_a = 0$ . Principals believe that agents do not hold any social preferences.

Turning to Figure B2b, where agents are only averse to differences in piece rates that make them worse-off, we can observe that egalitarians still perform worse as agents become more other-regarding. Output-maximizers are, however, nearly as efficient as intermediate principals. This is because the high-ability agent does not retaliate to receiving a higher piece rate whereas the low-ability agent does. However, given his low ability, this is not very costly. The opposite is true for the egalitarian principal who pays a higher piece rate to the low-ability agent and



the high-ability agent retaliates. This has a significant effect on profits, as characterized by the graph.

**(2) Agents have a preference for ex post equality** These agents can be modeled as being difference-averse in their expectations. Hence, they care about inequality in the income that individuals are able to attain – their potential income – under a given piece rate. In other words, they care about the inequality of outcomes that would occur if both agents best-responded to incentives  $(y_i(w_i, e_i^*), y_R(w_R, e_R^*))$ :<sup>34</sup>

$$M(w_i, w_R) = -\alpha_a(y_R(w_R, e_R^*) - y_i(w_i, e_i^*)|y_i^* < y_R^*) - \beta_a(y_i(w_i, e_i^*) - y_R(w_R, e_R^*)|y_R^* \leq y_i^*) \quad (2.8)$$

For a given level of  $\alpha_a$  and  $\beta_a$  agents will reduce their effort level if the dispersion of *potential* income becomes too high. Indeed their optimal effort decision becomes

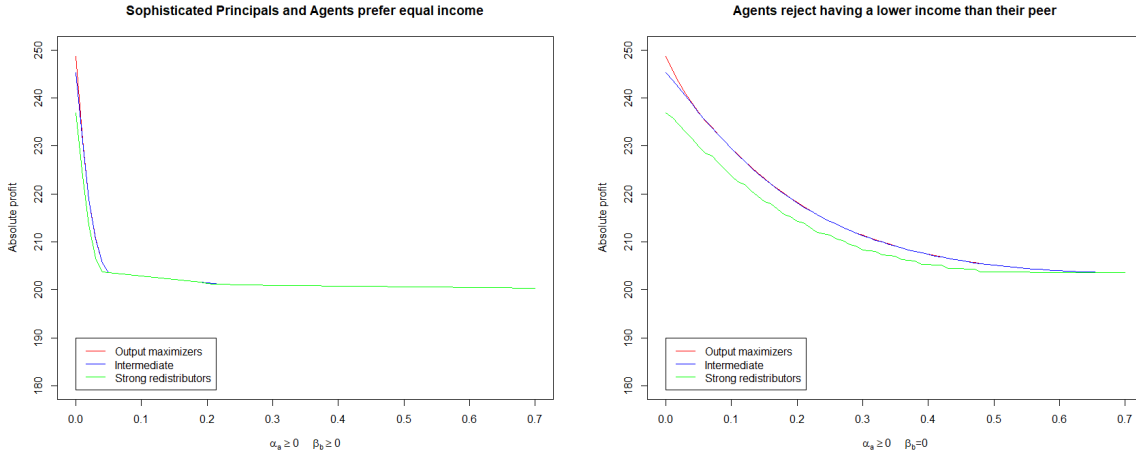
$$e_i = \left[ \frac{w_i - \alpha_a(y_R(w_R, e_R^*) - y_i(w_i, e_i^*)|y_i^* < y_R^*) - \beta_a(y_i(w_i, e_i^*) - y_R(w_R, e_R^*)|y_R^* \leq y_i^*)}{k} \right]^{1/s_i} \quad (2.9)$$

We will further examine principals' decisions when they correctly anticipate the agents' behavior, and when they are naive about agents' other-regarding preferences.

---

<sup>34</sup>This modelling choice takes an ex ante perspective and argues that agents care more about what they would have earned if both had exerted their optimum effort levels. We prefer this approach to one assuming that agents care about equilibrium levels of inequality, i.e. the distribution of income after reacting to the choice of contract *and* its distributive consequences. We make this decision because it reflects the idea that agents care about being able to earn the same ex-post income. If agents care about equilibrium levels of inequality, then we would end up with multiple equilibria, including cases in which the low-ability agent increases his effort to compensate for having a low piece rate. This would amount to rewarding the principal for her unequal treatment. We do not consider this to be a realistic situation as it does not capture the fact that agents mostly care about the principal's intentions.

Figure B3: Firm’s relative profit if agents care about potential ex-post income and principals are sophisticated

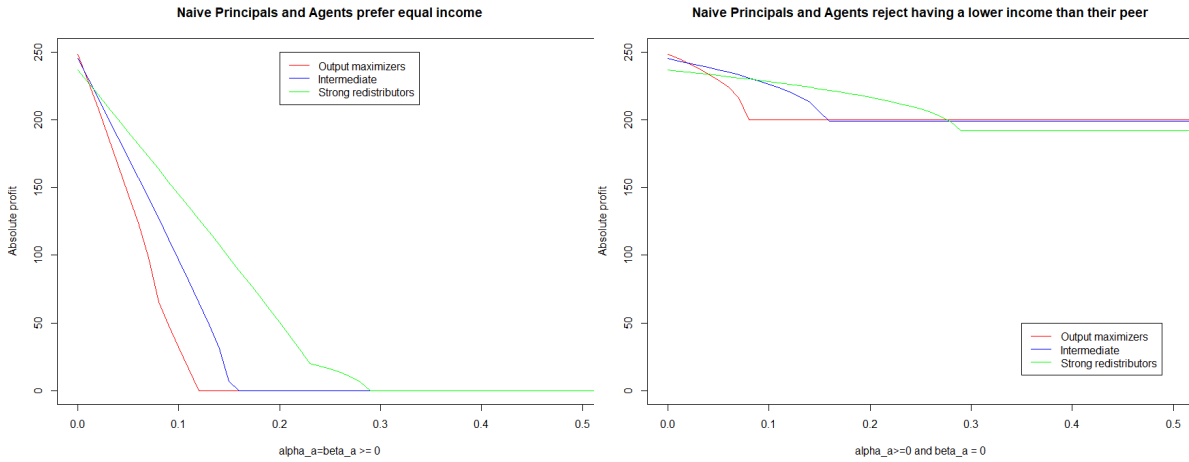


(a) Agents reject inequality in potential income (b) Agents dislike having lower potential income and principal anticipates correctly

*Notes:* The graphs display the simulated profit for each of the three preference types. Principals choose piece rates based on the preferences identified in the previous section. Agents hold preferences characterized by equation (2.8). The y-axis displays the absolute profit. The x-axis displays the variation in  $\alpha_a$  and  $\beta_a$ . Figure B3a simulates agents with  $\alpha_a = \beta_a \geq 0$ , while Figure B3b simulates agents with  $\alpha_a \geq 0$  and  $\beta_a = 0$ . Principals correctly anticipate their agents’ behavior.

**Sophisticated Principals** For each distributive type, Figure B3 plots the absolute profits for different values of  $\alpha_a$  and  $\beta_b$ , now assuming that agents hold preferences as in equation (2.8). Figure B3a considers a case in which  $\alpha_a = \beta_b$  and agents care about differences in potential income symmetrically. We can see a similar convergence in behavior to the previous case but the egalitarian principals now become indistinguishable from the output maximizers as other-regarding concerns grow stronger. This stems from the fact that agents punish deviations more severely for higher  $\alpha_a$  and  $\beta_b$ . This becomes very costly for all principals and consequently, the optimal behavior now becomes egalitarian with the low-ability worker receiving a higher piece rate in order to harmonize *ex post* the workers’ income. This is why the output-maximizing principals are behaving like egalitarian ones. Even though we still see the same convergence as in Figure B3b, it occurs more slowly if agents only care about being *worse* off than their peers. Taking a closer look at the income levels, we can see that output-maximizing principals are quick to reduce inequality by giving a lower piece rate to the high-ability worker and a higher piece rate to the low-ability worker. This is due to the fact that the *low-ability* worker’s rejection of inequality becomes a much stronger response than the gain from giving marginally higher incentives to the high-ability worker.

Figure B4: Firm’s relative profit if agents care about potential ex post income and principals are naive



(a) Agents reject inequality in potential income (b) Agents dislike having lower potential income and principal anticipates correctly

*Notes:* The graphs display simulated profit for each of the three preference types. Principals choose piece rates using the preferences identified in the previous section. Agents hold preferences characterized by equation (2.8). The y-axis displays absolute profit. The x-axis displays the variation in  $\alpha_a$  and  $\beta_a$ . Figure B4a simulates agents with  $\alpha_a = \beta_a \geq 0$ , while Figure B4b simulates agents with  $\alpha_a \geq 0$  and  $\beta_a = 0$ . Principals believe that agents do not hold any social preferences.

**Naive Principals** Figure B4 simulates profits for naive principals and inequality-averse agents. Figure B4a assumes that agents care equally about disadvantageous and advantageous inequality. As agents become more inequality-averse, strong redistributors become the most efficient type in relative terms. This is intuitive because the contract they prefer remains that which equalizes ex post incomes, even if they expect agents to be neutral with respect to their co-workers. At some point, other-regarding concerns become so strong that agents do not work at all, even under a contract chosen by a naive egalitarian principal.<sup>35</sup> Hence, agents with high  $\alpha_a$  and  $\alpha_b$  will eventually retaliate in response to even a small gap in potential income.

Figure B4b assumes that agents dislike being worse-off than their peers. For modest levels of other-regarding concerns, the naive strong redistributors do better than the other two types. However, at some point, low-ability agents no longer exert any effort at all, even under a contract preferred by the egalitarian principal. As in the previous figure, this is due to the fact that there is a small difference in potential income, even in contracts implemented by naive egalitarians. Low-ability agents, who receive a slightly lower potential income, will eventually sabotage this contract if  $\alpha_a$  becomes too large. Then, only high-ability agents will work (because they do not care about advantageous inequality) and we return to the situation in which the naive output-maximizing principal is the most efficient.

The simulations have shown that intermediate principals become indistinguishable from output-maximizing principals as we increase workers’ distaste for piece rate inequality. Egalitarian principals, however, still prefer suboptimal incentives that, ex post, yield lower inequality. How-

<sup>35</sup>The egalitarian principals that we identified in our data are not “perfect” egalitarians and still face a residual trade-off.

ever, if we assume that agents dislike inequality in ex post income, we find that all three types become indistinguishable in equilibrium. If we assume that principals are naive about workers' social preferences and falsely believe that workers will best respond to incentives, we observe that intermediate principals are more efficient if agents only care about differences in piece rates, and egalitarian principals become more efficient if workers are egalitarian in expected income. These results demonstrate that the manner in which other-regarding preferences held by principals conflict with optimality is crucially dependent on the setting in which the principal operates.

## 2.E Instructions

Thank you for participating in this experiment. Please read the following instructions carefully. Your answers will remain anonymous throughout the experiment. Please refrain from talking to your neighbors, and turn off your cellphones. If you choose your answers carefully, you may earn a substantial payoff.

The currency used in this experiment is the ECU. At the end of the experiment, you will be paid in euros using the following conversion rate: 1 euro = 10 ECU.

### 2.E.1 Principals

This experiment takes place in a firm. There are two possible roles: being the principal of the firm or being one of the two employees. Your role has been drawn randomly; you are the principal of the firm.

**The employees** As the principal of the firm, you have to choose the wage paid to both employees. These two people are also participating in this experiment at the same time as you. Although you are in the same room, you will never know who they are, and they will never know who you are. Your identity and their identity will remain anonymous throughout the experiment.

*[Stakeholder treatment]* You will receive compensation of 60 ECU for your participation. In addition, you will obtain a variable wage that will depend on the production level of both employees. You will obtain the revenues generated by the sales of the units produced by the employees. You will also have the opportunity to earn more money if you correctly guess your employees' behavior.

*[Spectator treatment]* You will receive a fixed wage of 200 ECU for your participation. You will also have the opportunity to earn more money if you correctly guess the behavior of your employees.

Both employees' wages are paid in two parts. They first receive a fixed participation fee of 90 ECU. The second part is variable and depends on the number of units they produced. Your task is to choose how this variable part is calculated.

**Employees’ effort level and production** Both employees will have to choose their effort levels for the performance of their jobs. Each effort level is associated with a production level. The higher the effort level chosen by the employees, the more they will produce.

*[Stakeholder treatment]* The more they produce, the more money you will earn. Each unit produced by the employees will earn you 0.5 ECU.

*[Spectator treatment]* Your own wage is completely independent of their performance. You will receive a fixed wage of 200 ECU.

Example of an effort-production table:

Figure D1: Effort-production table *[Stakeholder treatment]*

Niveau d'effort	0	0.5	1	1.5	2	2.5	3	3.5	4	4.5	5
Production	0	25	50	75	100	125	150	175	200	225	250
Votre revenu	0	13	25	38	50	63	75	88	100	113	125

Figure D2: Effort-production table *[Spectator treatment]*

Niveau d'effort	0	0.5	1	1.5	2	2.5	3	3.5	4	4.5	5
Production	0	25	50	75	100	125	150	175	200	225	250
Votre revenu	200	200	200	200	200	200	200	200	200	200	200

**Choice of the wage compensation scheme** You will have to define the details of both employees’ employment contracts. You will have to decide on the piece-rate wage that each employee will receive. We will show you several examples at the end of the instructions.

**Your employees’ ability** You will obtain information about the ability of both employees. One of them will be more productive than the other. In other words, for the same effort level, one of them will produce a larger quantity than the other.

We will show you a table for each employee describing how their efforts translate into units produced for both employees. You will be able to refer to these tables when you make your wage compensation choices.

The employees’ ability will be determined by an aptitude test that they will take at the beginning of the experiment. The higher their grade in the test, the higher their productivity.

This test is a multiple-choice questionnaire consisting of 3 French questions, 3 logic questions and 3 general knowledge questions. They will have 5 minutes to complete the test.

At the end of the instructions, you will also have the opportunity to answer the questions of this test in order to better understand how your employees' productivity has been determined.

**Individual choices** The employees choose their effort level in complete independence; they will never communicate with each other, nor with you, during the experiment.

They will know the piece rate you chose for them but will be unaware that you have hired another employee. They will not know which piece rate you chose for the other employee. They are not informed that there is another employee.

**Effort cost** Employees choose an effort level after they have each discovered their piece-rate wage. The higher the effort level they choose, the more it will cost them. Each effort level is associated with a cost in ECU. Therefore, if they choose a high effort level, they will have a higher effort cost to deduct from the wage you will pay them. The cost of the effort is identical for both employees.

Example of an effort-production table

Figure D3: Effort-production-cost table

Niveau d'effort	0	0.5	1	1.5	2	2.5	3	3.5	4	4.5	5
Production	0	25	50	75	100	125	150	175	200	225	250
Coût de l'effort	0.0	1.3	5.0	11.3	20.0	31.3	45.0	61.3	80.0	101.3	125.0

Hence, if this employee chooses an effort level of 1.5, it will cost 11.3 ECU. If she chooses an effort level of 5, it will cost 125 ECU.

**Impact of your choices** You will choose between several employment contracts for the two employees chosen randomly from among the participants in this experiment today. Your choices have real consequences for both participants. One of your wage choices for both employees will be drawn randomly and will be implemented. You will be the sole decision-maker for both employees.

*[Stakeholder treatment]* Your own income will correspond to the sales of the unit produced by both employees. Each unit produced will earn you 0.5 ECU. You may additionally earn money for guessing the effort level that your employees will choose in response to various piece-rate wages.

*[Spectator treatment]* On top of your fixed wage of 200 ECU, you may earn money for guessing the effort levels that your employees will choose when confronted with various piece-rate wages.

## 2.E.2 Workers

This experiment takes place in a firm. There are two possible roles: being the principal of the firm or being an employee. Your role has been drawn randomly: you are an employee. You will receive a fixed wage of 90 ECU for participating. You can also obtain an additional wage that will depend on your decisions.

**Firm** You work in a firm. A principal who has been drawn at random from the people present in this room will offer you a work contract describing your wage for each unit you will produce (piece-rate wage). You must choose an effort level that will be associated with a quantity of units produced. The higher the effort level you choose, the more you will produce. The more you produce, the higher your income will be.

The table below illustrates hypothetically how effort may translate into production for several different effort levels.

Here is an example of an effort-production table:

Figure D4: Effort-production table

<b>Niveau d'effort</b>	0	0.5	1	1.5	2	2.5	3	3.5	4	4.5	5
<b>Production</b>	0.0	25.0	50.0	75.0	100.0	125.0	150.0	175.0	200.0	225.0	250.0

**Ability** You will have the opportunity to influence how your choice of effort level translates into the quantity produced. You will take an aptitude test that will determine your ability level. This test is a multiple-choice questionnaire consisting of 3 French questions, 3 logic questions and 3 general knowledge questions. Participants will have 5 minutes to complete the test. The higher your performance at this test, the higher your production level will be for a given effort level. You will have an opportunity to familiarize yourself with this type of test by answering 9 other similar questions for 10 minutes.

**Effort cost** If you choose a high effort level, you will produce more but this will be more costly for you as well. Each effort level is associated with a cost in ECU. Therefore, if you choose a high effort level, you will have a higher effort cost to deduct from your income.

Example of an effort-production-cost table

Figure D5: Effort-production-cost table

Niveau d'effort	0	0.5	1	1.5	2	2.5	3	3.5	4	4.5	5
Production	0.0	25.0	50.0	75.0	100.0	125.0	150.0	175.0	200.0	225.0	250.0
Coût de l'effort	0.0	1.3	5.0	11.3	20.0	31.3	45.0	61.3	80.0	101.3	125.0

Therefore, if you choose an effort level of 1.5, it will cost you 11.3 and you will produce 75 units. If you choose an effort level of 4, it will cost you 80 and you will produce 200 units.

**Your income** You will be paid a fixed amount for each unit produced. You will be informed of this piece rate before choosing your effort level. In the example below, we show you your variable income (net of effort cost) for a piece rate of 0.4 ECU. Your net variable income corresponds to the production multiplied by the piece rate minus the effort cost. In summary, your net variable income = production x piece-rate - effort cost.

Figure D6: effort-production-cost-income table

Niveau d'effort	0	0.5	1	1.5	2	2.5	3	3.5	4	4.5	5
Production	0.0	25.0	50.0	75.0	100.0	125.0	150.0	175.0	200.0	225.0	250.0
Coût d'effort	0.0	1.3	5.0	11.3	20.0	31.3	45.0	61.3	80.0	101.3	125.0
Votre rémunération variable (nette du coût de l'effort) avec un salaire à la pièce de 0.4	0.0	18.7	15.0	18.8	20.0	18.7	15.0	8.7	0.0	-11.3	-25.0

**Impact of your choices** You will be asked to choose effort levels for several employment contracts. At the same time, the principal of the firm will choose one of these contracts. You will be paid according to the choice made by the principal. The principal will choose a contract without knowing which effort level you chose. You will be unable to communicate with the principal of the firm during the experiment, and will not know his or her identity. Therefore, the principal will be unable to influence your choices. You are completely free to choose your effort level and the principal will be unaware of your choice when making his or her employment contract decision.

## 2.F Comprehension test

The principals' comprehension tests were composed of 3 sets of questions of increasing difficulty (tests 1, 2 and 3). Each test consisted of 2 to 6 questions. For each of the three tests, subjects could take three trial tests with hints and feedback on each question to improve their understanding. After the three tests, they had to answer simple True-False questions in order to assess their overall understanding of the rules of the experiment.



Workers also had to take a comprehension test based on the same format, but the questions were adapted to their own choice environment.

Workers and principals were given different tests since their choices were very different. The workers’ test ensured that workers were capable of reading the effort-cost-income table (as in Figure A2). We asked them to determine how much income they would obtain under various piece-rate wages and effort choices. The principals’ comprehension tests ensured that they were capable of reading the double table describing the characteristics of Workers A and B (as in Figure A3). We asked them to determine the differences between worker A and B (which is the more productive?) and to determine their output, how much income each worker would receive, and their own income in various situations. The Spectators’ test was slightly easier since their income is 20 euros in all cases.

### 2.F.1 Questions principals

Before moving on to your final choices, we will first ask you a few questions in order to assess your understanding. This test will have no impact on the rest of the experiment. We just want to make sure that you fully understand how the experiment works. You can raise your hand at any time, and someone will come to answer your questions.

#### Test 1

Let’s take the following example. Here is the information about your employees A (first table) and B (second table). The left-hand columns show the production, cost of effort and your income for low effort levels, and the right-hand columns give this information for higher effort levels.

*[Stakeholder treatment]* Therefore, for employee A, we can see that if he or she chooses an effort level of 2, he or she will produce 100 units. It will cost him or her 20 ECU. For employee B, if he or she chooses an effort level equal to 2, he or she will produce 50 units. It will cost him or her 20 ECU.

*[Spectator treatment]* Therefore, for employee A, we can see that if he or she chooses an effort level of 2, he or she will produce 100 units. It will cost him or her 20 ECU and you will earn income of 50 ECU. Indeed, each unit produced is sold at 0.5 ECU. For employee B, if he or she chooses an effort level equal to 2, he or she will produce 50 units. It will cost him or her 20 ECU and you will earn income of 25 ECU. Your total income from the sales of the units produced will thus be equal to  $50 + 25 = 75$  ECU

Figure E1: Effort-production-cost table *[Stakeholder treatment]*

Employé A											Employé B												
Niveau d'effort	0	0.5	1	1.5	2	2.5	3	3.5	4	4.5	5	Niveau d'effort	0	0.5	1	1.5	2	2.5	3	3.5	4	4.5	5
Production	0.0	25.0	50.0	75.0	100.0	125.0	150.0	175.0	200.0	225.0	250.0	Production	0.0	12.5	25.0	37.5	50.0	62.5	75.0	87.5	100.0	112.5	125.0
Coût de l'effort	0.0	1.3	5.0	11.3	20.0	31.3	45.0	61.3	80.0	101.3	125.0	Coût de l'effort	0.0	1.3	5.0	20.0	20.0	31.3	45.0	61.3	80.0	101.3	125.0
Votre revenu	0.0	12.5	25.0	37.5	50.0	62.5	75.0	87.5	100.0	112.5	125.0	Votre revenu	0.0	6.3	12.5	18.8	25.0	31.3	37.5	43.8	50.0	56.3	62.5

Figure E2: Effort-production-cost table [*Spectator treatment*]

Employé A											Employé B												
Niveau d'effort	0	0.5	1	1.5	2	2.5	3	3.5	4	4.5	5	Niveau d'effort	0	0.5	1	1.5	2	2.5	3	3.5	4	4.5	5
Production	0.0	25.0	50.0	75.0	100.0	125.0	150.0	175.0	200.0	225.0	250.0	Production	0.0	12.5	25.0	37.5	50.0	62.5	75.0	87.5	100.0	112.5	125.0
Coût de l'effort	0.0	1.3	5.0	11.3	20.0	31.3	45.0	61.3	80.0	101.3	125.0	Coût de l'effort	0.0	1.3	5.0	20.0	20.0	31.3	45.0	61.3	80.0	101.3	125.0

**Question 1:** Which employee is of higher ability (who is the more productive employee)?

Imagine that employee A chose an effort level of 0.5 and employee B an effort level of 3.

**Question 2:** What is the total production?

**Question 3 [*Stakeholder treatment only*]:** How much income do you earn from employee B (what is your income due to the production of employee B)?

**Question 4 [*Stakeholder treatment only*]:** What is your total income? (add up the income that you earn from both employee A and employee B)

**Test 2**

You clearly understand how production works in your firm. Now we are going to show you wage simulations to help you make your choices. These examples will have no impact on the rest of the experiment. Let's consider a first choice between two employment contracts. Contract 1 pays employee A 0.6 ECU per unit produced and employee B 0.4 ECU per unit produced. Contract 2 pays employee A 0.4 ECU and employee B 0.6 ECU. We have added two lines to the table, which show the variable wage (net of effort cost) of your employees for both contracts. We have deleted the lines showing your employees' production and effort cost in order to simplify the tables. Remember that the variable wage (net of effort cost) is equal to the production multiplied by the piece-rate wage minus the effort cost.

Figure E3: Effort-production-cost table [*Stakeholder treatment*]

Employé A											Employé B												
Niveau d'effort	0	0.5	1	1.5	2	2.5	3	3.5	4	4.5	5	Niveau d'effort	0	0.5	1	1.5	2	2.5	3	3.5	4	4.5	5
rémunération variable (nette du coût de l'effort) de l'employé A avec un salaire à la pièce de 0.6	0.0	13.8	25.0	33.8	40.0	43.8	45.0	43.8	40.0	33.8	25.0	rémunération variable (nette du coût de l'effort) de l'employé B avec un salaire à la pièce de 0.4	0.0	3.8	5.0	3.8	0.0	-3.3	-15.0	-26.3	-40.0	-56.3	-75.0
rémunération variable (nette du coût de l'effort) de l'employé A avec un salaire à la pièce de 0.4	0.0	18.7	15.0	18.8	20.0	18.7	15.0	8.7	0.0	-11.3	-25.0	rémunération variable (nette du coût de l'effort) de l'employé B avec un salaire à la pièce de 0.6	0.0	6.3	10.0	11.3	10.0	6.3	0.0	-8.8	-20.0	-33.8	-50.0
Votre revenu	0.0	12.5	25.0	37.5	50.0	62.5	75.0	87.5	100.0	112.5	125.0	Votre revenu	0.0	6.3	12.5	18.8	25.0	31.3	37.5	43.8	50.0	56.3	62.5

Figure E4: Effort-production-cost table [*Spectator treatment*]

Employé A												Employé B											
Niveau d'effort	0	0.5	1	1.5	2	2.5	3	3.5	4	4.5	5	Niveau d'effort	0	0.5	1	1.5	2	2.5	3	3.5	4	4.5	5
rémunération variable (nette du coût de l'effort) de l'employé A avec un salaire à la pièce de 0.6	0.0	13.8	25.0	33.8	40.0	43.8	45.0	43.8	40.0	33.8	25.0	rémunération variable (nette du coût de l'effort) de l'employé B avec un salaire à la pièce de 0.4	0.0	3.8	5.0	3.8	0.0	-6.3	-15.0	-26.3	-40.0	-56.3	-75.0
rémunération variable (nette du coût de l'effort) de l'employé A avec un salaire à la pièce de 0.4	0.0	18.7	15.0	18.8	20.0	18.7	15.0	9.7	0.0	-11.3	-25.0	rémunération variable (nette du coût de l'effort) de l'employé B avec un salaire à la pièce de 0.6	0.0	6.3	10.0	11.3	10.0	6.3	0.0	-8.8	-20.0	-33.8	-50.0

Imagine that you choose Contract 1, hence a rate of 0.4 ECU for employee A and 0.6 ECU for employee B.

Imagine that employee A chose an effort level of 2.5 and employee B an effort level of 1.

**Question 1:** What is the variable wage (net of the effort cost) of employee A?

**Question 2:** What is the variable wage (net of the effort cost) of employee B?

**Question 3 [*Stakeholder treatment*]:** What is your own total income?

**Test 3**

[Same tables as in Test 2] Imagine that you choose Contract 1, hence a rate of 0.4 ECU for employee A and 0.6 ECU for employee B.

**Question 1:** For this piece-rate wage of 0.6 ECU, which effort level would employee A choose if he or she wanted to make as much money as possible?

**Question 2:** For this piece-rate wage of 0.4 ECU, which effort level would employee B choose if he or she wanted to make as much money as possible?

**Question 3 [*Stakeholder treatment*]:** What would be your total income if both employee A and employee B chose the effort levels that maximize their revenues?

Imagine that you choose Contract 2, hence a rate of 0.6 ECU for employee A and 0.4 ECU for employee B.

**Question 4:** For this piece-rate wage of 0.4 ECU, which effort level would employee A choose if he or she wanted to make as much money as possible?

**Question 5:** For this piece-rate wage of 0.6 ECU, which effort level would employee B choose if he or she wanted to make as much money as possible?

**Question 6** [*Stakeholder treatment*]: What would be your total income if both employee A and employee B chose the effort levels that maximize their revenues?

**True-False**

To make sure that you understand the general rules of the experiment, here are several assertions. You have to determine which ones are correct and which ones are wrong.

1. You are matched with 3 employees.
2. Employees choose their effort level according to the piece-rate wages you offer them. You cannot force your employees to choose a particular effort level.
3. Your employees obtain compensation of 90 ECU for their participation.
4. Your employees will not know the piece-rate that you offered the other employee.
5. Both employees are identical.
6. [*Spectator treatment*]: You will earn a fixed wage of 200 ECU. You can earn more money by correctly guessing your employees' reactions.
7. A contract giving the highest piece-rate to the higher ability employee leads to a higher production level but implies larger variable wages differences relative to productivity differences.
8. A contract giving the same piece-rate to both employees causes variable wages to become proportional to the quantity that the employees respectively produce.
9. A contract giving a higher piece-rate to the low-ability employee leads to a lower production level but reduces the differences in the variable wages of both employees.

**2.F.2 Workers' questions**

**Test 1**

Imagine that you can transform effort into production according to the table below. The left-hand columns indicate production and the cost of effort for low effort levels, and the right-hand columns give this information for higher effort levels.

Figure E5: Effort-production-cost table

Niveau d'effort	0	0.5	1	1.5	2	2.5	3	3.5	4	4.5	5
Production	0.0	25.0	50.0	75.0	100.0	125.0	150.0	175.0	200.0	225.0	250.0
Coût de l'effort	0.0	1.3	5.0	11.3	20.0	31.3	45.0	61.3	80.0	101.3	125.0

**Question 1:** How much would you produce if you chose effort level 2?

**Question 2:** What is the cost associated with effort level 2?

**Question 3:** How much would you produce if you chose effort level 4?

**Question 4:** What is the cost associated with effort level 4?

**Test 2**

Now imagine that we pay you 0.4 ECU per unit produced. The table below has an additional line compared to the previous one. This line describes your variable wage (net of effort cost) for each production level. Your variable wage (net of effort cost) corresponds to the production multiplied by the piece-rate minus the effort cost.

Figure E6: Effort-production-cost table

Niveau d'effort	0	0.5	1	1.5	2	2.5	3	3.5	4	4.5	5
Production	0.0	25.0	50.0	75.0	100.0	125.0	150.0	175.0	200.0	225.0	250.0
Coût de l'effort	0.0	1.3	5.0	11.3	20.0	31.3	45.0	61.3	80.0	101.3	125.0
Votre rémunération variable (nette du coût de l'effort) avec un salaire à la pièce de 0.4	0.0	18.7	15.0	18.8	20.0	18.7	15.0	8.7	0.0	-11.3	-25.0

**Question 1:** How much would you produce if you chose effort level 3?

**Question 2:** What is the cost associated with effort level 3?

**Question 3:** What effort level allows you to obtain the highest variable wage (net of effort cost)?

**Question 4:** What effort level allows you to obtain the lowest variable wage (net of effort cost)?

**Test 3**

Now imagine that we pay you 0.6 ECU per unit produced. The last line of the table below describes your variable wage (net of effort cost) for each production level with this piece-rate wage.

Figure E7: Effort-production-cost table

Niveau d'effort	0	0.5	1	1.5	2	2.5	3	3.5	4	4.5	5
Production	0.0	25.0	50.0	75.0	100.0	125.0	150.0	175.0	200.0	225.0	250.0
Coût de l'effort	0.0	1.3	5.0	11.3	20.0	31.3	45.0	61.3	80.0	101.3	125.0
Votre rémunération variable (nette du coût de l'effort) avec un salaire à la pièce de 0.6	0.0	13.8	25.0	33.8	40.0	43.8	45.0	43.8	40.0	33.8	25.0

**Question 1:** What effort level allows you to obtain the highest variable wage (net of effort cost)?

**Question 2:** What effort level allows you to obtain the lowest variable wage (net of effort cost)?

### True-False

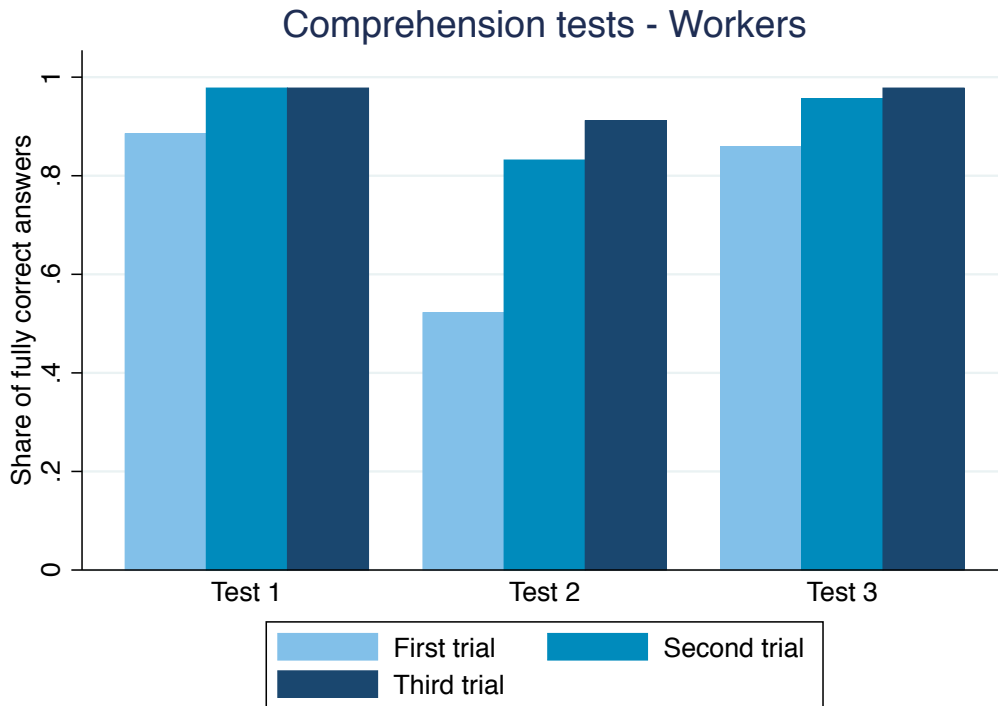
To make sure that you understand the general rules of the experiment, here are several assertions. You have to determine which ones are correct and which ones are wrong.

1. Each effort level costs the same in ECU.
2. You must choose the preferred effort level of the firm's principal.
3. You must choose effort levels for several employment contracts, but in the end, only one employment contract will be implemented so that you can be paid.
4. You receive a fixed wage of 90 ECU on top of your variable wage.
5. Your fixed wage of 90 ECU will be paid to you once only.

### 2.F.3 Comprehension test performance

Overall, subjects managed to complete the comprehension tests without any major difficulty and obtained fairly high scores. For each test, the majority of the subjects' answers were completely correct at the first try. Subsequent attempts with feedback improved scores substantially. For the last trials, the share of completely correct answers was always above 83% for all three tests. There were minor variations across Spectators and Stakeholders: principals in the Spectator treatment tended to perform slightly better. This can be easily explained by the fact that the comprehension test for Stakeholders had a few more questions and was harder because we also asked them to compute their own income under various scenarios, which was not necessary for Spectators.

Figure E8: Workers' comprehension tests



Notes: each bar displays the share of principals achieving a perfect score for each Test and trial. There are three trials per test. The first test has 2 (4) questions for Spectators (Stakeholders), the second test has 2 (3) questions for Spectators (Stakeholders) and the third test has 4 (6) questions for Spectators (Stakeholders)

Table E1: True-False average score

Variable	Obs	Mean	Std. Dev.	Min	Max
Average score Stakeholders	58	.877	.111	.5	1
Average score Spectators	55	.907	.0998	.556	1
Average score Workers	226	.857	.186	.2	1

Notes: The average score is calculated as follows. We create binary variables for each question of the True-False test that are equal to 1 if the subject answered correctly. The average score is the mean of these binary variables.

## 2.G Aptitude test

*Translated from French to English by the Authors.*

### 2.G.1 French Questions

**Question 1:** A hyperbole is a figure of speech in which the expression of an idea or reality is exaggerated in order to highlight it (example: this man is as handsome as an angel). Among the five sentences below, only one does not include hyperbole. Which one?

1. I've been waiting for you for an eternity!
2. Your story is as old as the hills: surely you don't expect anyone to believe you?

3. He came in soaked to the bones because of the storm that was raging outside.
4. **I finished this book in three hours, I devoured it.**

**Question 2:** Which of the following assertions is the odd one out?<sup>36</sup>

1. All his work is just a drop in the ocean of the work that remains to be done.
2. His explanation was as clear as a mountain stream.<sup>37</sup>
3. **There is a chasm between the world champion and his rivals.**
4. The sea is your mirror, you contemplate your soul in its infinitely rolling waves.

**Question 3** Which of the following words is a synonym of eminent?<sup>38</sup>

1. **Remarkable**
2. Immediate
3. Indiscreet
4. Boaster

### 2.G.2 Logic questions

**Question 4:** David has capital of 10,000 euros that he decides to invest in a savings account. After withdrawing his investment with interest two years later, he has total capital of 12,100 euros. What is the annual interest rate on the savings account?

1. 7%
2. **10%**
3. 11%
4. 13%

**Question 5:** The group formed by the words "triangle", "glove", "clock", "bicycle", corresponds to the group formed by the following numbers:

1. 1,2,3,4
2. **10,4,7,2**

---

<sup>36</sup>Subjects had to realize that all sentences except one uses a water-related semantic field. Sentences are translated word for word to make this clearer but obviously, these French expressions using water elements do not always have an exact English counterpart.

<sup>37</sup>Crystal-clear would be the correct translation but then this sentence would be an intruder too

<sup>38</sup>In French immediate can be translated by "imminent" and thus many people are confused about the difference between "éminent" and "imminent"



3. 4,8,10,12

4. **3,5,12,2**

**Question 6:** Complete the following series 5V - 4Q - 3L - 2G -?

1. 1A

2. 1B

3. 1C

4. 1D

### 2.G.3 General knowledge

**Question 7:** Simone Veil<sup>39</sup>

1. Was an attorney
2. Had been convicted for anti-Semitic statements
3. **Was the first woman President of the European Parliament**
4. Entered the Panthéon in September 2017

**Question 8** The Schengen Agreement is treaty about:

1. The European flag
2. The introduction of the Euro
3. The project of European Constitution
4. **The free movement of people**

**Question 9:** NASDAQ is a stock market located:

1. **In the United States**
2. In Asia
3. In the United Kingdom
4. In Germany

---

<sup>39</sup>Simone Veil was a judge but not a lawyer. She entered the French Panthéon in 2018

## Chapter 3

# Motivating Beliefs in a Just World

### Abstract

This chapter studies whether individuals distort their beliefs about the relative importance of effort and luck to motivate themselves to exert effort. To that end, I develop a novel experimental design where past experience of success or failure serves as a noisy signal about the true importance of effort in achieving success. To test whether individuals distort their beliefs to motivate future effort, I vary the moment in time when subjects are informed about an effortful task to be performed later in the experiment. Subjects who receive the information before belief elicitation face an incentive to distort their beliefs to motivate effort in the later task. The results show that such individuals are more likely to believe that their effort is important for success. Motivating belief distortion is particularly pronounced for subjects who receive disincentivizing news about the true state of the world, i.e. that success depends on luck rather than on effort. I additionally test whether motivating belief distortion affects subjects' willingness to distribute money between two *other* individuals as a third-party spectator. I find no evidence that distributive behavior differs across the two treatment groups. These results suggest that individuals' luck-effort beliefs not only depend on past or current events that inform about the true state of the world but are also endogenous to the incentive structure individuals expect to face.

---

I thank Nicolas Jacquemet, Claudia Senik, Kai Barron, Beatrice Boulu-Reshef, Alessandra Casella, Sophie Cetre, Liza Charroin, Urs Fischbacher, Marc Fleurbaey, Florian Grosset, Jeanne Hagenbach, Nina Hestermann, Shachar Kariv, Fabrice Le Lec, Suanna Oh, Leonardo Pejsachowicz, Stefano Piasenti, Chris Roth, Morten Støstad, Charlotte Saucet, Robert Stüber, Michael Thaler, Bertil Tungodden, Nikhil Vellodi, Florian Zimmermann, as well as seminar and conference participants at Aarhus University, Berlin School of Economics, NHH Bergen, Paris School of Economics and conference participants at the RGS Doctoral Conference, ADRES Doctoral Conference, MBEEES and the HEIRS Workshop “Illusion of Merit”, for very helpful comments and discussions. Funding by CEPREMAP (Program 3: Public economics and redistribution) and EUR grant ANR-17-EURE-0001 is gratefully acknowledged. This experiment was pre-registered (AEARCTR-0006421) and it was approved by PSE's IRB (2020-15).

## 1 Introduction

Research in recent decades has made great strides in identifying factors that shape attitudes towards redistribution. Such factors include beliefs about the source of inequality and the extent to which income is determined by luck rather than effort (luck-effort beliefs) (Alesina and Giuliano, 2011; Almås et al., 2020). While the literature has recognized the importance of these beliefs, empirical evidence about how they are formed remains limited.

Beliefs about the relative importance of luck and effort may be motivated. If beliefs are motivated, individuals distort their beliefs about the true relationship between effort and success in order to fulfill a certain goal. For example, if individuals have a strong desire to maintain a positive self-image by believing that they are more talented and productive than their peers, they want to attribute failure to luck and success to effort (Deffains et al., 2016). Similarly, if individuals have a strong desire to overcome a lack of motivation, they have an incentive to maintain the view that effort is important for success (Bénabou and Tirole, 2006).

This chapter studies whether individuals distort luck-effort beliefs to motivate effort. I call this form of motivated beliefs *motivating* beliefs. Bénabou and Tirole (2002) propose that individuals distort beliefs to counter an under-provision of effort due to self-control problems. These same authors created a model that shows how motivating beliefs affect preferences for redistribution (Bénabou and Tirole, 2006). Specifically, they showed that, if an economic agent with self-control problems expects low levels of redistribution, exerting low levels of effort can become very costly. This creates a demand to motivate future effort by distorting beliefs. To validate this model, it is important to provide empirical evidence of motivating belief distortion, because such evidence would show that luck-effort beliefs could be shaped by expectations about future levels of redistribution. This evidence would, thus, advance our understanding of the dynamic interaction between inequality, redistributive preferences, and beliefs about the importance of luck and effort. Such evidence also would imply that the causal relationship between luck-effort beliefs and redistribution runs both ways: Beliefs affect the demand for redistribution and expected levels of redistribution affect beliefs by shaping incentives.<sup>1</sup>

To test the prediction that future incentives distort beliefs about the importance of effort in achieving success I use an online experiment. In the experiment, subjects begin by performing a cumbersome real effort task. This task is completed in an uncertain environment, where the payment rule depends on the state of the world (Environment) that was drawn at the beginning of the experiment. If the subject is in the Performance-Environment condition, the likelihood of winning a prize for completing the task is an increasing function of their performance on the task. Specifically, the subject participates in a noisy tournament against a randomly drawn competitor, where the chance of winning the prize is equal to 80% if the subject transcribes more

---

<sup>1</sup>Beliefs affect preferences for redistribution in several ways: First, there are selfish reasons to ask for less redistribution if one distorts beliefs to motivate future effort. An individual who believes that she will be a net contributor if her effort is reflected in the pre-tax income distribution would be less likely to support redistribution after engaging in motivating belief distortion. Second, motivating belief distortion may affect preferences for redistribution for other-regarding reasons. Meritocratic individuals who distort their beliefs in a motivating way are less likely to believe that an initially unequal distribution is due to luck and if they accept inequalities that reflect differences in effort, they should, hence, opt for less redistribution compared to a situation where they did not distort their beliefs to motivate future effort.

images than her competitor, while the chance is equal to 20% if she transcribes fewer images. If the subject is in the Chance-Environment condition, her performance on the task has no effect on her likelihood of winning the prize; the subject wins the prize with 50% probability no matter how many words she transcribes.

Upon completing the task, subjects receive a noisy signal informing them about which condition—Performance-Environment or the Chance-Environment—they were in. The aim of this signal is to induce variation in baseline beliefs regarding the state of the world. I induce these beliefs by giving subjects feedback about the outcome of the first task. The feedback comprises two pieces of information: (1) whether a subject won the prize and (2) whether she transcribed more or fewer images than her competitor. Using this information, subjects can form posterior beliefs about the likelihood of being in the Chance- or Performance-Environment. For example, a subject who learns that she transcribed more images than the competitor but did not win the prize is likely to perceive herself as having a high probability of being in the Chance-Environment, that is, the condition in which success is unrelated to effort. By contrast, if the same person learns that she won the prize, she should perceive herself as having a high probability of being in the Performance-Environment where effort does influence likelihood of success.<sup>2</sup> After subjects receive the signal, I elicit their probabilistic beliefs about the environment (Chance or Performance).

To test whether subjects distort baseline beliefs to motivate themselves to exert effort, I introduce a second task that subjects can complete at the end of the experiment and that serves as an incentive (and, hence, motive) to distort beliefs for motivating purposes. As in the first task, the payment rule depends on the environment that was drawn at the beginning of the experiment: If the subject is in the Performance-Environment, effort determines whether one receives a reward for performance on the second task, while for subjects in the Chance-Environment, effort has no effect on the likelihood of receiving a reward. Subjects who know about the second task may overestimate their likelihood of being in the Performance-Environment in order to motivate themselves to work hard on the second task.

To identify motivating belief distortion, I vary the point in time at which I inform subjects about the second task: Subjects who are assigned to the Motive treatment group are informed about the second task *before* belief elicitation and, hence, have an incentive to distort beliefs to motivate effort. Subjects in the No-Motive treatment group receive this information *after* belief elicitation. The latter subjects have no incentive to distort beliefs to motivate effort because they do not know that they will be completing a second task in the experiment. This variation allows me to test the main hypothesis of the experiment: Motive-group subjects, who know that they will be completing a second task, are, on average, more confident of being in the Performance-Environment than are No-Motive-group subjects who do not know that they have to exert effort in the future.

---

<sup>2</sup>The signal mimics real life experiences that people may use to infer the importance of effort for success in life: One's colleague may get a promotion even though one considers oneself more talented and productive than the person who got the promotion; other people may get a position to which they applied, knowing that they only got the position because of their personal ties to the company's CEO; still other individuals may win an award for their work knowing that they worked harder and performed better than the other people who were short-listed for the award.

The design, shown in Figure 3.1, allows me to test for motivating belief distortion non-parametrically by comparing posterior beliefs across the two treatment groups. Nonetheless, I can go further and ask what type of signal leads to motivating belief distortion. First, I can ask whether subjects are more or less likely to engage in motivating belief distortion when receiving a signal that suggests that they are in the Performance-Environment rather than the Chance-Environment. Second, the design allows me to test whether events that are non-informative about the true state of the world affect beliefs. Specifically, I ask whether individuals are more likely to believe in the importance of effort if they won—rather than lost—a reward, holding the informational content of the event constant. This allows me to infer what type of events induce motivating belief distortion.

I further ask whether motivating belief distortion affects decisions about redistribution between two other individuals, as these beliefs are strong predictors of the demand for redistribution for meritocratic individuals. After the first phase of the experiment, in which subjects receive a signal about the environment to which they were assigned (i.e., Chance or Performance) and in which Motive-group subjects are informed about the second task, I give subjects the opportunity to redistribute an initially unequal bonus allocation between two uninvolved participants. These participants were previously recruited to perform the same first task as the decision maker herself. I truthfully tell participants that the initial allocation was determined by the same payment rule they themselves just faced. Subjects can then redistribute this initial allocation. By exploiting variation across treatment groups and signals, I can test (a) whether motivating belief distortion affects inequality-acceptance for other-regarding motives and (b) whether past experiences affect redistributive decisions above and beyond the experiences' informational content.

My results show that subjects distort beliefs to motivate future effort. Subjects who know they will perform another task in the same environment are significantly more confident (seven percentage points) of being in the Performance-Environment. This average effect masks heterogeneity by signal type. Motive-group subjects who received a disincentivizing signal indicating that reward is unrelated to effort, i.e. that they were assigned to the Chance-Environment, are significantly more confident (nine percentage points) that they are in the Performance-Environment compared to control group subjects who received the same signal. My results do not show any difference in beliefs across treatment groups for subjects who received an incentivizing signal indicating that reward is a function of effort, i.e. that they were assigned to the Performance-Environment. This shows that motivating belief distortion is particularly frequent if people receive information that is disincentivizing, i.e., information that indicates that effort is not important for success. By exploiting independent variation in the event that leads to a given signal, I show that motivating belief distortion is particularly pronounced for individuals who know (or believe) that they would have done well in a world that actually rewards effort, i.e., people who learned that the outcome of the task was not justified by their relative performance. Overall, the results of my experiment provide strong evidence that individuals distort their own luck-effort beliefs to motivate themselves for the task they expect to face in the future.

Turning to the results on the distribution decision, I show that motivating belief distortion

does not significantly affect distributive behavior. This suggests that beliefs may be instrumental for motivating future effort but this shift in beliefs is not strong enough to be reflected in aggregate distribution behavior. Even though subjects who are confident of being in the Performance-Environment are less likely to redistribute, I find that past experiences tend to matter a great deal for redistributive decisions in this context. Importantly, I find that subjects who did not win a prize and who performed worse than their competitor redistribute larger amounts than do other subjects, even though the former received a signal indicating a higher likelihood of being in the Performance-Environment. This result highlights the importance of taking into account event characteristics that are not informative about the relative importance of luck and effort when analyzing distributive behavior.

**Contribution to the Literature** Recent decades have produced a large amount of evidence that beliefs about luck and effort influence attitudes towards redistribution (see Cappelen et al., 2020, for a recent survey). Earlier work, using survey evidence, has demonstrated a robust correlation between the belief that economic inequality is due to luck or effort and the willingness to redistribute (Fong, 2001; Alesina and La Ferrara, 2005; Alesina and Giuliano, 2011). This correlation has been replicated in a more controlled setting using laboratory experiments. These experiments typically vary exogenously regarding whether differences in the initial allocation are due to luck or effort (e.g. Konow, 2000; Cappelen et al., 2007; Krawczyk, 2010; Durante et al., 2014; Lefgren et al., 2016). In these studies, the source of inequality is typically known to the subjects and there is no uncertainty about the true role of luck or effort. Cappelen et al. (2019) look at intensive margin differences in probabilistic luck-effort beliefs by informing subjects about the true probability that success is within one’s control. The present chapter contributes to this literature by testing empirically how individuals form beliefs about the role of luck or merit and, more specifically, how this belief formation interacts with future incentives faced by the decision maker.

By providing evidence about how individuals form luck-effort beliefs, the chapter helps build the micro-foundation of canonical models that explain distributive equilibria. Piketty (1995), for example, studies a model where individuals learn about the relative importance of luck and effort from their own or their ancestors’ past mobility experiences. Bénabou and Tirole (2006) study how motivated just-world beliefs may function as a commitment device. Alesina and Angeletos (2005) argue that differences in beliefs originate in historical experiences. Recent work has tested some predictions from these models by exogenously providing individuals with information about mobility and testing the effect of such information on beliefs and preferences for redistribution (Alesina et al., 2018; Gärtner et al., 2019). Gärtner et al. (2019) additionally test and find evidence in support of Bénabou and Tirole (2006)’s prediction that parents transmit distorted beliefs to their children if they expect relatively low levels of redistribution in the future. However, it remains unclear to what extent luck-effort beliefs are used as a motivational device.

Studies using lab experiments have focused mainly on an attribution bias in luck-effort beliefs in a static setting where beliefs have no motivating value (Deffains et al., 2016; Cassar and Klein, 2019; Fehr et al., 2020). These studies found that subjects who lost a contest attribute their

failure to luck and demand more redistribution ex-post, while those who won attribute their success to differences in effort and are less willing to redistribute income. Cassar and Klein (2019) additionally identify an ingroup bias in distributive decisions. A recent experiment by Valero (2020) shows that individuals attribute failure to luck and success to effort but she does not find any evidence that individuals distort beliefs to morally justify the self-serving implementation of low levels of redistribution. Erkal et al. (2021) show that an attribution bias also extends to the evaluation of others' decisions, such that outcomes that are bad for a group are more likely to be attributed to bad decisions while good outcomes are more likely to be attributed to luck. I advance this literature by testing whether luck-effort beliefs are distorted to motivate future effort and whether they are shaped by the incentive structure individuals face. This work, thus, advances our understanding of how the economic environment shapes luck-effort beliefs.

I contribute to a growing literature on motivated belief formation. While the recent empirical literature has made advances in identifying *how* individuals distort beliefs (e.g. Zimmermann, 2020; Chew et al., 2020), we still lack empirical evidence as to *why* individuals want to distort beliefs. Past work has shown that individuals distort beliefs to deceive others (Schwardmann and van der Weele, 2019; Charness et al., 2018) or to justify selfish behavior (Di Tella et al., 2015). Coutts (2019a) and Barron (2021) show that it is unlikely that individuals distort non-ego-relevant beliefs about the likelihood of being in a given state of the world, purely because they expect a higher income in that state of the world. A large part of this literature, starting with Möbius et al. (2014), looks at asymmetric updating after receiving feedback about relative ability and has found mixed results. Convincing explanations for this heterogeneity in results remains lacking (Benjamin, 2019). I advance this literature by explicitly testing whether individuals distort beliefs to motivate future effort while controlling what subjects typically learn about themselves from the feedback they receive. The idea of beliefs as a motivating device was first introduced by Carrillo and Mariotti (2000) and Bénabou and Tirole (2002) to model why individuals are persistently overconfident, which has been shown to have far-reaching consequences, such as excess entry into business (Camerer and Lovallo, 1999).

Few empirical studies explicitly examine whether individuals distort beliefs to motivate effort. Chen and Schildberg-Hörisch (2019) show empirically that higher confidence in one's own ability is related to higher effort provision. This validates a necessary theoretical condition such that there is a demand to distort beliefs about one's own ability for motivating purposes. Banerjee et al. (2020) study the effect of feedback spillovers across unrelated task on confidence in one's own ability to do well in a contest using an artefactual field experiment. While they find evidence that individuals engage in motivated belief distortion, their heterogeneity analysis indicates that this is driven by hedonic rather than instrumental motives. Huck et al. (2018) study performance and information avoidance in the presence of uncertain incentive schemes using a lab experiment. They find preferences for and against information and show that information avoiders outperform information receivers. König et al. (2019) provide field evidence that is consistent with the prediction that beliefs about the return to effort are inflated to motivate effort. Ambuehl (2017) shows that incentives to undertake an unpleasant task in the future affects information

acquisition about how pleasant the task is. The latter three studies are closest to mine because they study how variation in the return to effort affects either information acquisition, beliefs, or behavior. Other studies in this literature, however, hold the motive behind beliefs constant and vary information. My design generates variation in the motive and holds information constant. By varying the motive across individuals, I avoid the problem that the existence of the motive itself is unobserved and endogenous. Importantly, my design allows me to explicitly isolate belief distortion for motivational purposes from affective motives (i.e., deriving utility from the mere fact of believing in one state of the world). Furthermore, I can analyze whether motivating belief distortion interacts with the information that subjects face, as well as with the content of the message, as I vary rank and outcome while holding the informational content of the signal constant.

The remainder of this chapter is organized as follows: Section 2 derives the hypotheses I want to test using a simple framework; Section 3 describes the experimental design; Section 4 characterizes the experimental procedures; Section 5 presents the results; and Section 6 concludes.

## 2 Framework and Hypotheses

To motivate the experimental design, I build a simple framework that mirrors the situation subjects face in the experiment. For an in-depth analysis of a similar setting, see Bénabou and Tirole (2006). The general set-up is borrowed from Bénabou (2015). The aim of this framework is to characterize what generates the treatment effect, what are the main explanatory variables that are necessary to observe motivating belief formation, and what other motives could be relevant in my setting. Note that in this experiment, I do not aim to explain *how* individuals distort beliefs in a motivated way.<sup>3</sup> However, recent contributions to the literature have found strong evidence that individuals forget or recode negative signals about their intelligence (Chew et al., 2020; Zimmermann, 2020).

In this chapter, I am interested in testing the hypothesis that, in order to motivate themselves to work hard, individuals overestimate the importance of effort for obtaining a reward. This requires studying a situation where individuals are initially willing to work. A related, but different, hypothesis is that individuals may underestimate the importance of effort in order to morally justify to themselves the decision to not exert effort. This can be rationalized as self-handicapping, i.e., as a strategy to protect one's self-confidence by avoiding the outcome of an activity that is too informative about one's own ability (e.g. Berglas and Jones, 1978). While this latter hypothesis is interesting, it will not be part of the present's work empirical design because it is an affective motive for belief distortion, driven by the demand to maintain a positive self-image, which is not directly related to the demand to distort beliefs for motivating purposes. Furthermore, if this type of belief distortion is relevant in my setting, it would predict

---

<sup>3</sup>Numerous theoretical models have characterized *how* individuals distort beliefs. In brief, they assume either that agents ignore, forget, or recode past information (e.g. Bénabou and Tirole, 2002, 2006, 2011) or individuals literally choose the beliefs they want to hold and trade them off against the material (Brunnermeier and Parker, 2005) or mental costs (Bracha and Brown, 2012) of belief distortion (see Coutts, 2019b, for an experiment that tests both mechanisms).



the opposite treatment effect. I also will not address purely affective motives for belief distortion (i.e., the desire to believe in a state of the world with higher income prospects). While such motives are certainly interesting to study, they do not capture the notion that individuals distort beliefs to motivate effort, which is the central idea of this chapter. In the experimental design section below, I explain how I control for this competing motive.

Agents undertake an activity in a world where the return to their effort is uncertain. Depending on the state of the world, the return  $\theta$  is either high or low ( $\theta_H \geq \theta_L$ ). I also assume that each state of the world gives the agent an effort-independent payoff  $\kappa$  ( $\kappa \in \{\kappa_H, \kappa_L\}$ ). To relate this setting to the bigger picture, one can say that  $\theta$  characterizes the degree to which effort is rewarded and  $\kappa$  is the part of income that is independent of effort.  $\kappa$  characterizes, for example, inherited income, genes, or unconditional transfers. Note that for a risk-neutral agent, what matters is how much more she is able to earn from exerting effort. Hence, if her salary  $Y$  comprises an unconditional part  $a$  and an effort-conditional bonus  $b$ ,  $\theta$  would be equal to  $b$ . Further, the effect of taxation on the return to effort is already included in  $\theta$ . I do not model how citizens choose taxation, nor do I explicitly model the underlying redistributive system. I am interested in mapping the return to effort,  $\theta$ , to belief distortion. These are the parameters that I will manipulate in the experiment.

Agents have time-inconsistent preferences that I characterize using hyperbolic discounting. In period  $t$ , they maximize:

$$U_t = \frac{\delta^t}{\beta} U_t + \sum_{\tau=t+1}^T \delta^\tau U_\tau \quad (3.1)$$

$\delta$  represents time-consistent discounting while  $\beta$  characterizes the degree of present bias. As  $\beta$  decreases, utility in period  $t$  increases. For simplicity, I assume  $\delta = 1$  which is what Augenblick and Rabin (2019) actually find for the task that I use.

The framework has three periods: In period 3, the agent receives her payoff  $U_3 = \theta e + \kappa$ .<sup>4</sup> In period 2, the agent chooses whether or not to exert effort, given her belief  $p_2$ . In my experiment, this is a binary decision because I use threshold incentives (subjects earn  $\theta$  if their effort exceeds a certain threshold), which allows me to control by design for affective motives for belief distortion. Exerting effort  $e$  costs the agent  $c(e, p)$ , where  $c(e = 0, p) = 0$ . The  $p$  term in  $c(\cdot)$  characterizes a psychological cost from exerting effort if the agent believes that his effort is wasted or is a “lost cause” and unnecessary.  $p$  is the likelihood that her effort actually matters and  $c(\cdot)$  decreases convexly in  $p$  ( $c'_p(e, p) \leq 0, c''_{p,p}(e, p) \geq 0$ ), meaning that the psychological cost increases with the likelihood that her effort has no effect on the outcome. The cost of effort, both psychological and real, is paid in period 2, while the returns are reaped in period 3. Thus, her decision payoff in period 2 is:

$$\max_e E_2(U(e, c, \theta, \kappa)) = E_2\left(\theta e + \kappa - \frac{c(e, p)}{\beta}\right)$$

In the experiment, there are two states of the world,  $H$  and  $L$ . In period  $t$ , the agent believes

---

<sup>4</sup>In Prolific, where I run the experiment, subjects typically receive the payment after 1-2 days.

that she is in state  $H$  with probability  $p_t$  and in state  $L$  with probability  $1 - p_t$ . We can then rewrite the problem as

$$\max_e E_2(U(e, c, \theta, \kappa)) = e(p_2\theta_H + (1 - p_2)\theta_L) + (p_2\kappa_H + (1 - p_2)\kappa_L) - \frac{c(e, p_2)}{\beta}. \quad (3.2)$$

The agent chooses to exert effort in period 2 if:

$$\begin{aligned} E_2(U(e = 1, c, \theta, \kappa)) &\geq E_2(U(e = 0, c, \theta, \kappa)) \\ (p_2\theta_H + (1 - p_2)\theta_L) + (p_2\kappa_H + (1 - p_2)\kappa_L) - \frac{c(e, p_2)}{\beta} &\geq (p_2\kappa_H + (1 - p_2)\kappa_L) \\ (p_2\theta_H + (1 - p_2)\theta_L) &\geq \frac{c(e, p_2)}{\beta} \end{aligned} \quad (3.3)$$

Note that effort-unconditional income  $\kappa$  cancels out in the decision of a risk-neutral agent. This obviously, would be different if the agent's utility function is concave.<sup>5</sup> The inequality is more likely to hold, the larger  $p_2$  is (because  $\theta_H \geq \theta_L$  and because of its psychological effect on  $c(\cdot)$ ) and the larger  $\theta$  is. It is also more likely to hold if the direct cost of effort  $c$  and the degree of present bias  $\beta$  are low. In the experiment, I create exogenous variation in  $\theta_H$  (across treatment groups). Note that  $p_2$  is the motivated belief and, thus, is endogenous in this model.

We now turn to the behavior of the agent in period 1. The agent in period 1 receives a signal  $\sigma$  that is informative about the true state of the world. She updates her belief in a “non-motivated” way and we denote her posterior  $p_1$ .<sup>6</sup> The agent now has the opportunity to ignore, forget, recode, or reinterpret the signal they receive.<sup>7</sup> I assume that this comes at a mental cost  $M(p_1, p_2)$ . In the spirit of Bracha and Brown (2012), I assume that the mental cost of self-deception is increasing in the absolute *difference* between the non-motivated belief she holds in period 1,  $p_1$ , and the belief she recalls in period 2,  $p_2$ . Additionally,  $M(p_1, p_2)$  can also stand for the cost forgone in the belief elicitation task. Hence, if she maintains her non-motivated belief, she does not incur any mental or material cost from distorting beliefs. A period-1 agent maximizes her expected utility with respect to  $p_2$ , which is her choice variable:

$$\begin{aligned} \max_{p_2} E_1(U(e, c, \theta, \kappa, p_2)) &= e(p_1\theta_H + (1 - p_1)\theta_L) + (p_1\kappa_H + (1 - p_1)\kappa_L) \\ &\quad - c(e, p_2) - \frac{M(p_1, p_2)}{\beta} \end{aligned} \quad (3.4)$$

Note that she maximizes expected utility by anticipating the effect that  $p_2$  has on the period-2 agent's decision about whether to exert effort. The internal solution to (3.4) is the value

<sup>5</sup>If utility is concave, the threshold  $p_2$  for which the inequality holds is higher compared to the linear case, *ceteris paribus*. This is due to the fact that the decision maker is now averse to making the mistake of paying the cost of effort without receiving a reward. Furthermore, (3.3) is less likely to hold for low values of  $\kappa_L$  because losses from exerting effort without receiving a reward now have a large impact on the margin. If  $\kappa_H$  is low and  $\kappa_L$  is high, the opposite is true, as losses have a low effect on the margin, compared to the gains from exerting effort. I want to highlight that the comparative statics that I focus on in this chapter remain the same: The inequality in (3.3) is more likely to hold, the larger  $p_2$  is.

<sup>6</sup>“Non-motivated” means that agents may form non-Bayesian beliefs but they are restricted to not act on the basis of any motive (affective or instrumental).

<sup>7</sup>See the beginning of this section for a discussion concerning the supply side of motivated beliefs.

for  $p_2$  that equalizes the marginal gain from belief distortion to its marginal cost ( $c'_{p_2}(e, p_2) = M'_{p_2}(p_1, p_2)$ ,  $p_2 \in [0, 1]$ ). Note that this is not necessarily the *optimal* belief, as the optimal belief has to satisfy the constraints an agent faces, which I detail below.

A period-1 agent has two reasons to distort beliefs: first, to counteract a misallocation of effort due to present bias and, second, to ease the psychological burden of working in an environment where it is uncertain whether it is actually worthwhile to undertake this activity. If time-inconsistent preferences are the main driver behind motivating belief distortion, one would only expect the period-1 agent to distort beliefs if she thinks that the period-2 agent is unlikely to work under non-motivated beliefs.<sup>8</sup> If the psychological channel is the relevant mechanism behind motivating belief distortion, then one would still expect subjects to distort beliefs for motivating purposes if they are not sophisticated about their present bias and if the period-2 agent would complete the task under non-motivated beliefs.

We can now ask under which conditions the period-1 agent is willing to distort beliefs. In our simple model, this will be the case if, in period 1, she plans to undertake the task herself in period 2 and if the cost of distorting her beliefs is not too high. Denoting the optimal belief  $p_2^*$  (or her initial belief,  $p_0$ , if we assume that agents distort by forgetting), we can formulate the necessary condition for distorting beliefs for motivating purposes as:

$$\begin{aligned}
 (p_1\theta_H + (1 - p_1)\theta_L) + (p_1\kappa_H + (1 - p_1)\kappa_L) - c(e = 1, p_2^*) - \frac{M(p_1, p_2^*)}{\beta} &\geq (p_1\kappa_H + (1 - p_1)\kappa_L) \\
 (p_1\theta_H + (1 - p_1)\theta_L) - c(e = 1, p_2^*) &\geq \frac{M(p_1, p_2^*)}{\beta} \quad (3.5) \\
 (p_1(\theta_H - \theta_L) + \theta_L) - c(e = 1, p_2^*) &\geq \frac{M(p_1, p_2^*)}{\beta}
 \end{aligned}$$

The agent distorts beliefs if she would want to work in period 2, evaluated from period-1 agent's point of view ( $LHS > 0$ ), and if the cost associated with distorting beliefs is not too high. Note that the period-1 agent is fully aware that the expected return to effort is a function of  $p_1$  and not of  $p_2$ . This property is important for the experimental design, as it implies that motivating belief distortion can only be identified if individuals are, in principle, willing to work under  $p_1$ . For example, if  $\theta_H = \theta_L = 0$  or if  $c(e, p_2 = 1)$  is very high, one should not expect individuals to engage in belief distortion because they would not want to engage in the task in the first place. While this is impossible to test empirically, because we only observe the period-2 agent's decision to exert effort, I can set the incentives for the task such that most of the subjects complete it, which implies that  $LHS \geq 0$  in (3.5). Furthermore, under  $p_2^*$ , (3.3) has to hold as well, which is relevant for agents with low  $\beta$ . For them, the cost of belief distortion may be too high for (3.3) and (3.5) to hold simultaneously.

Next, I ask what constitutes  $p_2^*$ . As mentioned previously,  $p_2^* = p_0$  if subjects distort beliefs by forgetting the signal they received. However, if subjects have more freedom in choosing the beliefs they want to hold, then  $p_2^*$  is either the belief that maximizes (3.4), denoted as  $p_2^{max}$ , or it is the minimal  $p_2$  that satisfies (3.3). The latter is the case if the period-2 agent would not be

<sup>8</sup>In this case, the demand to distort beliefs arises from the need to correct for a misallocation of effort due to present bias. If present bias has no significant impact on the effort chosen, then a subject would not distort beliefs for this reason.

willing to work under  $p_2 = p_2^{max}$  but would also be willing under  $p_2^*$ , which then has to satisfy both (3.3) and (3.5). This would be the case if the agent has a relatively strong present bias (i.e.,  $\beta$  is low) compared to the psychological burden characterized by  $c(\cdot, p)$ .

How does variation in  $p_1$  affect  $p_2^*$ ? In both cases described above, one would expect more belief distortion if  $p_1$  is low, conditional on (3.5) holding. If countering present bias is the driving force behind belief distortion, (3.3) is less likely to hold for low  $p_1$ , triggering a higher demand for belief distortion for these subjects. If the psychological cost is the relevant mechanism, we would expect more belief distortion for low  $p_1$  because (3.3) is less likely to hold under  $p_1$  and because the return to belief distortion is higher under the assumption that  $c(\cdot)$  is convex in  $p$ .

**Variation induced in the experiment** This experiment varies  $p_1$  and  $\theta_H$ , while keeping  $\theta_L = 0$ . It varies  $p_1$  within treatment groups by varying (Bayesian) posteriors. It varies  $\theta_H$  across treatment groups: The no-motive-group has  $\theta_H = \theta_L = 0$  because it does not know that it will face a task, while for the motive-group  $\theta_H > 0, \theta_L = 0$ . At first it is useful to note that no-motive-group subjects should never distort beliefs to motivate future effort because (3.3) never holds (i.e., the return to future effort is always 0 no matter the state of the world because there is no second task). This means that, for no-motive-group subjects,  $p_1 = p_2$ . Motive-group subjects, however, have an incentive to distort beliefs in order to counter the under-provision of effort due to their time-inconsistent preferences or to weaken the psychological burden of potentially working for no good reason.

I further vary  $p_1$  across subjects. This variation allows me to assess what type of information triggers motivating belief distortion. Because I create a situation where exerting effort is desirable for the period-1 agent, I expect belief distortion to be stronger for subjects with low  $p_1$ , i.e. those who received a disincentivizing signal.

This leads me to my first set of hypotheses (H1):

H1.1 Motive-group subjects are, on average, more confident that the state of the world is  $H$  than are no-motive-group subjects;

H1.2 Belief distortion is more prevalent if the signal is disincentivizing;

H1.3 Subjects with high effort costs or low returns to effort distort luck-effort beliefs.

In the experiment, I test H1.1 by comparing beliefs across treatment groups. I test H1.2 by comparing belief distortion among individuals who received an incentivizing signal with those who received a disincentivizing signal. I test H1.3 by testing whether motivating belief distortion is driven by subjects who either are not very good at the task at hand or do not enjoy the task.

**Belief distortion and redistributive behavior** The second set of hypotheses tests predictions concerning distributive decisions by spectators with regard to two *other* individuals. To derive these predictions, I build on the evidence that individuals hold normative fairness views and care about the distribution of income across other individuals (see e.g. Almås et al., 2020). I follow the general framework introduced by Cappelen et al. (2013), which was later applied

to a setting very similar to the one studied here (Cappelen et al., 2019), and I refer to their analysis in the remainder of this section.

Concretely, one agent (spectator) is matched with two workers who have different levels of *initial* income. One of the workers (worker  $H$ ) has an initially higher income  $y_H^{pre}$  than the other worker (worker  $L$ ), i.e.  $y_H^{pre} > y_L^{pre}$ . The spectator then has the choice to redistribute income from worker  $H$  to worker  $L$ . The final distribution of income is characterized by the pair  $\mathbf{y}^{post} = (y_H^{post}, y_L^{post})$ . I further assume that the spectator cares about the final distribution of income and seeks to implement an ex-post distribution of income that is as close as possible to her normative fairness views specified as  $\mathbf{y}^{fair} = (y_H^{fair}, y_L^{fair})$ . This decision is characterized by the spectator minimizing the following loss function by choosing  $\mathbf{y}^{post}$ :

$$V(\mathbf{y}^{post}) = -(|\mathbf{y}^{post} - \mathbf{y}^{fair}|) \quad (3.6)$$

The literature has identified three common fairness types (see Almås et al., 2020): (1) libertarian fairness ideal, where the spectator considers the initial distribution of income as fair, i.e.  $\mathbf{y}^{fair} = \mathbf{y}^{pre}$ ; (2) egalitarian fairness ideal, where both workers should get the same amount of income, i.e.  $\mathbf{y}^{fair} = (\frac{y_H^{pre} + y_L^{pre}}{2}, \frac{y_H^{pre} + y_L^{pre}}{2})$ ; and (3) the liberal egalitarian or meritocratic fairness ideal, where the worker that produced more units should also get a higher income ex-post, thus  $y_H^{fair} > y_L^{fair}$  if worker  $H$  produced more units and  $y_L^{fair} > y_H^{fair}$  if  $L$  produced more units. A meritocrat prefers an equal split if they produced the same amount of units.

In the setting I am interested in, the spectator faces uncertainty about which of the two workers produced more units. While this will not affect the decision made by libertarians or egalitarians, it does affect the decision by meritocrats. Their decision depends on their *beliefs* about which of these two workers was the better performer. These beliefs are affected by spectator's assessment of whether the initial distribution of income reflects differences in effort. If the spectator believes that the initial distribution of income  $\mathbf{y}^{pre}$  reflects differences in effort, then  $y_H^{fair} > y_L^{fair}$ .

If we now assume that there are two states of the world, one where effort matters for the initial allocation of income (true state of the world with probability  $p$ ) and another where one's income is based entirely on luck (true state of the world with probability  $1 - p$ ), then the initial allocation of income serves as a noisy signal regarding who was the better performer. Concretely, the likelihood that the agent with higher initial earnings is also the higher performer increases with  $p$ . Following Cappelen et al. (2019), I rewrite the expected value function for meritocratic spectators, where the posterior distribution of income  $\mathbf{y}^{post}$  is characterized as the share of total income given to the initially high earner, denoted as  $s_H$ , while the total income is normalized at 1. This implies that the share of the initially low earner is  $s_L = 1 - s_H$ .  $s_{Best}^{Fair}$  is the share of income a spectator prefers to allocate to the best performer.

$$\begin{aligned} E(V(s_H^{Post}, s_{Best}^{Fair}, p)) &= -P(\text{H is best}|p)(|s_H^{Post} - s_{Best}^{Fair}|)^\alpha \\ &\quad - (1 - P(\text{H is best}|p))(|s_H^{Post} - (1 - s_{Best}^{Fair})|)^\alpha \end{aligned} \quad (3.7)$$

The first part in (3.7) characterizes the loss function for a given amount redistributed if  $H$  was indeed the better performer, while the second part characterizes the loss function if  $L$  was the better performer.  $\alpha$  is an elasticity parameter. Importantly,  $P(H \text{ is best}|p)$  depends on the true state of the world and the likelihood that the high performer wins the initial prize in either state of the world. One can thus characterize it as

$$P(H \text{ is best}|p) = p^M p + p^C (1 - p). \quad (3.8)$$

$p^M$  is the likelihood that a high performer wins the initial prize in the state of the world where effort matters and  $p^C$  is the likelihood that a high performer wins the initial prize in the state of the world where winning is entirely a matter of luck. To capture this notion, I set  $p^M > 0.5$  and  $p^C = 0.5$  and  $P(H \text{ is best}|p)$  is then strictly *increasing* in  $p$ . This implies that the first part of (3.7) increases as  $p$  increases while the second part decreases, which induces meritocratic spectators to allocate a higher share to  $H$  as the likelihood that  $H$  is actually the high performer increases. These spectators thus want to further decrease the gap between  $s_H^{Post}$  and  $s_{Best}^{Fair}$ . The extent to which variation in  $p$  leads to inequality acceptance depends on  $\alpha$ . For  $\alpha = 1$ , meritocratic spectators would not redistribute at all, since the initial distribution is informative about who was the better performer. As  $\alpha \rightarrow \infty$ , meritocratic spectators will always redistribute, unless  $p = 1$ .

In the experiment, I study how a variation in  $p$  due to motivating belief distortion affects the redistributive behavior of spectators. I do this by comparing redistributive decisions made by subjects who distort beliefs for motivating purposes and those who do not hold motivating beliefs. Because meritocratically inclined subjects are the only subjects who should react to variation in  $p$ , I expect that spectators are, in the aggregate, weakly less willing to redistribute from  $H$  to  $L$ . Whether the variation in  $p$  due to motivating belief distortion is strong enough is an empirical question because it depends on  $\alpha$  and also on the share of meritocrats in the population. Furthermore, any potential difference in beliefs across the treatment groups should be mediated through variation in beliefs.

This leads to the second set of hypotheses (H2) that I want to test through my design:

H2.1 Motive-group subjects redistribute less than no-motive-group subjects;

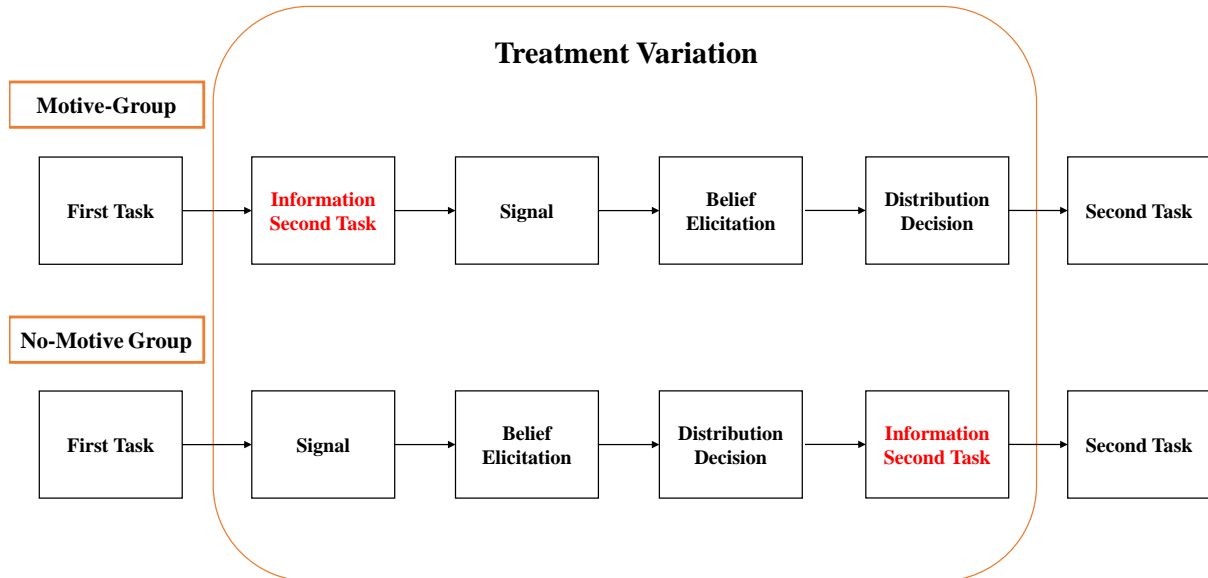
H2.2 The treatment effect on redistribution is mediated by differences in beliefs;

H2.3 The treatment effect is stronger for subjects who hold meritocratic ideals.

As mentioned above, motivating belief distortion may not be strong enough to significantly shift average beliefs around the prior and the relationship through which  $p$  affects inequality acceptance may be convex. This means that, on the one hand, evidence in favor of my hypothesis is relatively strong. On the other hand, it means that, in case of a null-result, a more granular design may be needed to identify the relationship between motivating belief distortion and redistributive preferences due to other-regarding reasons.

### 3 Experimental Design

Figure 3.1: Schematic diagram characterizing the stages of the experiment



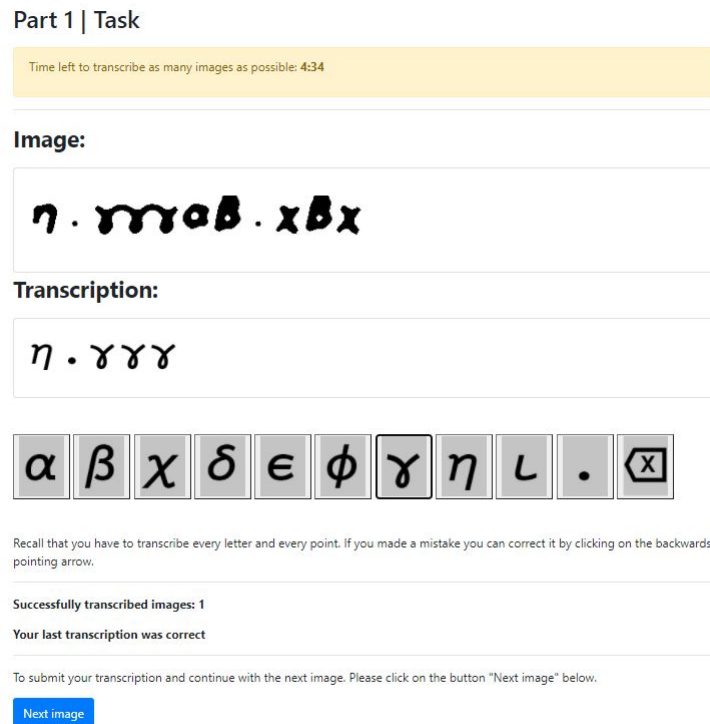
The basic set-up of the experimental design is shown in the schematic diagram in Figure 3.1. Subjects begin the experiment by engaging in a tedious real effort task (First Task) in an uncertain environment. They either complete the task in a setting where the likelihood of receiving a reward is increasing in their effort (Performance-Environment), or is based entirely on luck (Chance-Environment). After finishing the task, subjects receive a noisy signal informing them whether they are in the Chance- or Performance-Environment. The aim of this signal is to induce variation in baseline beliefs about the state of the world; I then elicit these beliefs. At the end of the experiment, subjects will engage in a second task. As in the first task, the payment rule depends on the environment that was drawn at the beginning of the experiment. The purpose of the second task is to provide a motive to distort beliefs for motivating purposes. Subjects that are informed about the second task may now overestimate the likelihood of being in the Performance-Environment in order to motivate themselves to work hard on the second task. To test for this hypothesis, I vary the point in time at which I inform subjects about the second task. Subjects assigned to the Motive-Group treatment are informed about the second task *before* belief elicitation. Therefore, they have an incentive to distort beliefs for motivating purposes. This is not the case for subjects in my control treatment group, the No-Motive-Group. These subjects are informed about the second task *after* belief elicitation. Therefore, they do not have an incentive to distort beliefs for motivating purposes. I thus vary the motive to distort beliefs for motivating purposes across treatment groups and by comparing beliefs across the two groups, I can identify whether subjects engage in motivating belief distortion. Note that the only difference across the two groups is the point in time at which they are informed about the second task. Importantly, the *objective* information they have about the state of

the world is equally distributed across the groups. To test whether motivating belief distortion affects inequality acceptance due to other-regarding motives, I pair subjects with two agents who previously participated in the first task and, importantly, were in the same environment—Chance or Performance—as the decision maker herself. One of the agents won the bonus in the first task, while the other did not win it. The subject is given the opportunity to redistribute income from the agent who won the initial prize to the other agent. By comparing the amount redistributed, I can infer whether motivating belief distortion affects inequality acceptance due to other-regarding motives. In the following subsections, I present each part of the experiment in detail.

### 3.1 Work Phase 1

Subjects start the experiment by engaging in a real effort task under (noisy) tournament incentives. The task was chosen because it is tedious and it is relatively independent of skill or intelligence. The incentive scheme was chosen to generate noisy signals about the environment to which the subject was assigned—Chance or Performance—that are not co-linear with effort or with success or failure in the task.

Figure 3.2: Screenshot from the real-effort task



The task, introduced by Augenblick et al. (2015), consists of transcribing as many images containing 11 Greek letters as possible in 5 minutes. I chose this task for three reasons: First, it is a very tedious task that has been shown to actually generate time-inconsistent behavior (indeed, this is the topic of Augenblick and Rabin (2019), which uses the same task) and may be particularly sensitive to limited willpower. Second, the task does not depend heavily on inherent skills, such as being good in math, which is important because it gives everybody a



realistic chance to complete the task. Third, relative performance on this task may be less likely to affect one’s self-image than other tasks used in the literature on overconfidence, such as typical IQ tasks. This third characteristic enables me to generate a benchmark finding in a setting where self-image concerns play a minor role.

Subjects know that they are in one of two environments. If the environment is Performance, there is a positive correlation between winning a prize of €7.50 and exerting more effort. If the environment is Chance, effort is orthogonal to winning the prize. Agents do not know the true state of the world but are informed that they are in either environment with probability 0.5. I chose a prior probability of 0.5 because this prior is easily understood and memorized by subjects and it allows me to generate a sufficiently large share of disincentivizing signals that yield a lower posterior likelihood of being in the Performance-Environment (see below).

Table 3.1: Likelihood of winning the prize by group rank and environment

Rank	Likelihood to win		
	Overall	If Performance	If Chance
1	0.65	0.8	0.5
2	0.35	0.2	0.5

After completing the task, the subject is matched with one other player (henceforth called “competitor”) who completed the identical task in a previous pilot study. As shown in Table 3.1, the likelihood of winning the prize varies by the environment and rank (whether the subject performed better than her competitor). If the subject is in the Performance-Environment, the likelihood of winning the prize is increasing in the effort she exerted: If she ranks first, the likelihood that she wins is equal to 0.8; if she ranks second, it is equal to 0.2. If she is in the Chance-Environment, the likelihood of winning the prize is always equal to 0.5. Taking both together, the *prior* likelihood of winning if she ranks first is equal to 0.65 and if she ranks second it is equal to 0.35.<sup>9</sup>

### 3.2 Feedback Phase

After completing the task, subjects receive two pieces of information that compose their signal: (1) Their effort rank within the group and (2) whether they won the prize. *Together*, these two pieces of information enable subjects to form posterior beliefs about the likelihood of being in the Chance- or Performance-Environment.

The structure of the signal is a crucial part of my design because it generates exogenous variation in (non-motivating) baseline beliefs. First, the mechanism will match individuals

<sup>9</sup>It may be that the prior likelihood of winning if ranked first is perceived as too low and the subjects do not want to exert effort in the first place. This type of behavior was, however, quite rare. Only 4.6% of subjects did not transcribe any image correctly in Work Phase 1. Beyond that, subjects are informed how they rank relative to their competitor in the subsequent part of the experiment. Hence, each signal is equally revealing for all subjects.

Table 3.2: Bayesian posterior likelihood of being in the Performance-Environment

Rank	Outcome	
	Win	Lose
1	0.62	0.29
2	0.29	0.62

**Note.** *Rank* is the subject's effort rank within her group. The column *win (lose)* is the Bayesian posterior probability of being in the Performance-Environment if the subject won (lost) the prize for every rank.

with equal performance to different competitors. This gives me exogenous variation in being ranked first or second. Second, there is exogenous variation in winning and losing the prize because a random component plays a role in the Performance-Environment. This implies that different subjects may perform equally well at the task but end up with signals that point in opposite directions, allowing me to study the heterogeneity in treatment effects across signals. Furthermore, I observe the signal's *objective* information, which does not depend on unobserved variables such as subjective beliefs about relative performance. This gives me the necessary level of control over baseline beliefs that I need to determine what type of information triggers motivating belief distortion.

Furthermore, both sources of variation enable me to test causally whether individuals update differently after receiving objectively the same information through a message that indicates whether or not they won the bonus. For example, an individual who transcribed fewer images than her competitor may not make the same inference from losing the prize as an individual who transcribed more images than her competitor and wins the prize. This enables me to elucidate whether the message that conveys the signal affects beliefs as such and whether this interacts with my main treatment variation. Seen from a broader perspective, this variation allows me to ask how economic experience, such as failure and success, affects belief in a just world above and beyond its informational content.

In sum, the feedback stage provides subjects with two pieces of information: (1) how many words they transcribed relative to their competitor and (2) whether they won or lost the prize. These two pieces of information together enable subjects to infer the relative likelihood of being in the Chance- or Performance-Environment.

### 3.3 Information Treatment

Before receiving the signal, half of the subjects are informed about a future part of the experiment. Though the remainder of the experiment is equal across the two treatment groups, their knowledge about it varies: Subjects in the motive-group are informed at this stage that they will engage in a second work-phase where they do the same task as in work-phase 1 but

under threshold incentives (see Section 3.6 for details). Crucially, the environment, drawn at the beginning of the experiment at the subject level, remains constant and matters for work-phase 2: If the subjects are in the Performance-Environment, they win a prize of €5 with probability 1 if they transcribe more than 20 images. If they transcribe fewer images, they do not get the reward. If they are in the Chance-Environment, they win the prize with probability 1, independent of the number of letters they transcribe. Hence, their expected earnings are equal across both states of the world if they plan on completing the task. This alleviates concerns that subjects distort beliefs because they anticipate a higher income in one environment compared to the other. Subjects in the no-motive-group face the identical task but are informed about this *after* the belief-elicitation stage.

This treatment variation enables me to identify whether *elicited* beliefs are distorted by the motive to motivate future effort because it provides a “clean counterfactual” (see Coutts, 2019a, for a methodological discussion). The experiment is identical for both groups up to the point that motive-group subjects are informed about the next task. Importantly, the objective information that subjects receive is equally distributed across the two treatment groups. Hence, even if subjects engage in other forms of non-Bayesian belief updating, a difference between the two treatment groups can only be explained through the fact that motive-group subjects know that they will have to redo the task in the same environment as before. This makes my results robust to functional form assumptions. Most importantly, priors and signals are balanced across treatment groups.<sup>10</sup>

While my hypothesis is that the treatment increases the motivational value of beliefs, it may also be the case that my treatment yields excuse-driven behavior that leads to a *negative* treatment effect (rather than the hypothesized positive effect). As briefly mentioned at the beginning of Section 2, subjects may want to bias their belief downwards in order to morally justify the decision not to work. This should be relevant for motive-group subjects who do not exert any effort at all or maintain positive self-confidence. This is, however, rarely the case, as I show at the end of Section 3.6.

Note that the motive-group subjects receive the information *before* they receive the signal. This implies that I equalize the time between learning about the task’s outcome and belief elicitation across subjects. This alleviates concerns that a potential treatment effect can be rationalized by differences in the time that has passed between the reception of the signal and belief elicitation.

### 3.4 Belief Elicitation Phase

In the next part of the experiment, I elicit subjects’ probabilistic beliefs about being in the Performance-Environment. Belief elicitation takes place before the redistribution stage (see above). Hence, at the point of belief elicitation, motive-group subjects are already informed about the second task, but this is not the case for no-motive-group subjects. To elicit beliefs, I

---

<sup>10</sup>My information condition may induce motive-group subjects to take on an external gaze and make more rational assessments of beliefs because they are primed to think more about the true state of the world. While this could indeed be a difference across the treatment groups, it would imply that motive-group subjects are more reactive to the treatment, which goes against my hypotheses.

follow Schwardmann and van der Weele (2019), who use a variant of the BDM to elicit beliefs. Importantly, the method is incentive-compatible to risk-aversion and asks subjects to choose a likelihood between zero and one-hundred that is divisible by ten.<sup>11</sup> Following Danz et al. (2020), I chose to not instruct subjects about the details of the payment mechanism, but I simply instructed them that it is in their best interest to select the likelihood that is closest to what they truly believe to be true. They can, however, click on a button that gives them detailed information about the payment mechanism and 23% of subjects did eventually click on the button. Subjects can earn €2.5 for this decision.

### 3.5 Redistribution Stage

To test how motivating belief distortion affects redistributive behavior among two *other* individuals, I ask subjects to redistribute income between two workers who were previously recruited on Prolific to do the same transcription task under the same payment rule as they faced themselves in work-phase 1. Building on the design used in canonical spectator experiments (e.g. Almås et al., 2020), I inform the subjects that one of the two workers with whom they are matched received a bonus of €4, while the other did not receive the bonus.<sup>12</sup> The two workers were each others' competitors and this information was explicitly communicated to the subjects. This implies that the subjects know that if they are in the Performance-Environment, the worker endowed with the initial bonus transcribed more images with 80% probability; if they are in the Chance-Environment, the worker endowed with the initial bonus transcribed more images with 50% probability. Beyond that, subjects are informed that the two workers do not know who earned the original bonus, nor did they know the state of the environment in which they performed the task. Subjects are also instructed that their decision will remain anonymous.

After reading the instructions, subjects could redistribute the initial endowment among the two subjects. Subjects had the option to either not redistribute at all, redistribute €1 from Worker A to Worker B, redistribute €2 from Worker A to Worker B, redistribute €3 from Worker A to Worker B, or redistribute €4 from Worker A to Worker B. After a subject made her decision, I ask to what extent she agrees with the statement that the worker that transcribed more images should also be paid more. This gives me a broad measure of whether the subject is a meritocrat.<sup>13</sup>

Observing subjects' distribution decision allows me to ask whether motivating belief distortion affects preferences for redistribution due to other-regarding motives. This can be done non-parametrically by comparing average amounts redistributed across the two groups. The

---

<sup>11</sup>Subjects are asked to indicate a probability  $P$  that is divisible by ten and makes them indifferent between winning a monetary prize with probability  $P$  and receiving the same prize if they are in the Performance-Environment. After having indicated this probability, the computer randomly chooses an integer  $x$  that is divisible by ten and lies between zero and one-hundred. If  $x$  is higher than  $P$ , the subject will participate in a lottery where she wins the prize of €2.5 with probability  $x$ . If  $x$  is lower than  $P$ , she wins the prize if she actually is in the Performance-Environment.

<sup>12</sup>The bonus is smaller than the one the subjects could earn themselves because the experiment involving the workers was shorter in duration.

<sup>13</sup>In the pilot study, I asked subjects how they would redistribute the endowment knowing that they were in the Performance- or the Chance-Environment using the strategy method. In the main experiment, I refrained from this to avoid diluting the incentives for the redistribution decision of interest.

comparison across the two groups should capture the effect of motivating belief distortion on inequality acceptance due to other-regarding motives.<sup>14</sup> I can also go further and conduct exploratory analyses where I ask whether the type of feedback affects distributive decisions. Importantly, I can analyze whether experiencing success oneself affects distributive decisions among others.

### 3.6 Work-Phase 2

Subjects engage in a second work phase after belief elicitation. This task serves as an incentive to distort beliefs for motivating purposes. The idea is that subjects distort baseline beliefs to motivate themselves to complete this task. The real-effort task is identical to the one used in Work-Phase 1 but incentives are different. The incentives are designed with three goals in mind that allow me to isolate the motive to distort beliefs for motivating purposes: First, the payment rule depends on the environment drawn at the beginning of the experiment; second, the incentives have to be strong enough such that all subjects want to engage in the second work phase to allow for the existence of the motive; and third, I want to equalize expected payoffs across the two environments, conditional on having completed the second task, to exclude the alternative hypothesis that subjects may prefer to believe that they are in the environment with higher income prospects.<sup>15</sup>

Thus, I chose to use the following incentive scheme for the second work-phase: Subjects have up to ten minutes to transcribe twenty images. The payment scheme depends on the environment drawn at the beginning of the experiment and that remains constant throughout the whole experiment. Subjects win a bonus of €5 with probability 1 if they are in the Performance-Environment and completed the task. If they do not complete the task, they will not win the bonus. They win the bonus with probability 1 if they are in the Chance-Environment, no matter how many words they transcribed. If they do not want to complete the task, they have the opportunity to quit the task and proceed with the post-experimental questionnaire.

I use threshold incentives because it allows me to control expected gains from exerting effort in a very salient manner. Importantly, the payoff is equalized across the two environments, conditional on planning to undertake the task. Furthermore, threshold incentives make the effort decision binary. Hence, I know whether the participation constraint, defined in my framework, holds or not. This is essential because it allows me observe whether subjects were actually willing to work on the second task, even if they received a disincentivizing signal.

Due to this design choice, I am not able to study the effect of motivating belief distortion on effort choices. As mentioned above, I can only identify motivating belief distortion for subjects who initially want to complete the second task, even if they received a disincentivizing signal.<sup>16</sup>

---

<sup>14</sup>One should take into account that the two groups differ in that motive-group subjects know that they will engage in another task and this may affect their preferences for redistribution. Specifically, motive-group subjects face higher income prospects for the rest of the experiment, which could have a countervailing effect on inequality acceptance. This would be a particular concern if most subjects decide to fully redistributed in the experiment.

<sup>15</sup>I want to control for this affective motive for belief distortion, which is different from the kind of motivated belief distortion that is the focus of this study.

<sup>16</sup>To put it differently, the period-1 agent in my framework, who holds non-distorted beliefs, should always want to attempt to transcribe the twenty images in Work-Phase 2.

This implies that I do not expect a difference in effort across the two groups and, for that reason, I did not include an analysis of Work-Phase 2 in my pre-analysis plan.

The notion that exerting effort in Work-Phase 2 is desirable for all subjects is further supported by my data. 93% of subjects did not exit the task prematurely. This shows that subjects are willing to attempt to reach the threshold. This is further supported by looking at the attempts subjects made. More than 90% of subjects attempted to transcribe at least 20 images and more than 95% attempted to transcribe at least 5 images.

### 3.7 Post-experimental questionnaire

At the beginning of the post-experimental questionnaire, I ask subjects to recall (a) their rank and (b) whether they won the bonus for Work-Phase 1 or not. Subjects also had the option to state “I do not recall.” This line of questioning allows me to observe whether subjects are more prone to forget some signals than others. This information also sheds light on the supply-side mechanism of motivating beliefs.

To obtain an incentivized measure of whether subjects use commitment devices to motivate future effort, I offer subjects the opportunity to perform the same task two weeks later under a piece-rate incentive scheme. They can earn up to €10 by transcribing 40 images. Crucially, if they transcribe fewer strings, they receive a partial payment because each string is rewarded with €0.25; i.e. if they transcribe 10 images, they receive €2.50. I first ask subjects how many images they would wish to transcribe if they were reinvited. I then give them the chance to choose a minimum effort level (minimum number of transcribed letters). If they transcribe fewer images than this threshold, they earn nothing. If they transcribe more images, the usual piece-rate scheme is applied. Hence, subjects can make it costly for their future selves if they choose to work less than the amount specified in their minimum production level. Putting both (plan and commitment) together allows me to construct a measure of demand for commitment.<sup>17</sup> The decision is implemented for 5% of subjects.

Subjects are then asked to fill out a post-experimental questionnaire. The questions cover demographics, risk and time preferences, and two attention checks. The question items are provided in Appendix 3.C.

### 3.8 Summary of the variation generated through the design

Before going through the results of the experiment, it is worth taking a step back to summarize the exogenous variation generated through my design. My main treatment variation is whether subjects have an incentive to distort beliefs for motivating purposes at the point of belief elicitation. The treatment is randomized at the subject level and each subject had an equal probability of being in the motive- or no-motive-group. In the experiment, 248 subjects were assigned to the no-motive-group, while 252 were assigned to the motive group. This variation is used to identify a motive-group effect.

---

<sup>17</sup>A contract in the same spirit was studied in the field by Kaur et al. (2015). They show that there is a demand for this type of “dominated” contract and that it significantly increases productivity and earnings.

Across subjects, I further vary the feedback that subjects receive. There, I exploit two sources of exogenous variation: First, winning or not winning the bonus is, by design, random once I take into account the performance rank. This means that two subjects may both be better or worse than their competitor but one ends up winning the bonus while the other does not win the bonus. Second, being ranked first or second is exogenous, once I control for the subjects' performance. The intuition is that two equally talented subjects may both win or lose the bonus but end up being ranked first and second because they got matched to different competitors.

This yields variation within treatment groups to obtain an incentivizing and a disincentivizing signal. Overall, 35% of subjects receive an incentivizing signal, while 65% of subjects receive a disincentivizing signal.<sup>18</sup> Holding the informational content of the feedback constant, it is equally likely that a subject receives an incentivizing (disincentivizing) signal through a message stating that she won the bonus and ranks first (second) or that she lost the bonus and ranks second (first). I exploit this variation in my analysis to test whether the informativeness of the signal interacts with my treatment effect.

## 4 Procedure

Subjects proceed through every part of the experiment (Work-Phase 1, Belief Elicitation, Redistribution, Work-Phase 2). At the end of the experiment, one part will be randomly drawn to become payoff relevant.<sup>19</sup> Subjects know that there are multiple parts at the beginning of the experiment and each part is explicitly introduced and concluded.

All subjects were recruited using Prolific (see Palan and Schitter, 2018, for information about using Prolific as a subject pool). Subjects could only participate in the experiment once. The interface was programmed using oTree (Chen et al., 2016). I only invited subjects who reside in the US or Great Britain and are citizens of either country. This ensured that subjects were fluent in English and capable of understanding the instructions. Beyond that, I only invited subjects who have completed 10 or more surveys and whose submissions had been accepted at least 95 percent of the time. As pre-specified, I do not include subjects in my analysis who did not pass the two attention checks. Furthermore, all subjects who finished the study passed the first comprehension test in no more than two tries. Subjects were informed before they started the comprehension test that they cannot continue with the study if they do not complete the task in at most two trials.<sup>20</sup>

In total, 500 subjects completed the experiment and passed both comprehension checks. They received a €3 show-up fee for completing the study. The median time to complete the study

---

<sup>18</sup>Motive-group subjects are weakly more likely to receive an incentivizing signal ( $p = 0.09$ , Fisher's exact test). Note, that this does *not* explain the difference across treatment groups as shown in Section 5.3, where I discuss heterogeneity.

<sup>19</sup>One interesting implication of my findings is that subjects may overestimate the likelihood that a high-stakes part of the experiment is actually implemented.

<sup>20</sup>Subjects who did not pass the comprehension test were paid a €0.50 show-up fee to reimburse them for the (short) time they spent with the initial instructions. To avoid selection effects, they were informed about this *after* failing the comprehension test twice. Overall, only 37 subjects out of 550 who attempted the comprehension test failed to answer all questions correctly after the second try.

was 30 minutes (as announced in the initial description of the study). The median earning in the experiment was €5.50. The experiment was implemented between September 18 and September 20, 2020. Twenty subjects were reinvited to perform the same task under a piece-rate scheme on October 4.

Table B2 in the Appendix presents subjects' individual characteristics. Subjects were on average 33 years old. 40% of subjects are female, 24% are students, and 43% work full time. Table B2 presents a balancing table across the motive- and the no-motive-group. Covariates are, overall, well-balanced across the treatment groups. Motive-group subjects are, however, slightly more risk seeking ( $p = 0.062$ , Wilcoxon rank-sum test). Furthermore, Table B3 in the Appendix presents results for balancing tests by incentivizing and disincentivizing signal. All covariates are balanced on this dimension. Table B4 in the Appendix presents results for balancing tests by winning or not winning the bonus. Respondents who won are slightly older ( $p = 0.1002$ , Wilcoxon rank-sum test;  $p = 0.083$ , t-test). This difference becomes insignificant, however, once I control for performance on task 1 ( $p = 0.739$ ). Furthermore, female subjects were more likely to win the bonus ( $p = 0.068$ , Fisher's exact test;  $p = 0.056$ , t-test). This correlation is likely to be spurious as the difference is nearly the same after conditioning on task 1 performance ( $p = 0.068$ ). That the latter correlation is likely spurious is further supported when comparing rank by gender as shown in Table B5 in the Appendix. Female subjects are not more likely to be ranked first ( $p = 0.628$ , Fisher's exact test;  $p = 0.648$ , t-test) and this correlation further weakens after conditioning on task 1 performance ( $p = 0.885$ , Fisher's exact test). Subjects who rank first are, however, significantly younger ( $p < 0.001$ , Wilcoxon rank-sum test). This correlation remains significant at the 10 percent level if I control for task 1 performance and is likely to be driven by a subset of younger subjects who outperformed the competitor sample (and hence were always ranked first). The other covariates are all balanced across ranks. Throughout the text, I will refer to regressions that control for observed heterogeneity, including risk aversion, age, and gender. I refer to the corresponding tables that are found in Appendix 3.B throughout the analysis. I additionally replicate tables that answer the central research questions for a sample that excludes people who did not answer a simple question about the instructions at the *end* of the experiment.<sup>21</sup> I refer to these tables in the table notes.

Finally, I additionally recruited 250 subjects to serve as workers for the redistribution task. These workers were recruited *before* the main experiment was implemented. Workers were paid a €1.00 show-up fee. They were only hired to do the 5-minute transcription task (Work-Phase 1) and were instructed that the initial distribution of the bonus might be redistributed by a third party. Workers received the implemented bonus after the completion of all sessions.

---

<sup>21</sup>The question asks: *The likelihood of receiving the bonus in Part 1 of the survey was independent of your performance if you were in the (a) Performance-Environment, (b) Chance-Environment, (c) None of the above.* 81 percent answered correct. The share of subjects that answered correct is not significantly different across treatment groups ( $p = 0.409$ , Fisher's exact test).



## 5 Results

In the following analysis, a signal is defined as incentivizing or disincentivizing purely based on its *informational* content. Hence, I say that a subject received an incentivizing signal if she learned that she ranked first and won the bonus or if she ranked second and did not win the bonus. Contrarily, I say that a subject received a disincentivizing signal if she won the bonus but was ranked second or if she did not win the bonus though she was ranked first.

The experimental design and analysis was pre-registered. The pre-analysis plan is provided in Appendix 3.E. Throughout the text, I will address whether and how I diverge from the pre-registered analysis. One general divergence is the use of non-parametric Wilcoxon rank-sum tests instead of t-tests, as the former perform better in smaller samples.

### 5.1 Are signals informative?

Before analyzing differences between the motive- and the no-motive-group, it is worth stepping back and asking to what extent baseline beliefs reflect information contained in the signals.

Figure 3.3: Mean elicited beliefs by signals (no-motive-group subjects only)

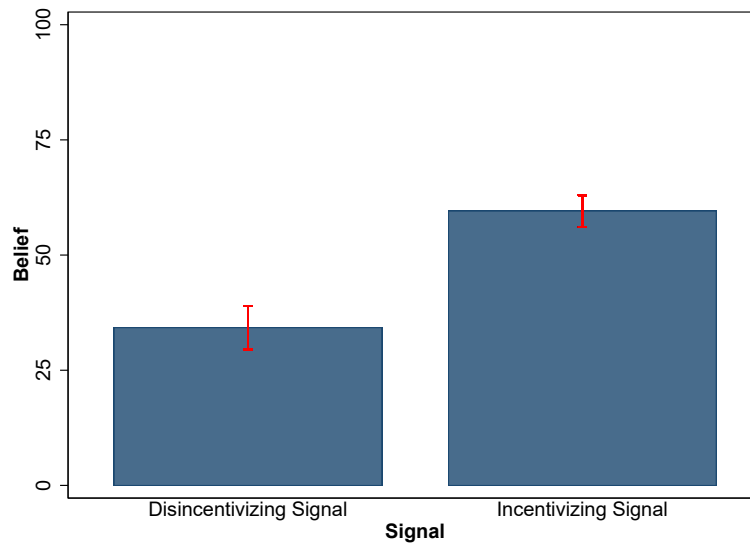
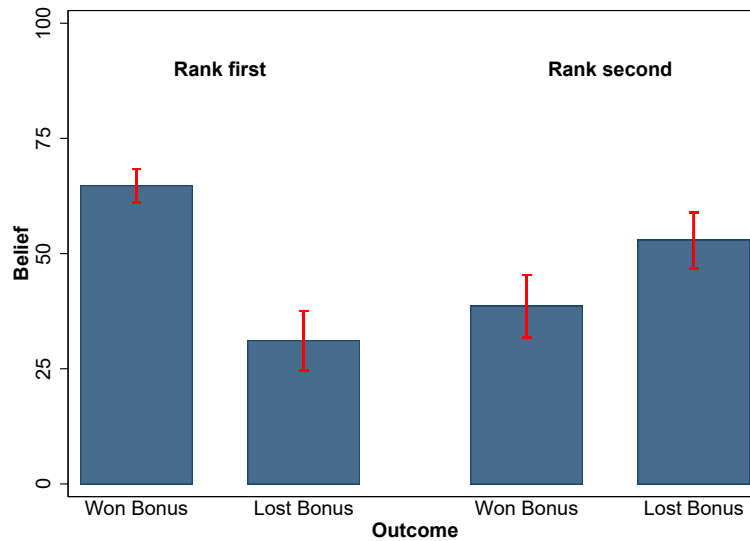


Figure 3.3 shows average beliefs by signal type for no-motive-group subjects. No-motive-group subjects who received an incentivizing signal are 25 percentage points more likely to believe that they are in the Performance-Environment than are the no-motive-group subjects who received a disincentivizing signal ( $p < 0.001$ , Wilcoxon rank-sum test). No-motive-group subjects who received a disincentivizing signal believe, on average, that they are in the Performance-Environment with 34 percent probability and those who received an incentivizing signal believe themselves to be in the Performance-Environment with 59 percent probability, on average. Both groups are, on average, slightly more conservative than the Bayesian prediction (0.62 and 0.29). Though the distribution of beliefs is relatively noisy, the difference is supported by looking at the full distribution of beliefs. Beliefs of subjects who received a disincentivizing signal are shifted

to the left of 50%, while those of the subjects who received an incentivizing signal are shifted to the right (see Figure A1 in the Appendix).

Figure 3.4: Mean beliefs by winning and losing the bonus and rank (no-motive-group subjects)



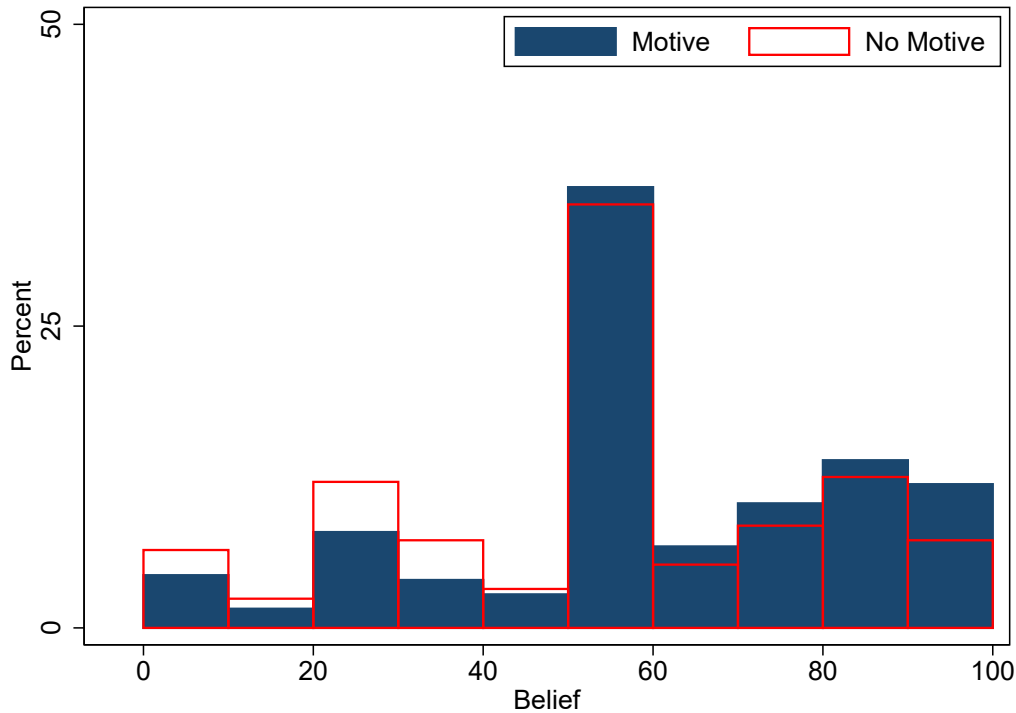
This difference in beliefs remains significant and meaningful when comparing beliefs across subjects who received a different signal and stratifying by whether the subjects ranked first or second, as shown in Figure 3.4. Subjects who did not win the bonus and ranked second are 20 percentage points more optimistic about being in the Performance-Environment than are subjects who did not win the bonus but ranked first ( $p < 0.001$ , Wilcoxon rank-sum test). Similarly, I find that subjects who won the bonus and ranked first believe with 26 percentage points greater probability that they are in the Performance-Environment than do subjects who won the bonus but ranked second ( $p < 0.001$ , Wilcoxon rank-sum test). Hence, the information subjects extract from the signal is similar in both cases, but the difference seems starker if the subject ranks first in terms of performance. I will return to the effect of winning the bonus later in the chapter when I discuss how it interacts with my main treatment variation. Overall, the evidence shows that subjects do, on average, take information provided by the signal into account. Table B6 replicates the results presented in this section controlling for observed heterogeneity.

## 5.2 Beliefs across motive- and no-motive-group

Figure 3.5 shows the distribution of beliefs by treatment group for the whole sample. The blue bars show the distribution of motive-group subjects' beliefs while the transparent bars with the red frame show the distribution of no-motive-group subjects' beliefs.

The plot shows that motive-group subjects believe, on average, that it is more likely that they are in the Performance-Environment than in the Chance-Environment compared to no-motive-group subjects. The distribution is slightly shifted to the right and this shift is reflected in average beliefs. Motive-group subjects believe, on average, that they have a 56% probability of being in the Performance-Environment, while for no-motive-group subjects, on average, the

Figure 3.5: Distribution of beliefs by motive- and no-motive-group subjects (full sample)



probability is 49%. Hence, motive-group subjects are 7 percentage points more optimistic about being in the Performance-Environment than are no-motive-group subjects ( $p = 0.004$ , Wilcoxon rank-sum test).<sup>22</sup>

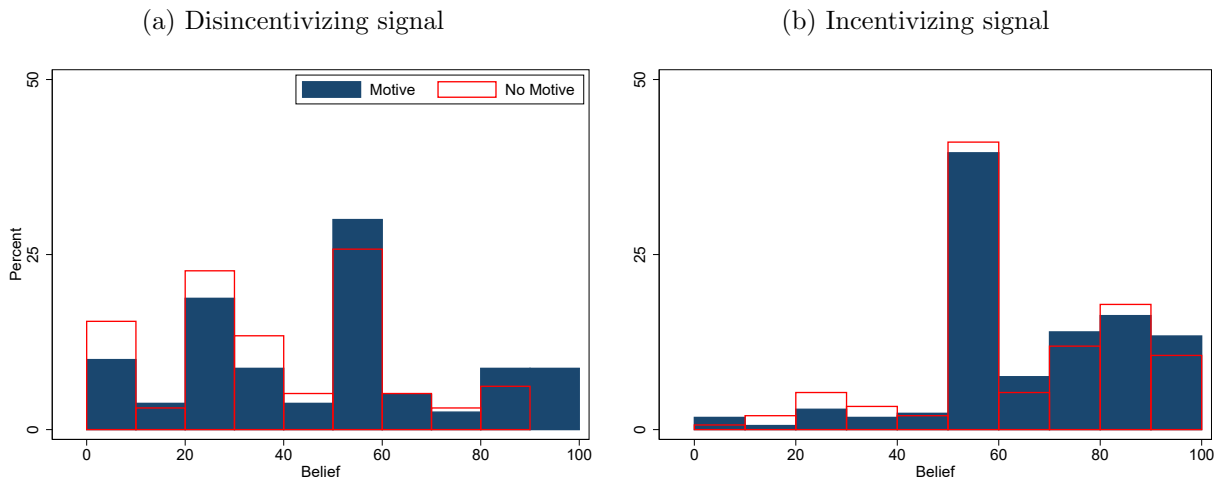
As shown in my framework, I expect motivating belief distortion to be stronger if the subject receives disincentivizing rather than incentivizing information. To test for this hypothesis, I compare beliefs across treatment groups, stratifying by the information subjects received.

Figure 3.6 plots the distribution of beliefs across treatment groups by subjects who received either a disincentivizing—sub-figure 6(a)—or an incentivizing signal—sub-figure 6(b). We start by comparing beliefs across subjects who observed an incentivizing signal. Sub-figure (b) shows that the differences across these groups is rather small and there is no systematic shift in the distribution. Overall, the difference in mean beliefs among these subjects is small. Motive-group subjects are 3 percentage points more optimistic about being in the Performance-Environment and I cannot reject the null-hypothesis that beliefs are equal across the two groups ( $p = 0.28$ , Wilcoxon rank-sum test).<sup>23</sup> Overall, this suggests that there is no systematic difference in updating behavior across treatment groups among subjects who received an incentivizing signal.

<sup>22</sup>Interestingly, average beliefs of no-motive-group subjects are not significantly different from 50% ( $p = 0.8220$ , t-test, c.i. = [46.5, 52.8]). Hence, there is no evidence that subjects have a general tendency to believe that they have control over their outcome because they derive comfort in believing that they have control over their own income, as postulated by theories that argue for affective motives in a just world (e.g. Lerner, 1980).

<sup>23</sup>This is further supported by comparing the share of subjects who (falsely) updated negatively after receiving an incentivizing signal. Even though the share of subjects who updated negatively is slightly higher (4 percentage points), I cannot reject the null-hypothesis that motive-group subjects are less likely to update negatively than no-motive-group subjects ( $p = 0.172$ , Fisher's exact test, one-sided) among subjects who received an incentivizing signal.

Figure 3.6: Distribution of beliefs by motive- and no-motive-group subjects across signals



**Note.** These figures plot the distribution of beliefs by treatment group. Sub-figure (a) plots the distribution for subjects who received a disincentivizing signal. Sub-figure (b) plots the distribution for subjects who received an incentivizing signal.

Sub-figure 6 (a) plots the distribution of beliefs by treatment groups among subjects who received a disincentivizing signal. The shift in the distribution indicates that motive-group subjects were less likely to update their beliefs negatively compared to no-motive-group subjects who received a disincentivizing signal. This is supported when comparing the share of subjects who updated negatively across treatment groups; 59 percent of no-motive-group subjects updated negatively, while only 45 percent of motive-group subjects updated negatively. Hence, motive-group-subjects are 14 percentage points less likely to update in the correct direction. This difference allows us to reject the null-hypothesis that motive-group subjects are more likely to update negatively than are no-motive-group subjects ( $p = 0.035$ , Fisher’s exact test, one-sided). This dynamic is also reflected when comparing average beliefs of subjects who received a disincentivizing signal across treatment groups. Among this group of subjects, no-motive-group subjects, on average, believed themselves to be in the Performance-Environment with 34% probability, while motive-group subjects, on average, believed this with 43% probability, leading to a significant 9-percentage-point difference in beliefs across the two groups ( $p = 0.008$ , Wilcoxon rank-sum test).

In sum, my data provide strong evidence that motive-group subjects are significantly more likely to believe that they are in the Performance-Environment than are no-motive-group subjects. Motive-group subjects, on average, believe with higher probability that they are in the Performance-Environment than do no-motive-group subjects and the former’s average belief is significantly higher than 50% ( $p < 0.001$ , t-test). This difference is especially pronounced for subjects who received a disincentivizing signal. Tables B7 and B8 in the Appendix confirm the main results of this section using linear regression and controlling for individual heterogeneity.<sup>24</sup>

<sup>24</sup>The motive-group treatment effect for subjects who received a disincentivizing signal becomes marginally significant ( $p = 0.074$ , two-sided) once I control for observable heterogeneity. Note, this is driven by limited power when regressing the treatment separately for subjects who received a disincentivizing and incentivizing signal. As shown in Table B7, the motive-group effect remains unchanged (both in magnitude and significance) if I control for observable heterogeneity using the complete sample.

Given that, by design, 65% of subjects received an incentivizing signal, it is remarkable to find a significant average effect for all subjects. Overall, motivating belief distortion may be even more important for average beliefs, if individuals hold false beliefs about their relative ability. Indeed, overconfident beliefs about one’s relative ability will generate disincentivizing signals about the true importance of effort if subjects refuse to update about their own relative ability.

**Updating relative to a Bayesian benchmark** Comparing the belief distortion across subjects who received an incentivizing or disincentivizing signal is complicated because disincentivizing signals are more informative about the true state of the world than are incentivizing signals. I circumvent this issue by characterizing motivating belief distortion relative to a Bayesian benchmark. To that end, I follow the experimental literature on motivated belief updating by calibrating a quasi-Bayesian updating regression (see e.g. Möbius et al., 2014; Coutts, 2019a; Barron, 2021). The outcome variable of the regression is the elicited logit belief ( $\tilde{\pi}_i$ ) at the subject level and the explanatory variable is the log-odds ratio of the signal a subject receives.<sup>25</sup> If a subject is Bayesian, the coefficient of the log-odds ratio would be equal to 1, indicating that her posterior does not differ from the posterior of a Bayesian updater. If the coefficient is larger than 1, she is updating more strongly in response to the signal than she would if she were a Bayesian updater. The latter is an example of a subject who over-responds to the signal. If the coefficient is smaller than 1, she is conservative and does not take all the information from the signal into account when updating beliefs. The coefficient of the prior indicates whether the subject’s belief updating is characterized by base-rate neglect or confirmatory bias. My design cannot inform about the latter biases in belief updating because prior beliefs are equal to 0.5 for all subjects. Hence, logit prior beliefs are equal to 0 and the coefficient cannot be identified.<sup>26</sup>

To characterize updating behavior relative to the Bayesian benchmark across incentivizing and disincentivizing signals ( $q$ ), as well as across motive-treatment groups, I interact the log-odds ratio ( $\tilde{q}$ ) with a dummy variable that indicates whether the subject received an incentivizing or disincentivizing signal. This results in two explanatory variables that characterize updating behavior relative to a Bayesian benchmark for subjects who received an incentivizing or disincentivizing signal separately (indicated by the subscript + and –). Furthermore, I add an additional interaction term, where I interact  $\tilde{q}_{+/-}$  with a motive-group (M) indicator. The coefficient of the resulting variable informs whether motive-group subjects are more or less likely to respond to the signal than are no-motive-group subjects. (3.9) characterizes the resulting model that can be estimated using OLS:

$$\tilde{\pi}_i = \alpha_+ \tilde{q}_+ + \beta_+ \tilde{q}_+ M + \alpha_- \tilde{q}_- + \beta_- \tilde{q}_- M \tag{3.9}$$

In total, the regression yields four parameters:  $\alpha_+$  and  $\alpha_-$  inform how no-motive-group subjects

---

<sup>25</sup>Note that logit-beliefs converge to (minus) infinity for subjects who believe that they are in the Performance-Environment with probability 1 (0). To include these subjects in my regressions, I assume that they hold probabilistic beliefs that are equal to .99 (.01).

<sup>26</sup>Note that the aim of the chapter is to study differences in belief updating due to the motive treatment rather than to provide a comprehensive characterization of belief updating. To that end, one should elicit beliefs after multiple signals.

update relative to a Bayesian benchmark if they receive an incentivizing and disincentivizing signal.  $\beta_+$  and  $\beta_-$  inform about how motive-group subjects respond to an incentivizing (disincentivizing) signal compared to the response of no-motive-group subjects who received the same signal.

Table 3.3: Results of the updating regression specified in (3.9)

Coefficient	Estimate
$\alpha_+$	1.197 (0.244)
$\beta_+$	0.424 (0.361)
$\alpha_-$	1.176 (0.193)
$\beta_-$	-0.787** (0.322)
<hr/>	
$H_0 : \alpha_+ = \alpha_- = 1$	$F(2, 496) = 0.74$
$H_0 : \beta_+ = \beta_-$	$F(1, 496) = 6.27$
<hr/>	
N	500
$R^2$	0.16

**Note.** This table shows coefficient estimates for the regression characterized in (3.9).  $\alpha$  coefficients test the null-hypothesis that  $\alpha = 1$  while  $\beta$  coefficients test the null-hypothesis that  $\beta = 0$ . Table B21 in the Appendix replicates this table excluding subjects that did not pass a basic comprehension test at the end of the experiment. Robust standard errors are in parentheses. *Significance levels:* \*10%, \*\*5%, \*\*\*1%.

Table 3.3 presents the results of this regression. The log-odds ratios indicate that no-motive-group subjects do not significantly depart from Bayesian posteriors. The null-hypothesis that the coefficient is equal to 1 cannot be rejected for no-motive-group subjects who received an incentivizing ( $p = 0.420$ , F-test) or disincentivizing signal ( $p = 0.361$ , F-test).<sup>27</sup>

The interaction terms with the treatment dummy variable show that motive-group subjects who receive an incentivizing signal react more strongly to the signal than do no-motive-group subjects who received an incentivizing signal. The coefficient is, however, not significantly different from zero ( $p = 0.240$ ). This changes when characterizing updating behavior by subjects who received a disincentivizing signal. They react less strongly ( $p = 0.015$ ) to the signal than do no-motive-group subjects who received the same signal. Combining the main effect with the interaction term results in a posterior logit belief of -0.38, which is significantly smaller than the posterior they would hold based on Bayesian beliefs ( $p = 0.018$ , F-test) and it is not significantly

<sup>27</sup>This does not mean that no-motive-group subjects update similarly to Bayesian agents on the individual level, but the results reflect average updating behavior.

different from 0 ( $p = 0.132$ , F-test). The latter means that I cannot reject the null-hypothesis that subjects do not update at all.

The next step is to compare the two interaction terms with each other. I reject the null-hypothesis that the two terms are equal ( $p = 0.013$ , F-test). This reveals significant differences in how treated subjects react to information after receiving an incentivizing or disincentivizing signal. The negative sign of  $\beta_-$  indicates that motive-group subjects *under-react* to the information they received if they received a disincentivizing signal. This is not the case for motive-group subjects who received an incentivizing signal. There, the positive, but insignificant,  $\beta_+$  indicates that subjects are not updating differently than the baseline; if anything, these subjects are over-inferring from the information. This dynamic is predicted by models of motivated belief distortion: i.e., subjects who receive information that goes *against* the motivated belief under-react to the information they receive, while those who receive information congruent the motivated belief are expected to weakly over-react to this information Bénabou (2015).<sup>28</sup>

### 5.3 Who distorts luck-effort beliefs to motivate effort?

The previous section established that subjects do distort beliefs to motivate future effort and that motivating belief distortion is particularly pronounced if subjects receive a disincentivizing rather than an incentivizing signal. First, I analyze whether motivating belief distortion interacts with the event observed by the subjects and then I characterize what type of subjects is more prone to distort luck-effort beliefs to motivate effort.<sup>29</sup>

**The effect of winning the bonus and performance rank on beliefs** To analyze how winning and losing affects beliefs, I plot average beliefs of subjects who won and lost the bonus for no-motive-group subjects in Figure 3.7.

Figure 3.7 shows a clear and significant difference ( $p < 0.001$ , Wilcoxon rank-sum test) in no-motive-group subjects' beliefs between those who won the bonus and those who lost it. This difference is equal to 13 percentage points and is somewhat less than half as large as the effect of receiving an incentivizing signal.<sup>30</sup> Overall, this indicates that experiencing a win induces a general tendency to believe that one's effort is likely to be rewarded.

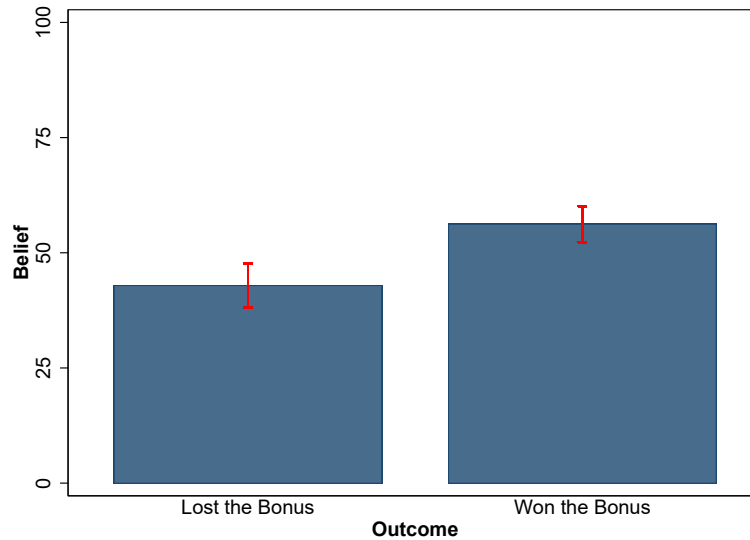
The next step is to ask whether the effect of winning the bonus on beliefs interacts in a meaningful way with the performance rank subjects hold. Figure 3.4, introduced in Section 5.1,

<sup>28</sup>Note that I can not reject the null hypothesis of no difference in the absolute *magnitude* of the interaction terms ( $p = 0.453$ , F-test). This means that the difference in updating is driven by a difference in direction rather than through a difference in magnitude. The reader should, however, take this null result with a grain of salt given the large standard errors and given that the magnitude of the parameters indicates that the magnitude of the distortion is larger for those subjects who received a disincentivizing signal.

<sup>29</sup>In the pre-analysis plan, I listed the first point under exploratory analysis. I added this part to the section on who engages in motivating belief distortion because it informs whether subjects who win or lose are more likely to engage in motivating belief distortion.

<sup>30</sup>Note that these average differences do not reflect the fact that winning the bonus is more informative than losing the bonus. While, for no-motive-group subjects, there is a weak correlation (0.13) between receiving an incentivizing signal and winning the bonus, this does not explain such a large difference in beliefs. This is also confirmed in Table B7 in the appendix, which reports results from linear regressions that explicitly control for the informational content of the signal. These regressions also control for task 1 performance. The coefficient of winning the bonus remains nearly unchanged, indicating that the effect from winning the bonus is unlikely to reflect unobserved heterogeneity that is correlated with ability.

Figure 3.7: Mean beliefs by winning and losing the bonus (no-motive-group subjects)



plots no-motive-group subjects' mean beliefs by winning and losing as well as rank. Though the informational content of the signal is reflected in the belief differences among those ranked second, the shift in beliefs due to the information is smaller. This can reflect subjects either over-responding to an incentivizing signal when being ranked first or under-responding to the same information when being ranked second. Though I do not have a suitable control group to definitely answer this question, I can say that subjects who ranked first and won or lost the bonus are relatively close to the Bayesian benchmark compared to those who ranked second and won or lost the bonus. The latter's beliefs are more condensed and the average beliefs of those who lost the bonus and rank second are at 55.7%, which is only slightly larger than 50% ( $p = 0.005$ , t-test). Hence, at the baseline, the main effect of winning or losing the bonus is likely to be driven by no-motive-group subjects under-responding to the information if they ranked second rather than first.

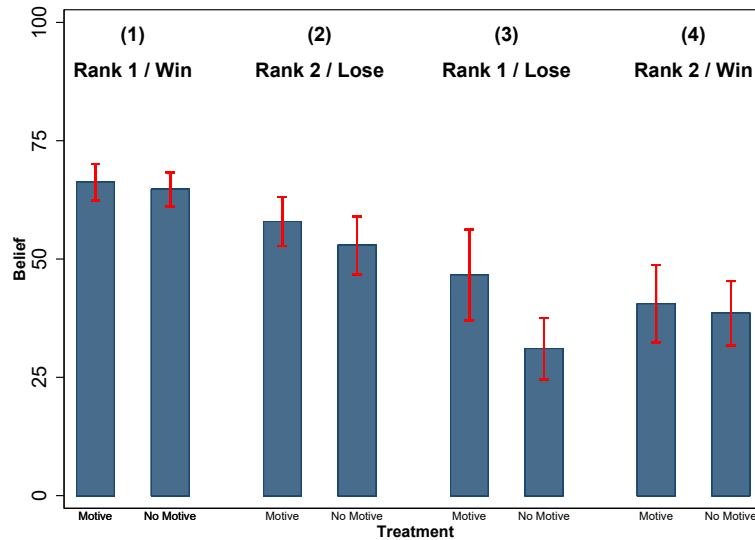
I now ask to what extent this effect interacts with being a motive-group subject. Figure 3.8 plots mean beliefs for each treatment group by every possible outcome of the task. Bars (1) and (2) feature outcomes where the informational value is incentivizing, while bars (3) and (4) feature outcomes where the motivational value is disincentivizing.

The figure clearly shows that the treatment effect is mainly driven by subjects who ranked first and lost the bonus, while there is no significant difference across treatment groups for subjects who ranked second but won the bonus. Among subjects included in bars (3), motive-group subjects are 15 percentage points more confident of being in the Performance-Environment compared to no-motive-group subjects ( $p = 0.014$ , Wilcoxon rank-sum test). These results were confirmed using linear regressions that control for observable heterogeneity and the informational content of the signal as presented in Table B9 in the Appendix.

This result provides evidence indicative of what type of event triggers motivating belief distortion. While I showed above that losing the bonus has a negative effect on beliefs at the baseline, this is not the case for motive-group subjects who received a disincentivizing signal.



Figure 3.8: Mean beliefs by winning and losing the bonus, rank, and motive treatment



This leads to the question of what type of individuals frequently observe such events outside the laboratory. Generally speaking, these are people who believe that, in a world that rewards effort, they should have won, despite the fact that they do not succeed. This type of event occurs if individuals think their productivity in a given task is above average. My results imply that, at least in the aggregate, these individuals, when facing such an event, do not update their beliefs about the state of the world to increase the role of luck. Furthermore, this is precisely the type of event that individuals face if they are persistently overconfident in their relative skill and ability. Such individuals would make relatively correct inferences about the true state of the world, holding their belief about their relative ability constant, if they could use beliefs as a motivating device. However, if they can motivate future effort by distorting their beliefs, they become significantly more confident that they are in a world that rewards effort. This equally implies that an attribution bias in luck-effort beliefs (e.g. Deffains et al., 2016) may be crowded out if beliefs can be distorted to motivate future effort.

**Individual characteristics and motivating belief distortion** While the above section showed how the treatment effect interacts with the type of signal, one can also ask whether certain types of individuals are more or less likely to engage in motivating belief distortion. The theoretical literature on the topic argues that individuals who are present-biased distort beliefs to overcome limited willpower (see Bénabou, 2015). Furthermore, subjects who are not very good at the task should be more likely to engage in motivating belief distortion because they need the extra motivation to perform better. Taking this as a starting point, I pre-specified to analyze whether my treatment effect interacts with work-phase-1 performance, demand to commit to the exertion of effort, and self-reported enjoyment of the task.

Table 3.4 shows the results from regressions that interact the motive-group dummy variable with the performance decile a subject occupied. Column (1) performs this regression for the whole sample while columns (2) and (3) split the sample by subjects receiving an incentivizing

Table 3.4: Regressions interacting the treatment variable with performance decile in Task 1

	(1)	(2)	(3)
	Belief	Belief	Belief
Motive group	13.45*** (4.975)	42.50*** (12.68)	2.351 (5.365)
Bonus win	12.52*** (3.239)	9.391* (5.142)	10.11** (4.563)
Task 1 perf. decile	0.831 (0.578)	0.756 (1.000)	0.543 (0.799)
Motive * bonus win	-8.675* (4.664)	-26.68*** (9.153)	-6.668 (5.854)
Motive * task 1 perf. decile	-0.412 (0.830)	-3.905** (1.643)	0.934 (1.019)
Constant	38.82*** (3.717)	26.36*** (7.090)	50.86*** (4.136)
Signal	All	Disincentivizing	Incentivizing
Observations	500	177	323

**Note.** The above table reports OLS regressions that interact the motive treatment variable with winning the bonus and performance decile in task 1. Column (1) reports results for the whole sample; column (2) reports results for a sample that only includes subjects who received a disincentivizing signal and column (3) reports results for a sample that only includes subjects who received an incentivizing signal. See Table B10 for results of a regression that includes the full set of results. Standard errors are robust to heteroskedasticity. *Significance levels:* \*10%, \*\*5%, \*\*\*1%.

or disincentivizing signal. All of the regressions control for the interaction effect of winning the bonus and being in the motive group. This control variable is crucial because the two effects may cancel each other out given the positive correlation between winning the bonus and performing well in the task that I discussed in the previous section. Furthermore, I showed above that the treatment effect is particularly pronounced for subjects who lost the bonus and ranked first. Hence, one must also control for the interaction effect of winning the bonus and being in the motive-group to identify an interaction effect for task 1 performance and the motive-treatment. For specifications that control for individual heterogeneity, see Table B10 in the Appendix.

The results are in line with the hypothesis that subjects who do not perform well on the task are more likely to engage in motivating belief distortion ( $p = 0.019$ ) if they received a discouraging signal. This is consistent with the predictions from my framework, where I showed that individuals with a relatively high cost of effort are expected to distort beliefs, as they demand extra motivation to attain the threshold. Subjects who can attain the threshold with ease do not necessarily need to distort beliefs to motivate their effort and, hence, are less willing to pay the cost associated with belief distortion.<sup>31</sup>

<sup>31</sup>This result goes in some sense against the finding by Banerjee et al. (2020) who study confidence spillovers across tasks. Their heterogeneity analysis does not suggest that belief distortion is more prevalent for those who should exhibit a higher demand for motivating belief distortion. I emphasize that I view these results as complementary rather than contradictory. Importantly, the setting they study is very different than mine because

The second characteristic I examine is whether subjects who make a revealed choice to commit to effort in a future transcription task are more likely to respond to the treatment. The idea is that these are people who have a general tendency to use a commitment device to motivate future effort. As described in Section 3.7, I asked subjects to state how many images they planned to transcribe if they were reinvited two weeks hence to redo the same task under piece-rate incentives. The subjects subsequently were asked to state a minimum production level. If they transcribe fewer images than their minimum production level, they receive no pay; if they transcribe more images, they will be paid according to the piece rate. I pre-specified that a subject demands a commitment device if her minimum production level is at least half as large as the number of images she plans on transcribing. Using this binary measure yields enough variation to identify a potential heterogeneity in the treatment effect as exactly 50% of the subjects chose to commit to exert effort in the future by this measure. Furthermore, the commitment measure is uncorrelated with task 1 performance ( $\rho = 0.06$ ,  $p = 0.901$ ) and weakly correlated with having won the bonus after the first part ( $\rho = 0.07$ ,  $p = 0.107$ ).

Table 3.5: Regressions interacting the treatment variable with demanding a commitment device

	(1)	(2)	(3)
	Belief	Belief	Belief
Motive group	11.73*** (3.249)	15.10*** (5.528)	5.254 (3.454)
Demands commitment	7.160** (3.186)	9.578** (4.641)	2.775 (3.496)
Motive * demands commitment	-10.37** (4.503)	-12.58 (7.813)	-5.038 (4.792)
Constant	46.29*** (2.278)	30.18*** (3.206)	58.16*** (2.403)
Signal	All	Disincentivizing	Incentivizing
Observations	500	177	323

**Note.** The above table reports OLS regressions that interact the motive treatment variable with a dummy variable that indicates the demand for a commitment device. Column (1) reports results for the whole sample; column (2) reports results for a sample that only includes subjects who received a disincentivizing signal and column (3) reports results for a sample that only includes subjects who received an incentivizing signal. See Table B11 for results of a regression that includes the full set of results. Standard errors are robust to heteroskedasticity. *Significance levels:* \*10%, \*\*5%, \*\*\*1%.

Table 3.5 presents the results of the regression and Table B11 in the Appendix replicates these regressions, controlling for observable characteristics. As before, I run the regression for the whole sample and I additionally split the sample by signal type. The results indicate that subjects who demand a commitment device are, generally speaking, more likely to believe that their effort is rewarded. The coefficients for the interaction term indicate, if anything, that subjects who demand a commitment device are *less* likely to engage in motivating belief distortion than are those who do not commit themselves to exert effort on the additional task.

their belief is ego-relevant (relative performance) and anticipatory utility motives—which I control for in my setting—may be relevant in their setting.

This goes against the pre-specified hypothesis that subjects who have a revealed preference for commitment engage in motivating belief distortion.

Finally, I test whether subjects who reported enjoying the task are less likely to engage in motivating belief distortion. The idea is that these subjects have a higher cost of effort and have low intrinsic motivation to undertake the task. Hence, it is precisely they who should inflate extrinsic incentives to motivate future effort. My findings indicate that the latter mechanism may be more relevant in this setting.

Table 3.6: Regressions interacting the treatment variable with enjoying the task

	(1)	(2)	(3)
	Belief	Belief	Belief
Motive group	5.779 (5.715)	-2.007 (10.85)	4.660 (6.317)
Bonus win	13.31*** (3.239)	6.421 (5.095)	8.412** (4.075)
Enjoys task	-4.033 (4.005)	-5.876 (6.856)	0.889 (4.603)
Motive * bonus win	-10.39** (4.502)	-16.18** (7.781)	-4.063 (4.786)
Motive * enjoys task	7.922 (5.907)	22.29** (11.07)	1.358 (6.208)
Task 1 perf. decile	0.628 (0.419)	-0.720 (0.809)	1.033** (0.507)
Constant	42.70*** (4.274)	40.07*** (7.704)	48.43*** (4.968)
Signal	All	Disincentivizing	Incentivizing
Observations	498	177	321

**Note.** The above table reports OLS regressions that interact the motive treatment variable with a dummy variable that indicates whether the subject won the bonus and whether they subject enjoyed the task. Column (1) reports results for the whole sample; column (2) reports results for a sample that only includes subjects who received an incentivizing signal and column (3) reports results for a sample that only includes subjects who received a disincentivizing signal. See Table B12 for results of a regression that includes the full set of results. Standard errors are robust to heteroskedasticity. *Significance levels:* \*10%, \*\*5%, \*\*\*1%.

Table 3.6 presents the results of regressions that interact the treatment variable with a dichotomized variable indicating whether the individual enjoyed the task or not. Table B12 in the Appendix replicates this table controlling for observable heterogeneity.<sup>32</sup> The results are, if anything, contrary to my expectation. Subjects who enjoyed the task are more likely to engage in motivating belief distortion, as indicated by the positive and significant interaction term for subjects who received a disincentivizing signal. This result may be attributable to the fact that

<sup>32</sup>I additionally control for winning the bonus and its interaction because one may expect that subjects who won the bonus also feel more satisfied with undertaking the activity, which may cancel out an interaction effect with enjoying the task.

subjects who distort beliefs to motivate future effort are also prone to telling themselves that they like the task for the same reason. This, however, needs further study in greater detail in future research.

What do these results imply about the relevant mechanism behind belief distortion? The first finding outlined in this section is consistent with the framework proposed above, because one would expect that those for whom the task is hard are more inclined to pay the mental cost of belief distortion because their expected returns from it are higher. This finding is consistent with a mechanism where beliefs are used as a commitment device because it helps those subjects who need the extra motivation to get over the threshold that need the extra motivation. The result regarding the commitment device is, however, less consistent with the idea that subjects use beliefs as a commitment device because one would expect that motivating belief distorters would use similar instruments to commit themselves to exert effort. Hence, it is more likely that an alternative mechanism is relevant in this context. One alternative mechanism that was outlined in Section 2 is that subjects have an aversion to engage in an activity when they believe that the activity is unnecessary or does not serve any purpose. To avoid this extra cost of effort, individuals may distort information that indicates that their effort is not needed to obtain the reward. My results suggest that this mechanism may be more relevant in this setting than the idea that belief distortion is driven by present-biased preferences, but more research is warranted to reach a conclusive answer to this question.

**Memory** To test whether the difference in beliefs is explained by motive-group subjects being more likely to forget or recode the signal they previously received, I asked subjects to recall the signal at the beginning of the post-experimental questionnaire. Overall, subjects were relatively good at recalling their signal, as 86% of subjects correctly recalled the signal. There is no significant difference across treatment groups in the propensity to recall the correct signal ( $p = 0.365$ , Fisher's exact test). Even though subjects were 15 percentage points more likely to not recall the correct signal if they received a disincentivizing signal ( $p < 0.001$ , Fisher's exact test), I cannot reject the null hypothesis that motive-group-subjects are equally likely to recall the correct signal conditional on having observed a disincentivizing signal ( $p = 0.375$ , Fisher's exact test) or incentivizing signal ( $p = 0.843$ , Fisher's exact test). These results are confirmed in Table B13 using linear regressions and controlling for observable heterogeneity.

Furthermore, I can ask whether motive-group subjects are more likely to recall the reverse signal. Among those who said that they recalled a signal, as opposed to saying that they did not recall any signal, I can construct the signal they claimed to recall and construct a variable which I call "perceived signal." Recalling an incorrect signal is relatively rare and only occurred for 22 subjects (5% of all subjects). Unsurprisingly, I cannot detect any group differences in the propensity to recall a false signal ( $p = 0.196$ , Fisher's exact test). This holds equally for subjects who received an incentivizing signal ( $p = 0.35$ , Fisher's exact test) or a disincentivizing signal ( $p = 0.249$ , Fisher's exact test). One can equally ask whether the motive-treatment effect is mediated by recalling a false signal. To that end, I regress beliefs on the signal they recall (conditional on claiming that they recalled any signal). As shown in Table B14 in the

Appendix, this does not explain the difference across treatment groups and the reverse seems to be true—i.e., the magnitude of the treatment effect becomes larger once we control for the recalled signal. These results indicate that the cognitive mechanism that yields the difference across treatment groups is unlikely to be amnesia but, rather, under-investment in decoding the information contained in the signal.

#### 5.4 Distributive Behavior

The previous section characterized updating behavior across treatment groups. We now ask whether this difference in updating behavior is reflected in the distributive decisions subjects make. Note that this decision does not affect their own payoff and isolates *other*-regarding motives in the decision to distribute money among two agents.

Figure 3.9: Redistribution by motive treatment

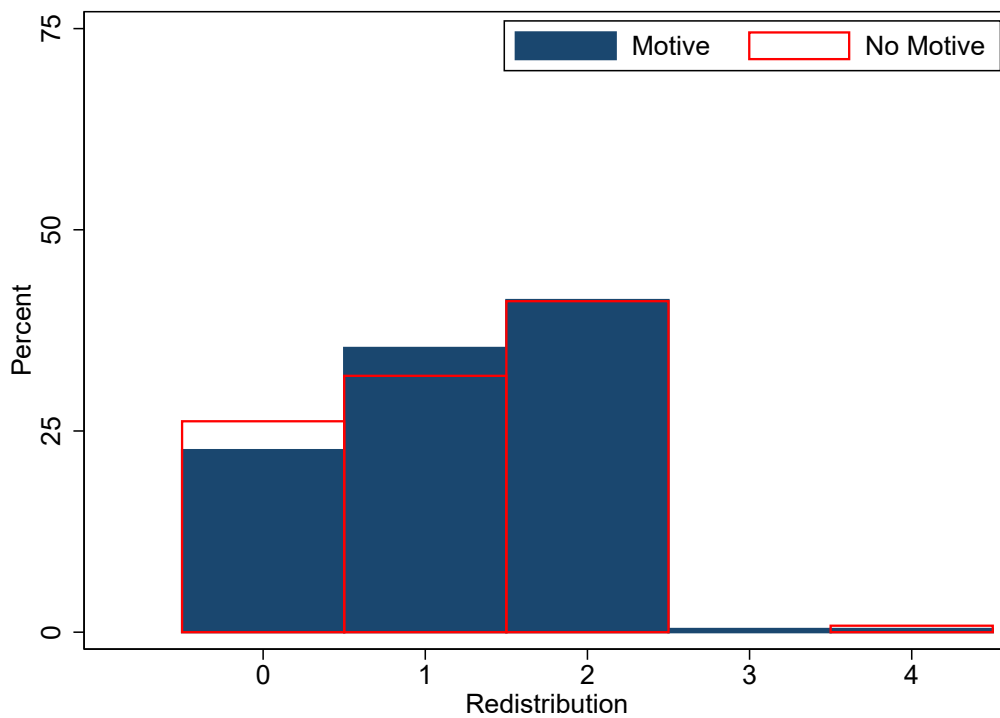
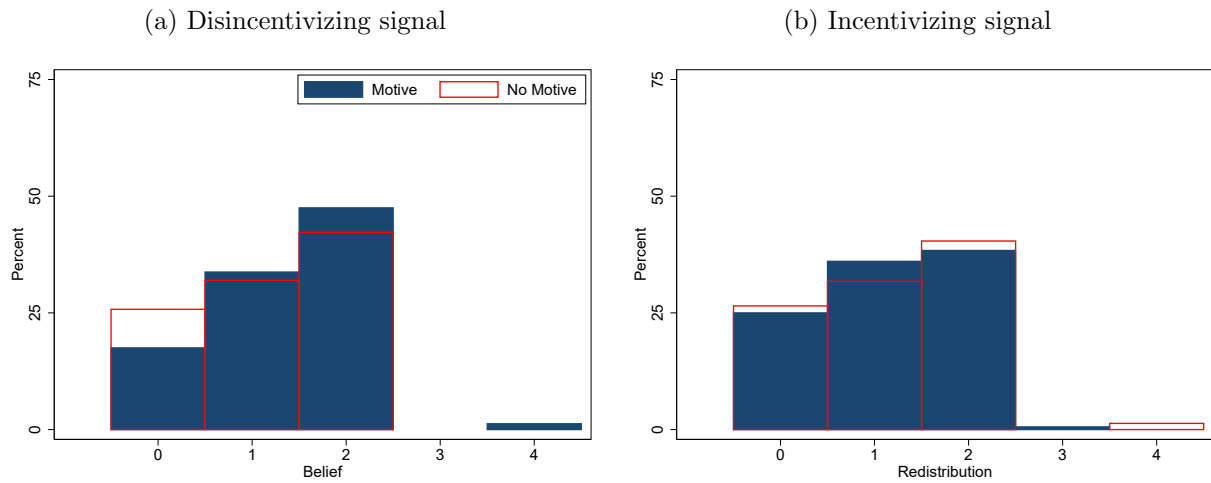


Figure 3.9 plots the distribution of redistributive decisions for motive- and no-motive-group subjects. A subject who redistributes 0 chooses to distribute €4 to the worker who won the initial bonus and to not give any bonus to the other worker. A subject who redistributes 2 chooses an equal split and redistributes half of the bonus to the worker who did not win the initial bonus. The graph shows that there is heterogeneity in distributive behavior. More than half of the subjects (58%) chose not to fully redistribute the initial distribution. 25% chose not to redistribute a positive amount and, instead, chose the initial distribution of income. Giving *more* to Worker B is very rare, as only 4 out of 500 subjects chose to reverse the initial inequality.

I cannot reject the null hypothesis that motive-group subjects redistribute the same amount

of money as no-motive-group subjects ( $p = 0.65$ , Wilcoxon rank-sum test). If anything, the data suggest that motive-group subjects are more likely to redistribute.

Figure 3.10: Redistribution by motive treatment and signal



**Note.** These figures plot the distribution of redistribution decisions by treatment group. Figure (a) plots the distribution for subjects who received a disincentivizing signal. Figure (b) plots the distribution for subjects who received an incentivizing signal.

Figure 3.10 plots the distribution of the redistribution decision for subjects who received an incentivizing or a disincentivizing signal in subfigures (a) and (b), respectively. In both cases, I cannot reject the null-hypothesis that redistributive behavior is equal across the two treatment groups ( $p = 0.21$  (disincentivizing signal),  $p = 0.81$  (incentivizing signal), Wilcoxon rank-sum test). Again, if anything the evidence indicates that motive-group subjects redistribute *more* than no-motive-group subjects, which goes against the hypothesis specified above. These results are confirmed by linear regressions that control for observable heterogeneity as presented in Table B15 in the Appendix.

Table B15 in the Appendix also presents regressions that interact the motive-treatment dummy variable with a dichotomized variable that indicates whether the subject self-identifies as a meritocrat. I do not find any evidence that belief distortion is stronger for subjects who self-identify as a meritocrat. This is true for subjects who received a disincentivizing or an incentivizing signal.

Overall, I do not find evidence that distributive behavior differs significantly across motive- and no-motive-groups. The induced distortion in beliefs through the motive treatment may not be sufficiently strong to provoke changes in distribution behavior. Most of the literature has studied distributive behavior and the source of inequality has varied the likelihood that subjects are indeed rewarded for their effort on the *extensive* margin.<sup>33</sup> Hence, an average difference of 7 percentage points may be too small to translate into average differences in inequality acceptance.<sup>34</sup> Further, Cappelen et al. (2019) show that the relationship between beliefs on the

<sup>33</sup>Most experiments that study how the source of inequality affects preferences regarding redistribution compare treatments where there is no uncertainty in how the initial allocation of income was distributed.

<sup>34</sup>Results from a pilot-experiment ( $N=62$ ) show that individuals are indeed concerned about merit, if there is certainty about the true state of the world. In that design, subjects were informed that Worker A was the

intensive margin and distributive behavior is convex in the probability that effort is rewarded. As mentioned above, belief distortion is particularly relevant for those who received a disincentivizing signal. Average belief probabilities (of being in the Performance-Environment) across the two treatment groups are 34% and 43%. While this variation is large in magnitude, it may not be around the point that is locally relevant to trigger differences in redistributive behavior. Future research should explore further whether this is actually the case. One could, for example, replicate the experiment but vary the prior that subjects hold. If inequality-acceptance and beliefs follow a convex relationship, one could imagine that a treatment may potentially be stronger around a 90% prior.

**Characterizing redistributive behavior** Though the motive-treatment does not directly affect distributive behavior, I do find that beliefs are significantly correlated with distributive behavior ( $p = 0.041$ ), as shown in Table B16 in the Appendix. This dynamic is largely driven by distributive behavior at the extensive margin: Beliefs are uncorrelated with implementing complete equality ( $p = 0.445$ ), but subjects who believe with greater certainty that they are in a Performance-Environment are significantly more likely to not redistribute at all ( $p = 0.005$ ).<sup>35</sup>

One can also ask how distributive behavior varies across events. Figure 3.11 plots average amounts redistributed by performance rank and prize received. Bars (1) and (2) show average amounts redistributed for subjects who received an incentivizing signal and bars (3) and (4) show average amounts redistributed for subjects who received a disincentivizing signal. The figure clearly shows that subjects who won the bonus and ranked first redistributed significantly less than all other subjects ( $p < 0.001$ , Wilcoxon rank-sum test). Beyond that, it is apparent that subjects who received a disincentivizing signal redistribute relatively similar amounts no matter whether they won the bonus and ranked second or whether they lost the bonus and ranked first in terms of performance. On the other hand, I find that subjects who ranked second and did not win the bonus redistribute *more* than the other subjects that received an incentivizing signal ( $p < 0.001$ , Wilcoxon rank-sum test) and weakly more than do subjects who received a disincentivizing signal ( $p = 0.081$ , Wilcoxon rank-sum test). Overall, this indicates that subjects who did not win the bonus and ranked second are not making redistributive decisions based on their induced beliefs. If we compare, however, distributive behavior across subjects who ranked

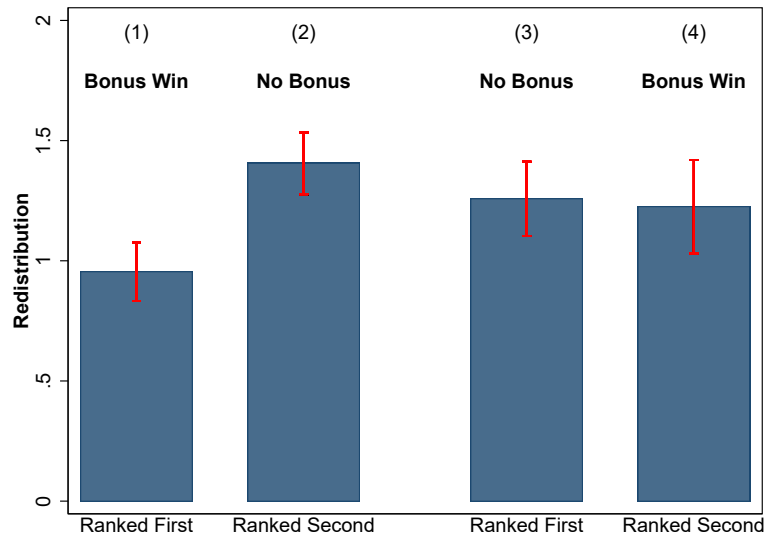
---

better performer with 100% probability if they were in the Performance-Environment and with 50% probability if they were in the Chance-Environment. Hence, there was considerably less uncertainty about who was the better performer conditional on being in the Performance-Environment. Furthermore, it included a strategy-method decision, where subjects made the identical distribution decision *conditional* on being in the Chance- or Performance-Environment that shows that subjects redistribute 40 cents less ( $p < 0.001$ , t-test) if they made the decision conditional on being in the Performance-Environment compared to when they made the same decision conditional on being in the Chance-Environment. Beyond that, the correlation between beliefs and redistribution in the unconditional decision is stronger than in the main treatment ( $\beta = -0.007$ ,  $p = 0.093$ ). Nonetheless, I still fail to find a significant negative relationship between redistribution and being in the motive group. If anything, it is positive (diff.=0.27,  $p = 0.301$ , Wilcoxon rank-sum test). This is not explained by a difference in the treatment effect, which is similar in size (diff. = 4.9,  $p = 0.386$ , Wilcoxon rank-sum test) but insignificant given the small sample.

<sup>35</sup>These correlations should be viewed with caution because I cannot reject the null-hypothesis that beliefs are uncorrelated with distributive behavior if I instrument beliefs by the signal subjects received ( $p = 0.284$ ). This may not be very surprising given that the signal predicts a 20-percentage-point shift in beliefs around 50%. Beyond that, the extensive margin regressions were not pre-registered. I present them because they provide a more comprehensive impression of distributive behavior.



Figure 3.11: Redistribution by event



first, I find that subjects who won the bonus distribute 30 cents less than subjects who did not win the bonus ( $p = 0.003$ , Wilcoxon rank-sum test) and they are 15 percentage points more likely to not redistribute at all ( $p = 0.005$ , Fisher's exact test). Regressions reported on Table B17 confirm these results after controlling for observable heterogeneity.

Concluding this section, I do not find evidence that distributive behavior is significantly affected by the motive treatment in my context. As mentioned above, the correlation between beliefs and distributive behavior is rather weak. This stems from the fact that redistributive behavior is also heavily affected by factors that are not directly related to beliefs, such as winning the prize. This makes it challenging to identify a significant relationship between motivating belief distortion and redistributive behavior.

Taking a step back, the results complement the finding by Cassar and Klein (2019) who show that experiencing failure or success has a pure effect on distributive decisions that cannot be explained through variation in beliefs. My result shows that this effect can dominate if it comes through failure and through negative information about one's own relative ability. This may spark a feeling of solidarity with the worker who was not allocated the initial bonus. Future research should further study how these experiences shape preferences concerning redistribution above and beyond the belief channel.

## 6 Conclusion

This chapter asks whether individuals distort their beliefs about the importance of effort in economic outcomes in order to motivate themselves to exert effort. I employ a novel experimental design where individuals receive a noisy signal about the true importance of effort in success. To identify motivating belief distortion, I vary the knowledge about a subsequent task across subjects. I find that subjects distort luck-effort beliefs to motivate future effort. Subjects who

know that they face a task in the future believe that it is more likely that they are in an environment where their effort is likely to be rewarded. This form of belief distortion is more pronounced for subjects who received a signal that indicates that it is rather unlikely that the true state of the world rewards effort.

My main result enhances our understanding of how individuals form luck-effort beliefs. I advance this literature by providing causal evidence that these beliefs are not only shaped by past experiences and current information about income inequality but also by the incentives individuals expect to face in the future. This has wide-ranging implications: It implies that individual beliefs about the relative importance of luck and effort may be inelastic to information about the shape of the income distribution or the degree of intergenerational mobility if individuals believe that overcoming a lack of willpower remains important (e.g. due to a lack of social security). This implies that individual beliefs about the relative importance of luck and effort depend on expected levels of redistribution and post-tax inequality. Hence, average beliefs about the importance of effort and luck may remain stable for a relatively long time, unless the electorate does not expect a shift in distributive policy in the near future. These beliefs and preferences for redistribution, which have been shown to be tightly connected, can explain why Americans' support for redistribution has been rather stable since the 1970's even though there has been a sharp increase in economic inequality over the past decades (Ashok et al., 2015).

My findings open up new avenues for research. First, it would be interesting to replicate this study while varying the prior likelihood of being in the Performance-Environment. One could expect a tighter correlation between changes in beliefs due to motivating belief distortion and preferences for redistribution if subjects hold a high prior. This would give us a more granular view of how motivating belief distortion affects preferences for redistribution. Second, it would be equally interesting to test how the magnitude of motivating belief distortion interacts with the shape of the pre-tax income distribution. The demand to distort beliefs should be higher if pre-tax inequality is expected to be high and if a significant share of low-skilled individuals are overconfident in their skill and ability. Both extensions would enable us to identify to what extent motivating belief distortion has a reinforcing effect on inequality acceptance, i.e. individuals who expect low levels of redistribution and high levels of post-tax inequality distort beliefs to motivate future effort, and this makes them less likely to vote for more redistribution, fostering inequality.

## 7 References

- Alesina, Alberto and George-Marios Angeletos. Fairness and redistribution: US vs. Europe. *American Economic Review*, 95(5):960–980, 2005.
- Alesina, Alberto and Paola Giuliano. Chapter 4 - Preferences for Redistribution. In Benhabib, Jess, Alberto Bisin, and Matthew O. Jackson, editors, *Handbook of Social Economics*, volume 1, pages 93–131. North-Holland, 2011.
- Alesina, Alberto and Eliana La Ferrara. Preferences for redistribution in the land of opportunities. *Journal of Public Economics*, 89(5-6):897–931, 2005.
- Alesina, Alberto, Stefanie Stantcheva, and Edoardo Teso. Intergenerational Mobility and Preferences for Redistribution. *American Economic Review*, 108(2):521–554, 2018.
- Almås, Ingvild, Alexander W. Cappelen, and Bertil Tungodden. Cutthroat Capitalism versus Cuddly Socialism: Are Americans More Meritocratic and Efficiency-Seeking than Scandinavians? *Journal of Political Economy*, 128(5):1753–1788, 2020.
- Ambuehl, Sandro. An Offer You Can’t Refuse? Incentives Change How We Inform Ourselves and What We Believe. *Working Paper*, 2017.
- Ashok, Vevikin, Ilyana Kuziemko, and Ebonya Washington. Support for redistribution in an age of rising inequality: New stylized facts and some tentative explanations. *Brookings Papers on Economic Activity*, pages 367–405, 2015.
- Augenblick, Ned and Matthew Rabin. An Experiment on Time Preference and Misprediction in Unpleasant Tasks. *The Review of Economic Studies*, 86(3):941–975, 2019.
- Augenblick, Ned, Muriel Niederle, and Charles Sprenger. Working over Time: Dynamic Inconsistency in Real Effort Task. *The Quarterly Journal of Economics*, 130(3):1067–1115, 2015.
- Banerjee, Ritwik, Nabanita Datta Gupta, and Marie Claire Villeval. Feedback spillovers across tasks, self-confidence and competitiveness. *Games and Economic Behavior*, 123:127–170, 2020.
- Barron, Kai. Belief updating: Does the ‘good-news, bad-news’ asymmetry extend to purely financial domains? *Experimental Economics*, 24(1), 2021.
- Bénabou, Roland. The Economics of Motivated Beliefs. *Revue d’économie politique*, Vol. 125 (5):665–685, 2015.
- Bénabou, Roland and Jean Tirole. Self-Confidence and Personal Motivation. *The Quarterly Journal of Economics*, 117(3):871–915, 2002.
- Bénabou, Roland and Jean Tirole. Belief in a Just World and Redistributive Politics. *The Quarterly Journal of Economics*, 121(2):699–746, 2006.
- Bénabou, Roland and Jean Tirole. Identity, Morals, and Taboos: Beliefs as Assets. *The Quarterly Journal of Economics*, 126(2):805–855, 2011.

- Benjamin, Daniel J. Chapter 2 - Errors in probabilistic reasoning and judgment biases. In Bernheim, B. Douglas, Stefano DellaVigna, and David Laibson, editors, *Handbook of Behavioral Economics: Applications and Foundations 1*, volume 2, pages 69–186. North-Holland, 2019.
- Berglas, Steven and Edward E. Jones. Drug choice as a self-handicapping strategy in response to noncontingent success. *Journal of Personality and Social Psychology*, 36(4):405–417, 1978.
- Bracha, Anat and Donald J. Brown. Affective decision making: A theory of optimism bias. *Games and Economic Behavior*, 75(1):67–80, 2012.
- Brunnermeier, Markus K. and Jonathan A. Parker. Optimal Expectations. *American Economic Review*, 95(4):1092–1118, 2005.
- Camerer, Colin and Dan Lovallo. Overconfidence and Excess Entry: An Experimental Approach. *The American Economic Review*, 89(1):306–318, 1999.
- Cappelen, Alexander W, Astri Drange Hole, Erik Ø Sørensen, and Bertil Tungodden. The pluralism of fairness ideals: An experimental approach. *The American Economic Review*, 97(3):818–827, 2007.
- Cappelen, Alexander W, James Konow, Erik Ø Sørensen, and Bertil Tungodden. Just luck: An experimental study of risk-taking and fairness. *The American Economic Review*, 103(4):1398–1413, 2013.
- Cappelen, Alexander W, Johanna Mollerstrom, Bjørn-Atle Reme, and Bertil Tungodden. A Meritocratic Origin of Egalitarian Behavior. *Working Paper*, 2019.
- Cappelen, Alexander W., Ranveig Falch, and Bertil Tungodden. Fair and Unfair Income Inequality. In Zimmermann, Klaus F., editor, *Handbook of Labor, Human Resources and Population Economics*, pages 1–25. Springer International Publishing, Cham, 2020.
- Carrillo, Juan D. and Thomas Mariotti. Strategic Ignorance as a Self-Disciplining Device. *The Review of Economic Studies*, 67(3):529–544, 2000.
- Cassar, Lea and Arnd H. Klein. A Matter of Perspective: How Failure Shapes Distributive Preferences. *Management Science*, 65(11):5050–5064, 2019.
- Charness, Gary, Aldo Rustichini, and Jeroen van de Ven. Self-confidence and strategic behavior. *Experimental Economics*, 21(1):72–98, 2018.
- Chen, Daniel L., Martin Schonger, and Chris Wickens. oTree—An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9:88–97, 2016.
- Chen, Si and Hannah Schildberg-Hörisch. Looking at the bright side: The motivational value of confidence. *European Economic Review*, 120:103302, 2019.
- Chew, Soo Hong, Wei Huang, and Xiaojian Zhao. Motivated False Memory. *Journal of Political Economy*, 128(10):3913–3939, 2020.

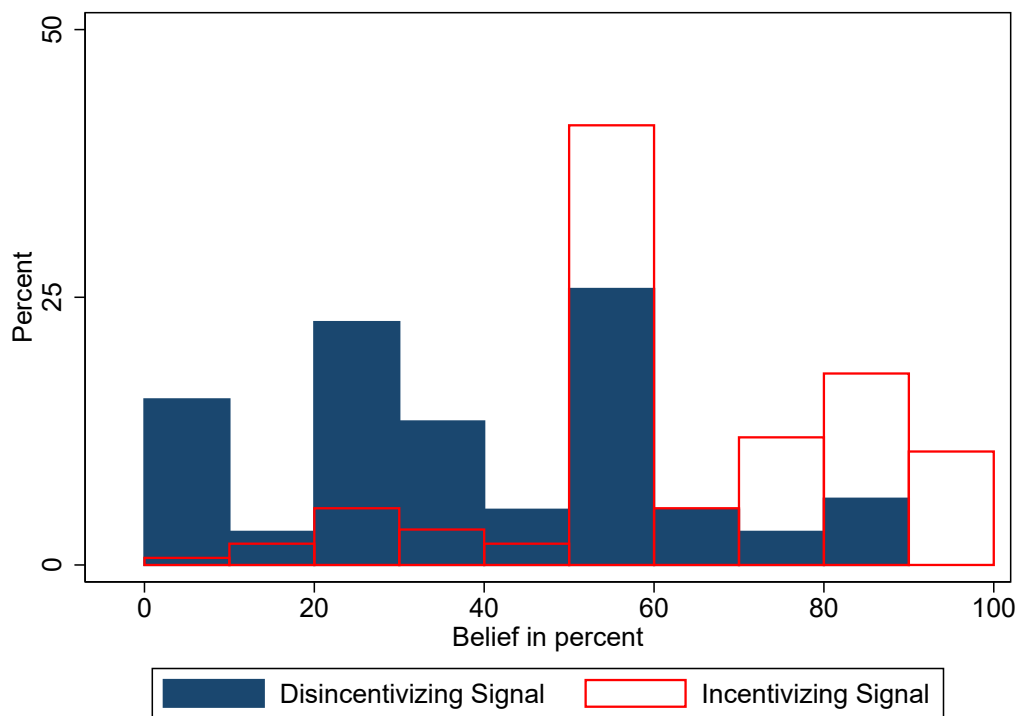
- Coutts, Alexander. Good news and bad news are still news: Experimental evidence on belief updating. *Experimental Economics*, 22(2):369–395, 2019a.
- Coutts, Alexander. Testing models of belief bias: An experiment. *Games and Economic Behavior*, 113:549–565, 2019b.
- Danz, David, Lise Vesterlund, and Alistair J. Wilson. Belief Elicitation: Limiting Truth Telling with Information on Incentives. *Working Paper*, 2020.
- Deffains, Bruno, Romain Espinosa, and Christian Thöni. Political self-serving bias and redistribution. *Journal of Public Economics*, 134:67–74, 2016.
- Di Tella, Rafael, Ricardo Perez-Truglia, Andres Babino, and Mariano Sigman. Conveniently Upset: Avoiding Altruism by Distorting Beliefs about Others’ Altruism. *American Economic Review*, 105(11):3416–3442, 2015.
- Dohmen, Thomas, Armin Falk, David Huffman, Uwe Sunde, Jürgen Schupp, and Gert G. Wagner. Individual Risk Attitudes: Measurement, Determinants, and Behavioral Consequences. *Journal of the European Economic Association*, 9(3):522–550, 2011.
- Durante, Ruben, Louis Putterman, and Joël van der Wee. Preferences for Redistribution and Perception of Fairness: An Experimental Study. *Journal of the European Economic Association*, 12(4):1059–1086, 2014.
- Erkal, Nisvan, Lata Gangadharan, and Boon Han Koh. By Chance or by Choice? Biased Attribution of Others’ Outcomes when Social Preferences Matter. *Working Paper*, 2021.
- Fehr, Dietmar, Hannes Rau, Stefan T. Trautmann, and Yilong Xu. Inequality, fairness and social capital. *European Economic Review*, 129:103566, 2020.
- Fong, Christina. Social preferences, self-interest, and the demand for redistribution. *Journal of Public Economics*, 82(2):225–246, 2001.
- Gärtner, Manja, Johanna Möllerström, and David Seim. Income Mobility, Luck/Effort Beliefs, and the Demand for Redistribution: Perceptions and Reality. *Working Paper*, 2019.
- Huck, Steffen, Nora Szech, and Lukas M Wenner. More Effort with Less Pay: On Information Avoidance, Optimistic Beliefs, and Performance. *Working Paper*, page 39, 2018.
- Kaur, Supreet, Michael Kremer, and Sendhil Mullainathan. Self-Control at Work. *Journal of Political Economy*, 123(6):1227–1277, 2015.
- König, Tobias, Sebastian Schweighofer-Kodritsch, and Georg Weizsäcker. Beliefs as a Means of Self-Control? Evidence from a Dynamic Student Survey. *Working Paper*, 2019.
- Konow, James. Fair Shares: Accountability and Cognitive Dissonance in Allocation Decisions. *American Economic Review*, 90(4):1072–1091, 2000.

- Krawczyk, Michal. A glimpse through the veil of ignorance: Equality of opportunity and support for redistribution. *Journal of Public Economics*, 94(1):131–141, 2010.
- Lefgren, Lars J., David P. Sims, and Olga B. Stoddard. Effort, luck, and voting for redistribution. *Journal of Public Economics*, 143:89–97, 2016.
- Lerner, Melvin J. The Belief in a Just World. In Lerner, Melvin J., editor, *The Belief in a Just World: A Fundamental Delusion*, Perspectives in Social Psychology, pages 9–30. Springer US, Boston, MA, 1980.
- Möbius, Markus, Muriel Niederle, Paul Niehaus, and Tanya Rosenblat. Managing Self-Confidence. *Working Paper*, 2014.
- Palan, Stefan and Christian Schitter. Prolific.ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, 17:22–27, 2018.
- Piketty, Thomas. Social mobility and redistributive politics. *The Quarterly Journal of Economics*, 110(3):551–584, 1995.
- Schwardmann, Peter and Joël van der Weele. Deception and self-deception. *Nature Human Behaviour*, 3(10):1055–1061, 2019.
- Valero, Vanessa. Redistribution and Beliefs about the Source of Income Inequality. *Working Paper*, 2020.
- Vischer, Thomas, Thomas Dohmen, Armin Falk, David Huffman, Jürgen Schupp, Uwe Sunde, and Gert G. Wagner. Validating an ultra-short survey measure of patience. *Economics Letters*, 120(2):142–145, 2013.
- Zimmermann, Florian. The Dynamics of Motivated Beliefs. *American Economic Review*, 110(2):337–363, 2020.

# Appendices

## 3.A Figures

Figure A1: Distribution of no-motive-group subjects' beliefs by signal



### 3.B Tables

Table B1: Summary statistics

	Mean	SD	Min	Max
Motive group	0.505	0.500	0	1
Age	33.808	12.380	18	78
Female	0.406	0.492	0	1
Employed full time	0.432	0.496	0	1
Student	0.244	0.430	0	1
Risk aversion	5.180	2.261	0	10
Patience	6.119	2.307	0	10
Right wing	0.246	0.431	0	1
Observations	495			

**Note.** This table presents mean, standard deviation, smallest observation, and largest observation for my sample. Motive group is a dummy, indicating whether the subject was assigned to the motive-group; age is the self-declared age; female is the a dummy indicating whether the subject self-identifies as a female; employed full-time is a dummy indicating whether the subject works full time; student is a dummy indicating whether the subject is currently a student; risk-aversion is a variable indicating the subject’s attitude towards risk as in Dohmen et al. (2011); patience is a variable indicating the subject’s patience Vischer et al. (2013); and rightwing is a dummy indicating whether the subject self-identifies as having a political orientation right of the center.

Table B2: Summary statistics by motive treatment

Variable	(1) No-motive-group	(2) Motive-group	(3) Difference
Age	33.439 (12.338)	34.156 (12.413)	0.717 (1.111)
Female	0.403 (0.492)	0.401 (0.491)	-0.002 (0.044)
Employed full time	0.427 (0.496)	0.440 (0.497)	0.013 (0.044)
Student	0.238 (0.427)	0.250 (0.434)	0.012 (0.038)
Risk aversion	5.000 (2.241)	5.381 (2.279)	0.381* (0.202)
Patience	6.211 (2.323)	6.048 (2.305)	-0.163 (0.207)
Right wing	0.262 (0.441)	0.234 (0.424)	-0.028 (0.039)
Observations	248	252	500

**Note.** This table presents results from a balancing test across motive treatment groups. The first column presents mean of the variable for no-motive-group subjects, while the second column presents the mean of the same variable for motive-group variables; the third column characterizes the difference between the two. Standard deviation is in parentheses. *Significance levels:* \*10%, \*\*5%, \*\*\*1%.



Table B3: Summary statistics by motive treatment

Variable	(1) Disincentivizing signal	(2) Incentivizing signal	(3) Difference
Age	33.273 (11.937)	34.091 (12.609)	0.818 (1.161)
Female	0.390 (0.489)	0.409 (0.492)	0.019 (0.046)
Employed full time	0.412 (0.494)	0.446 (0.498)	0.033 (0.046)
Student	0.254 (0.437)	0.238 (0.427)	-0.016 (0.040)
Risk aversion	4.989 (2.369)	5.303 (2.203)	0.315 (0.212)
Patience	6.220 (2.249)	6.078 (2.349)	-0.143 (0.217)
Right wing	0.254 (0.437)	0.245 (0.431)	-0.010 (0.040)
Observations	177	323	500

**Note.** This table presents results from a balancing test across subjects that received a disincentivizing and incentivizing signal. The first column presents mean of the variable for subjects that received a disincentivizing signal, while the second column presents the mean of the same variable for subjects that received an incentivizing signal; the third columns characterizes the difference between the two. Standard deviation is in parentheses. *Significance levels:* \*10%, \*\*5%, \*\*\*1%.

Table B4: Summary statistics by winning the bonus

Variable	(1) Not win bonus	(2) Win bonus	(3) Difference	(4) Conditional difference
age	34.782 (12.849)	32.858 (11.838)	-1.924* (1.109)	-0.350 (1.052)
female	0.359 (0.481)	0.443 (0.498)	0.084* (0.044)	0.081* (0.045)
employed full time	0.424 (0.495)	0.443 (0.498)	0.019 (0.044)	0.021 (0.045)
student	0.208 (0.407)	0.278 (0.449)	0.070* (0.038)	0.053 (0.039)
risk aversion	5.151 (2.279)	5.231 (2.257)	0.080 (0.203)	0.163 (0.206)
patience	6.090 (2.422)	6.165 (2.207)	0.075 (0.207)	0.110 (0.211)
rightwing	0.245 (0.431)	0.251 (0.434)	0.006 (0.039)	0.018 (0.039)
Observations	245	255	500	

**Note.** This table presents results from a balancing test across subjects that received a disincentivizing and incentivizing signal. Column (1) presents mean of the variable for subjects that received a disincentivizing signal, while column (2) presents the mean of the same variable for subjects that received an incentivizing signal; column (3) characterizes the difference between the two; and column (4) is the difference conditional on work-phase1 performance decile. Standard deviation is in parentheses. *Significance levels:* \*10%, \*\*5%, \*\*\*1%.

Table B5: Summary statistics by rank

Variable	(1) Rank first	(2) Rank second	(3) Difference	(4) Conditional difference
age	30.814 (10.331)	37.339 (13.613)	6.525*** (1.077)	2.414* (1.272)
female	0.412 (0.493)	0.390 (0.489)	-0.021 (0.044)	-0.008 (0.053)
employed full time	0.441 (0.497)	0.425 (0.495)	-0.016 (0.045)	-0.031 (0.053)
student	0.265 (0.442)	0.219 (0.415)	-0.045 (0.039)	0.019 (0.045)
risk aversion	5.099 (2.192)	5.303 (2.350)	0.203 (0.203)	-0.047 (0.234)
patience	6.129 (2.321)	6.128 (2.307)	-0.001 (0.208)	-0.141 (0.240)
rightwing	0.232 (0.423)	0.268 (0.444)	0.036 (0.039)	0.000 (0.045)
Observations	272	228	500	500

**Note.** This table presents results from a balancing test across subjects that received a disincentivizing and incentivizing signal. Column (1) presents mean of the variable for subjects that received a disincentivizing signal, while column (2) presents the mean of the same variable for subjects that received an incentivizing signal; column (3) characterizes the difference between the two; and column (4) is the difference conditional on work-phase1 performance decile. Standard deviation is in parentheses. *Significance levels:* \*10%, \*\*5%, \*\*\*1%.

Table B6: Regression of belief on winning on incentivizing signal and control variables (no-motive-group only)

	(1)	(2)	(3)
	Belief	Belief	Belief
Incentivizing signal	22.11*** (2.288)	24.63*** (3.113)	23.30*** (3.069)
Task 1 perf. decile		0.860 (0.579)	0.743 (0.665)
Age		-0.0566 (0.151)	-0.0406 (0.149)
Female		-0.284 (2.958)	-1.102 (2.935)
Employed full time		6.659** (2.973)	6.166** (2.944)
Student		5.353 (3.748)	3.689 (3.764)
Risk aversion		0.414 (0.631)	0.420 (0.617)
Patience		-0.230 (0.654)	-0.399 (0.647)
Right wing		-8.089** (3.546)	-8.289** (3.460)
Bonus win			8.697*** (2.992)
Ranks first			-0.807 (3.585)
Constant	36.18*** (2.036)	29.70*** (8.024)	28.86*** (8.012)
Education FE	No	Yes	Yes
Signal	All	All	All
Observations	500	245	245

**Note.** This table reports OLS regressions that regress beliefs on receiving an incentivizing signal. It only includes subjects that are in the no-motive-group. Table B18 replicates this table with subjects that passed a basic comprehension test at the end of the experiment. Standard errors are robust to heteroskedasticity and shown in parentheses. *Significance levels:* \*10%, \*\*5%, \*\*\*1%.

Table B7: Regression of belief on winning on motive group dummy and control variables

	(1)	(2)	(3)
	Belief	Belief	Belief
Motive group	6.712*** (2.263)	6.924*** (2.237)	6.478*** (2.221)
Bonus win		7.674*** (2.516)	7.223*** (2.519)
Ranks first		2.433 (3.049)	1.698 (3.083)
Age			-0.0944 (0.101)
Female			-1.988 (2.324)
Employed full time			2.651 (2.420)
Student			3.412 (3.012)
Risk aversion			1.117** (0.529)
Patience			-0.297 (0.505)
Right wing			-4.594* (2.700)
Constant	49.64*** (1.611)	42.29*** (2.843)	43.71*** (7.002)
Education FE	No	No	Yes
Observations	500	500	495

**Note.** This table reports OLS regressions that regress beliefs on the motive treatment variable, a dummy variable that indicates whether the subject won or lost the bonus, and a dummy that indicates whether the subject ranked first in terms of performance. Table B19 replicates this table with subjects that passed a basic comprehension test at the end of the experiment. All regressions are run on the whole sample. Standard errors are robust to heteroskedasticity and shown in parentheses. *Significance levels:* \*10%, \*\*5%, \*\*\*1%.

Table B8: Regression of belief on winning on motive group dummy and control variables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Belief	Belief	Belief	Belief	Belief	Belief	Belief	Belief
Motive group	9.398** (3.930)	9.599** (3.958)	2.731 (2.391)	3.610 (2.335)	7.189* (4.106)	7.578* (4.212)	2.313 (2.277)	3.145 (2.258)
Bonus win		-0.990 (4.451)		6.556** (2.903)		-1.497 (4.597)		5.057* (2.844)
Task 1 perf. decile		-0.749 (0.809)		1.048** (0.504)		-0.506 (0.902)		0.768 (0.512)
Age					0.153 (0.193)	0.124 (0.209)	-0.365*** (0.103)	-0.245** (0.106)
Female					1.954 (4.242)	2.249 (4.365)	-3.288 (2.373)	-3.983* (2.331)
Employed full time					-3.016 (4.032)	-2.913 (4.087)	4.877* (2.495)	5.087** (2.411)
Student					4.705 (5.307)	4.702 (5.333)	3.839 (3.287)	3.679 (3.227)
Risk aversion					1.786** (0.844)	1.687* (0.878)	0.0761 (0.544)	0.284 (0.542)
Patience					-0.109 (0.907)	-0.127 (0.906)	0.216 (0.517)	0.117 (0.512)
Right wing					-2.781 (4.812)	-2.607 (4.867)	-4.585 (3.040)	-4.783 (2.954)
Constant	34.23*** (2.368)	38.51*** (6.166)	59.54*** (1.746)	50.08*** (3.047)	22.49** (10.33)	26.80** (11.96)	70.49*** (6.705)	59.54*** (7.528)
Education FE	No	No	No	No	Yes	Yes	Yes	Yes
Signal	Disincentivizing	Disincentivizing	Incentivizing	Incentivizing	Disincentivizing	Disincentivizing	Incentivizing	Incentivizing
Observations	177	177	323	323	176	176	319	319

**Note.** This table reports OLS regressions that regress beliefs on the motive treatment variable and a dummy variable that indicates whether the subject won or lost the bonus. Columns (1), (2), (5), and (6) report results for regressions that include subjects that received a disincentivizing signal, while columns (3), (4), (7), and (8) report results for regressions that include subjects that received an incentivizing signal. Table B20 replicates this table excluding subjects that did not pass a basic comprehension test at the end of the experiment. Standard errors are robust to heteroskedasticity and shown in parentheses. *Significance levels:* \*10%, \*\*5%, \*\*\*1%.

Table B9: Regression beliefs on the interaction of being in the motive group and winning the bonus

	(1)	(2)	(3)	(4)	(5)	(6)
	Belief	Belief	Belief	Belief	Belief	Belief
Motive group	11.52*** (3.450)	16.27*** (5.807)	5.609 (3.952)	10.75*** (3.458)	14.65** (5.842)	4.193 (3.871)
Bonus win	12.73*** (3.177)	5.174 (4.995)	8.503** (3.996)	11.82*** (3.204)	5.091 (5.315)	6.090 (3.913)
Task 1 perf. decile	0.615 (0.417)	-0.899 (0.816)	1.054** (0.502)	0.491 (0.456)	-0.632 (0.897)	0.774 (0.512)
Motive * bonus win	-9.098** (4.457)	-14.45* (7.827)	-3.666 (4.783)	-8.407* (4.511)	-15.48* (8.025)	-1.905 (4.782)
Age				-0.0955 (0.102)	0.147 (0.209)	-0.243** (0.107)
Female				-2.080 (2.320)	1.985 (4.352)	-3.996* (2.339)
Employed full time				2.482 (2.425)	-3.425 (4.131)	5.061** (2.420)
Student				2.831 (3.027)	4.300 (5.437)	3.554 (3.234)
Risk aversion				1.103** (0.528)	1.761** (0.864)	0.280 (0.543)
Patience				-0.349 (0.504)	-0.120 (0.903)	0.0957 (0.517)
Right wing				-4.521* (2.687)	-2.486 (4.890)	-4.780 (2.954)
Constant	39.87*** (3.163)	36.67*** (6.070)	48.95*** (3.491)	42.08*** (7.100)	23.76* (12.25)	59.04*** (7.661)
Education FE	No	No	No	Yes	Yes	Yes
Signal	All	Disincentivizing	Incentivizing	All	Disincentivizing	Disincentivizing
Observations	500	177	323	495	176	319

**Note.** This table reports OLS regressions that regress beliefs on the motive treatment variable, winning a bonus dummy, and the interaction of the motive treatment and winning the bonus. Columns (1) and (4) reports regressions for the whole sample; columns (2) and (5) report regressions for subjects that received a disincentivizing signal; and columns (3) and (6) report regressions for subjects that received an incentivizing signal. Table B22 replicates this table excluding subjects that did not pass a basic comprehension test at the end of the experiment. Standard errors are robust to heteroskedasticity and shown in parentheses. *Significance levels:* \*10%, \*\*5%, \*\*\*1%.

Table B10: Regression beliefs on the interaction of being in the motive group and doing well in the task

	(1)	(2)	(3)	(4)	(5)	(6)
	Belief	Belief	Belief	Belief	Belief	Belief
Motive group	13.45*** (4.975)	42.50*** (12.68)	2.351 (5.365)	12.92*** (4.940)	37.97*** (13.16)	3.135 (5.125)
Bonus win	12.52*** (3.239)	9.391* (5.142)	10.11** (4.563)	11.59*** (3.268)	9.109* (5.477)	6.657 (4.426)
Task 1 perf. decile	0.831 (0.578)	0.756 (1.000)	0.543 (0.799)	0.730 (0.608)	0.810 (1.050)	0.604 (0.774)
Motive * bonus win	-8.675* (4.664)	-26.68*** (9.153)	-6.668 (5.854)	-7.941* (4.723)	-26.31*** (9.411)	-2.908 (5.862)
Motive * task 1 perf. decile	-0.412 (0.830)	-3.905** (1.643)	0.934 (1.019)	-0.461 (0.828)	-3.436** (1.658)	0.307 (0.987)
Age				-0.0978 (0.102)	0.115 (0.218)	-0.242** (0.107)
Female				-2.044 (2.325)	1.785 (4.363)	-4.048* (2.334)
Employed full time				2.521 (2.429)	-3.558 (4.212)	5.022** (2.426)
Student				2.735 (3.010)	2.316 (5.494)	3.517 (3.248)
Risk aversion				1.086** (0.528)	1.629* (0.866)	0.283 (0.545)
Patience				-0.366 (0.505)	-0.0879 (0.899)	0.109 (0.522)
Right wing				-4.518* (2.683)	-1.774 (4.937)	-4.764 (2.959)
Constant	38.82*** (3.717)	26.36*** (7.090)	50.86*** (4.136)	41.19*** (7.241)	16.35 (12.52)	59.58*** (7.884)
Education FE	Yes	Yes	Yes	Yes	Yes	Yes
Signal	All	Disincentivizing	Incentivizing	All	Disincentivizing	Incentivizing
Observations	500	177	323	495	176	319

**Note.** This table reports OLS regressions that regress beliefs on the motive treatment variable, winning a bonus dummy, performance decile, and the interaction of the motive treatment with the latter two variables. Columns (1) and (4) reports regressions for the whole sample; columns (2) and (5) report regressions for subjects that received a disincentivizing signal; and columns (3) and (6) report regressions for subjects that received an incentivizing signal. Table B23 replicates this table excluding subjects that did not pass a basic comprehension test at the end of the experiment. Standard errors are robust to heteroskedasticity and shown in parentheses. *Significance levels:* \*10%, \*\*5%, \*\*\*1%.

Table B11: Regression beliefs on the interaction of being in the motive group and demanding committing to exert future effort

	(1)	(2)	(3)	(4)	(5)	(6)
	Belief	Belief	Belief	Belief	Belief	Belief
Motive group	11.73*** (3.249)	15.10*** (5.528)	5.254 (3.454)	11.93*** (3.208)	14.45** (5.808)	4.200 (3.263)
Demands commitment	7.160** (3.186)	9.578** (4.641)	2.775 (3.496)	8.248*** (3.171)	11.14** (5.160)	2.204 (3.330)
Motive * demands commitment	-10.37** (4.503)	-12.58 (7.813)	-5.038 (4.792)	-11.62** (4.528)	-15.54* (8.347)	-3.777 (4.691)
Age				-0.140 (0.0950)	0.164 (0.186)	-0.358*** (0.101)
Female				-1.405 (2.367)	2.734 (4.382)	-3.410 (2.452)
Employed full time				2.285 (2.457)	-3.722 (3.952)	4.762* (2.510)
Student				3.607 (3.038)	4.673 (5.205)	3.729 (3.311)
Risk aversion				1.029* (0.525)	1.708** (0.856)	0.104 (0.553)
Patience				-0.269 (0.512)	-0.195 (0.924)	0.206 (0.522)
Right wing				-4.884* (2.745)	-3.775 (4.931)	-4.712 (3.057)
Constant	46.29*** (2.278)	30.18*** (3.206)	58.16*** (2.403)	47.29*** (6.263)	17.88* (10.78)	69.18*** (6.730)
Education FE	Yes	Yes	Yes	Yes	Yes	Yes
Signal	All	Disincentivizing	Incentivizing	All	Disincentivizing	Incentivizing
Observations	500	177	323	495	176	319

**Note.** This table reports OLS regressions that regress beliefs on the motive treatment variable, enjoying the commitment dummy, and their interaction. Columns (1) and (4) reports regressions for the whole sample; columns (2) and (5) report regressions for subjects that received a disincentivizing signal; and columns (3) and (6) report regressions for subjects that received an incentivizing signal. Table B24 replicates this table excluding subjects that did not pass a basic comprehension test at the end of the experiment. Standard errors are robust to heteroskedasticity and shown in parentheses. *Significance levels:* \*10%, \*\*5%, \*\*\*1%.



Table B12: Regression beliefs on the interaction of being in the motive group and enjoying the task

	(1)	(2)	(3)	(4)	(5)	(6)
	Belief	Belief	Belief	Belief	Belief	Belief
Motive group	5.779 (5.715)	-2.007 (10.85)	4.660 (6.317)	5.959 (5.675)	-2.394 (11.13)	3.541 (6.116)
Bonus win	13.31*** (3.239)	6.421 (5.095)	8.412** (4.075)	12.41*** (3.266)	6.235 (5.456)	5.861 (3.955)
Motive * bonus win	-10.39** (4.502)	-16.18** (7.781)	-4.063 (4.786)	-9.527** (4.544)	-16.86** (8.056)	-2.200 (4.739)
Enjoys task	-4.033 (4.005)	-5.876 (6.856)	0.889 (4.603)	-3.626 (4.030)	-5.841 (7.667)	1.811 (4.287)
Motive * enjoys task	7.922 (5.907)	22.29** (11.07)	1.358 (6.208)	6.687 (5.856)	20.85* (11.68)	0.941 (5.823)
Task 1 perf. decile	0.628 (0.419)	-0.720 (0.809)	1.033** (0.507)	0.506 (0.460)	-0.470 (0.892)	0.739 (0.523)
Age				-0.107 (0.103)	0.137 (0.209)	-0.244** (0.106)
Female				-2.131 (2.331)	2.811 (4.267)	-3.769 (2.352)
Employed full time				2.411 (2.433)	-4.001 (4.041)	5.180** (2.456)
Student				2.271 (3.037)	3.210 (5.379)	3.267 (3.254)
Risk aversion				1.150** (0.530)	1.606* (0.861)	0.232 (0.556)
Patience				-0.283 (0.509)	-0.0538 (0.947)	0.0940 (0.518)
Right wing				-4.379 (2.698)	-2.089 (4.856)	-4.998* (3.000)
Constant	42.70*** (4.274)	40.07*** (7.704)	48.43*** (4.968)	44.62*** (7.563)	27.61** (12.51)	58.26*** (8.174)
Education FE	No	No	No	Yes	Yes	Yes
Signal	All	Disincentivizing	Incentivizing	All	Disincentivizing	Incentivizing
Observations	498	177	321	493	176	317

**Note.** This table reports OLS regressions that regress beliefs on the motive treatment variable, winning a bonus dummy, enjoying the task dummy, and the interaction of the motive treatment with the latter two dummy variables. Columns (1) and (4) reports regressions for the whole sample; columns (2) and (5) report regressions for subjects that received a disincentivizing signal; and columns (3) and (6) report regressions for subjects that received an incentivizing signal. Table B25 replicates this table excluding subjects that did not pass a basic comprehension test at the end of the experiment. Standard errors are robust to heteroskedasticity and shown in parentheses. *Significance levels:* \*10%, \*\*5%, \*\*\*1%.

Table B13: Regressing false recall on motive group, signal, its interaction term and observables

	(1)	(2)	(3)	(4)
	False recall	False recall	False recall	False recall
Motive group	-0.0302 (0.0309)	-0.0190 (0.0301)	-0.0680 (0.0637)	-0.0610 (0.0638)
Incentivizing signal		-0.152*** (0.0355)	-0.189*** (0.0503)	-0.176*** (0.0500)
Motive * pos. signal			0.0758 (0.0708)	0.0562 (0.0721)
Task 1 perf. decile				-0.0144** (0.00580)
Age				-0.000346 (0.00167)
Female				-0.0647** (0.0305)
Employed full time				-0.0297 (0.0311)
Student				-0.0332 (0.0383)
Risk aversion				0.00351 (0.00757)
Patience				0.00128 (0.00671)
Right wing				0.00643 (0.0385)
Constant	0.153*** (0.0229)	0.246*** (0.0357)	0.268*** (0.0452)	0.366*** (0.108)
Education FE	No	No	No	Yes
Signal	All	All	All	All
Observations	500	500	500	495

**Note.** This table reports OLS regressions that regress not recalling the correct signal on a motive group dummy, receiving an incentivizing signal, and their interaction. Not recalling the correct signal includes (i) recalling a false signal or (ii) declaring that one did not remember the signal. The sample includes subjects all subjects. Standard errors are robust to heteroskedasticity and shown in parentheses. *Significance levels:* \*10%, \*\*5%, \*\*\*1%.

Table B14: Regression beliefs controlling for recalled signal

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Belief	Belief	Belief	Belief	Belief	Belief	Belief	Belief
Motive group	5.298** (2.211)	6.152*** (2.213)	10.82** (4.226)	11.49*** (4.220)	2.421 (2.532)	2.500 (2.543)	9.642** (4.456)	2.803 (2.397)
Incentivizing signal	23.82*** (2.454)							
Recalled signal		23.95*** (2.545)		10.12* (5.774)		12.36*** (1.742)	11.81* (6.430)	14.54*** (4.299)
Task 1 perf. decile							0.00380 (0.863)	1.315*** (0.469)
Age							0.111 (0.234)	-0.280** (0.112)
Female							2.892 (4.569)	-3.007 (2.467)
Employed full time							-2.266 (4.180)	5.127** (2.541)
Student							7.021 (5.631)	4.478 (3.369)
Risk aversion							1.744* (0.919)	0.176 (0.570)
Patience							0.194 (0.937)	0.204 (0.554)
Right wing							-0.181 (5.193)	-3.195 (3.218)
Constant	34.51*** (2.198)	32.94*** (2.373)	32*** (2.560)	30.33*** (2.819)	59.86*** (1.850)	47.50*** (2.543)	16.15 (12.72)	44.71*** (9.791)
Education FE	No	No	No	No	No	No	Yes	Yes
Signal	All	All	Disincentivizing	Disincentivizing	Incentivizing	Incentivizing	Disincentivizing	Disincentivizing
Observations	454	454	156	156	298	298	156	295

**Note.** This table reports OLS regressions that regress belief on the motive treatment variable. Column (1) and (2) report results from a regression that excludes subjects that did not forget the signal. Incentivizing signal denotes the actual signal and recalled signal denotes the signal they claim to recall. Column (3) regresses beliefs on the treatment dummy for subject that recalled a signal for subjects that received a disincentivizing signal. Columns (4) and (7) regress beliefs on the treatment dummy controlling for the signal subjects recalled for subjects that received a disincentivizing signal. Column (5) regresses beliefs on the treatment dummy for subject that recalled a signal for subjects that received an incentivizing signal. Columns (6) and (8) regress beliefs on the treatment dummy controlling for the signal subjects recalled for subjects that received a disincentivizing signal. Standard errors are robust to heteroskedasticity and shown in parentheses. *Significance levels:* \*10%, \*\*5%, \*\*\*1%.

Table B15: Redistribution regression interacting the treatment variable with merit dummy

	(1)	(2)	(3)	(4)	(5)	(6)
	Redistr.	Redistr.	Redistr.	Redistr.	Redistr.	Redistr.
Motive group	0.0139 (0.210)	0.264 (0.298)	-0.110 (0.268)	0.000668 (0.209)	0.124 (0.307)	-0.0600 (0.276)
Meritocr.	-0.317** (0.140)	-0.165 (0.247)	-0.405** (0.169)	-0.294** (0.135)	-0.181 (0.226)	-0.402** (0.176)
Motive * meritocr-	0.0386 (0.225)	-0.0915 (0.326)	0.109 (0.286)	0.0578 (0.223)	-0.00500 (0.332)	0.0708 (0.295)
Age				0.00292 (0.00344)	0.00678 (0.00602)	0.000239 (0.00428)
Female				-0.234*** (0.0812)	-0.245* (0.141)	-0.205** (0.100)
Employed full time				-0.0574 (0.0828)	-0.0487 (0.140)	-0.0514 (0.103)
Student				0.00302 (0.0968)	-0.0380 (0.155)	0.0256 (0.124)
Risk aversion				0.00798 (0.0182)	-0.00215 (0.0282)	0.00780 (0.0238)
Patience				-0.00317 (0.0157)	-0.0602** (0.0241)	0.0287 (0.0205)
Right wing				-0.0576 (0.0912)	-0.230 (0.150)	0.0379 (0.118)
Constant	1.444*** (0.127)	1.308*** (0.230)	1.522*** (0.150)	1.507*** (0.204)	1.723*** (0.345)	1.470*** (0.255)
Signal	All	Disincentivizing	Incentivizing	All	Disincentivizing	Incentivizing
Education FE	No	No	No	Yes	Yes	Yes
Observations	500	177	323	495	176	319

**Note.** This table reports OLS regressions that interact the motive treatment variable with a dummy variable that indicates whether the subject self-identifies as a meritocrat (agrees with 7 out of 10 that subjects who transcribed more images should receive more money). Columns (1) and (4) report results for the whole sample; column (2) and (5) report results for a sample that only include subjects who received an incentivizing signal and column (3) and (6) reports results for a sample that only include subjects who received a disincentivizing signal. Table B26 replicates this table excluding subjects that did not pass a basic comprehension test at the end of the experiment. Standard errors are robust to heteroskedasticity and shown in parentheses. *Significance levels:* \*10%, \*\*5%, \*\*\*1%.

Table B16: Regression on extensive margin redistribution behavior

	(1)	(2)	(3)	(4)	(5)	(6)
	Redistr.	No redistrib.	Full redistrib	Redistr.	No redistrib.	Full redistrib
Belief	-0.00295** (0.00144)	0.00219*** (0.000772)	-0.000654 (0.000855)	-0.00262* (0.00145)	0.00221*** (0.000774)	-0.000317 (0.000877)
Age				0.000806 (0.00372)	-0.000473 (0.00201)	0.00105 (0.00216)
Female				-0.243*** (0.0815)	0.141*** (0.0419)	-0.127*** (0.0459)
Employed full time				-0.0459 (0.0825)	0.00690 (0.0419)	-0.0143 (0.0488)
Student				0.0249 (0.0985)	-0.0945* (0.0518)	-0.0410 (0.0588)
Risk aversion				0.00704 (0.0180)	0.000877 (0.00889)	0.000901 (0.0104)
Patience				-0.00655 (0.0155)	-0.000222 (0.00785)	-0.00740 (0.00968)
Right wing				-0.0957 (0.0919)	0.0692 (0.0476)	-0.0592 (0.0530)
Constant	1.346*** (0.0827)	0.128*** (0.0428)	0.447*** (0.0508)	1.596*** (0.221)	0.0323 (0.118)	0.615*** (0.138)
Perform. decile FE	No	No	No	Yes	Yes	Yes
Observations	500	500	500	495	495	495

**Note.** This table reports OLS regression that characterize the effect of beliefs on redistribution behavior. Columns (1) and (4) regress amount redistributed on beliefs; columns (2) and (5) regress a dummy variable that indicates whether the subject has redistributed at all; columns (3) and (6) regresses a dummy variable that indicates an equal split on beliefs. These regressions are run for the whole sample. Standard errors are robust to heteroskedasticity and shown in parentheses. *Significance levels:* \*10%, \*\*5%, \*\*\*1%.

Table B17: Regression on extensive margin redistribution behavior

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Redistr.	Redistr.	Redistr.	Redistr.	Redistr.	Redistr.	Redistr.	Redistr.
Bonus win	-0.0285 (0.156)	-0.507*** (0.109)			0.0187 (0.150)	-0.467*** (0.114)		
Task 1 perf. decile	0.00141 (0.0256)	0.0171 (0.0187)	0.00971 (0.0160)	0.00804 (0.0162)	0.0196 (0.0251)	0.00467 (0.0202)	0.0161 (0.0170)	0.00985 (0.0171)
Incentivizing signal			0.180* (0.0935)	-0.300*** (0.0906)			0.160* (0.0947)	-0.268*** (0.0924)
Age					0.00924 (0.00642)	-0.00524 (0.00444)	0.00310 (0.00437)	0.00260 (0.00460)
Female					-0.266* (0.139)	-0.180* (0.0980)	-0.230** (0.102)	-0.239** (0.0956)
Employed full time					-0.0554 (0.142)	-0.0425 (0.0997)	-0.136 (0.104)	0.0331 (0.0968)
Student					-0.0135 (0.157)	0.0389 (0.122)	-0.104 (0.123)	0.113 (0.109)
Risk aversion					0.00832 (0.0269)	-0.00338 (0.0230)	0.0206 (0.0211)	-0.00482 (0.0200)
Patience					-0.0585** (0.0240)	0.0266 (0.0196)	-0.0244 (0.0177)	-0.0131 (0.0193)
Right wing					-0.276* (0.146)	0.0288 (0.116)	-0.111 (0.114)	-0.150 (0.104)
Constant	1.248*** (0.191)	1.347*** (0.0927)	1.192*** (0.105)	1.201*** (0.105)	1.391*** (0.373)	1.563*** (0.255)	1.361*** (0.256)	1.342*** (0.267)
Education FE	No	No	No	No	Yes	Yes	Yes	Yes
Signal	Disincentivizing	Incentivizing	All	All	Disincentivizing	Incentivizing	All	All
Observations	177	323	325	352	176	319	322	349

**Note.** This table reports OLS regressions that characterize redistribution behavior across different events observed by subjects. Columns (1) and (5) characterizes the effect of winning the bonus for subjects that received a disincentivizing signal; columns (2) and (6) characterizes the effect of winning the bonus for subjects that received an incentivizing signal; columns (3) and (7) characterizes the effect of receiving an incentivizing signal excluding subjects that ranked first and won the bonus; columns (4) and (8) characterizes the effect of receiving an incentivizing signal excluding subjects that ranked second and did not win the bonus. Standard errors are robust to heteroskedasticity and shown in parentheses. *Significance levels:* \*10%, \*\*5%, \*\*\*1%.

Table B18: Regression of belief on winning on incentivizing signal and control variables (no-motive-group only), replication

	(1)	(2)	(3)
	Belief	Belief	Belief
Incentivizing signal	23.57*** (2.464)	26.63*** (3.338)	24.85*** (3.397)
Task 1 perf. decile		0.623 (0.636)	0.908 (0.720)
Age		-0.104 (0.174)	-0.119 (0.165)
Female		0.439 (3.179)	0.381 (3.168)
Employed full time		5.411* (3.229)	5.043 (3.167)
Student		6.515* (3.707)	4.949 (3.736)
Risk aversion		0.476 (0.665)	0.557 (0.655)
Patience		-0.0956 (0.759)	-0.204 (0.746)
Right wing		-9.297** (3.787)	-10.00*** (3.636)
Bonus win			8.201** (3.236)
Ranks first			-4.094 (3.945)
Constant	35.70*** (2.176)	29.31*** (9.109)	28.16*** (8.989)
Education FE	No	Yes	Yes
Signal	All	All	All
Observations	413	198	198

**Note.** This table reports OLS regressions that regress beliefs on receiving an incentivizing signal. It only includes subjects that are in the no-motive-group. This table replicates Table B6 with subjects that passed a basic comprehension test at the end of the experiment. Standard errors are robust to heteroskedasticity and shown in parentheses. *Significance levels:* \*10%, \*\*5%, \*\*\*1%.

Table B19: Regression of belief on winning on motive group dummy and control variables, replication

	(1)	(2)	(3)	(4)
	Belief	Belief	Belief	Belief
Motive group	5.896** (2.514)	6.206** (2.498)	5.786** (2.492)	5.987** (2.485)
Bonus win		7.917*** (2.758)		7.058** (2.747)
Ranks first		3.058 (3.522)		1.684 (3.558)
Task 1 perf. decile		0.332 (0.583)		0.279 (0.618)
Age			-0.189* (0.112)	-0.131 (0.119)
Female			-0.558 (2.594)	-1.135 (2.571)
Employed full time			1.835 (2.780)	2.037 (2.738)
Student			4.725 (3.291)	4.443 (3.243)
Risk aversion			1.055* (0.577)	1.100* (0.586)
Patience			-0.163 (0.595)	-0.234 (0.585)
Right wing			-3.443 (3.197)	-3.448 (3.145)
Constant	50.00*** (1.763)	42.02*** (3.468)	51.36*** (6.729)	43.63*** (8.155)
Education FE	No	No	Yes	Yes
Signal	All	All	All	All
Observations	413	413	409	409

**Note.** This table reports OLS regressions that regress beliefs on the motive treatment variable, a dummy variable that indicates whether the subject won or lost the bonus, and a dummy that indicates whether the subject ranked first in terms of performance. This table replicates Table B7 with subjects that passed a basic comprehension test at the end of the experiment. All regressions are run on the whole sample. Standard errors are robust to heteroskedasticity and shown in parentheses. *Significance levels:* \*10%, \*\*5%, \*\*\*1%.



Table B20: Regression of belief on winning on motive group dummy and control variables, replication

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Belief	Belief	Belief	Belief	Belief	Belief	Belief	Belief
Motive group	9.389** (4.169)	9.719** (4.185)	2.337 (2.608)	3.242 (2.590)	7.679* (4.352)	8.158* (4.436)	2.024 (2.492)	2.722 (2.504)
Bonus win		-3.065 (5.267)		6.287* (3.258)		-3.441 (5.503)		3.699 (3.160)
Task 1 perf. decile		-0.972 (0.929)		0.791 (0.575)		-0.699 (1.004)		0.517 (0.592)
Age					0.154 (0.229)	0.143 (0.253)	-0.384*** (0.109)	-0.295** (0.116)
Female					2.777 (4.612)	3.050 (4.718)	-3.736 (2.568)	-4.194 (2.555)
Employed full time					-3.282 (4.424)	-3.216 (4.528)	4.507* (2.662)	4.930* (2.650)
Student					7.184 (5.919)	7.357 (6.054)	5.087 (3.170)	5.165 (3.191)
Risk aversion					1.641* (0.913)	1.519 (0.959)	0.230 (0.605)	0.328 (0.602)
Patience					0.186 (1.006)	0.158 (1.001)	0.258 (0.597)	0.174 (0.591)
Right wing					-3.768 (5.284)	-3.448 (5.405)	-2.830 (3.418)	-2.955 (3.390)
Constant	33.54*** (2.516)	39.89*** (7.434)	60.66*** (1.854)	52.08*** (3.699)	19.23* (11.23)	24.92* (13.80)	69.35*** (7.186)	61.29*** (8.552)
Education FE	No	No	No	No	Yes	Yes	Yes	Yes
Signal	Disincentivizing	Disincentivizing	Incentivizing	Incentivizing	Disincentivizing	Disincentivizing	Incentivizing	Incentivizing
Observations	154	154	259	259	153	153	256	256

**Note.** This table reports OLS regressions that regress beliefs on the motive treatment variable and a dummy variable that indicates whether the subject won or lost the bonus. Columns (1), (2), (5), and (6) report results for regressions that include subjects that received a disincentivizing signal, while columns (3), (4), (7), and (8) report results for regressions that include subjects that received an incentivizing signal. This table replicates Table B8 with subjects that passed a basic comprehension test at the end of the experiment. Standard errors are robust to heteroskedasticity and shown in parentheses. *Significance levels:* \*10%, \*\*5%, \*\*\*1%.

Table B21: Results of the updating regression specified in (3.9) (subjects passed comprehension test at the end of experiment)

Coefficient	Estimate
$\alpha_+$	1.36 (0.244)
$\beta_+$	0.32 (0.361)
$\alpha_-$	1.14 (0.193)
$\beta_-$	-0.71** (0.322)
<hr/>	
$H_0 : \alpha_+ = \alpha_- = 1$	$F(2, 409) = 1.12$
$H_0 : \beta_+ = \beta_-$	$F(1, 409) = 3.77$
<hr/>	
N	413
$R^2$	0.16

**Note.** This table replicates Table 3.3 excluding subjects that did not answer correctly a comprehension question at the end of the experiment. It shows coefficient estimates for the regression characterized in (3.9).  $\alpha$  coefficients test the null-hypothesis that  $\alpha = 1$  while  $\beta$  coefficients test the null-hypothesis that  $\beta = 0$ . Robust standard errors are in parentheses. *Significance levels:* \*10%, \*\*5%, \*\*\*1%.

Table B22: Regression beliefs on the interaction of being in the motive group and winning the bonus (subjects passed comprehension test at the end of experiment)

	(1)	(2)	(3)	(4)	(5)	(6)
	Belief	Belief	Belief	Belief	Belief	Belief
Motive group	10.96*** (3.964)	17.03*** (5.941)	3.793 (4.696)	10.74*** (3.941)	16.50*** (5.940)	2.230 (4.503)
Bonus win	13.03*** (3.481)	4.568 (5.739)	6.764 (4.446)	11.95*** (3.487)	5.223 (6.225)	3.258 (4.276)
Task 1 perf. decile	0.637 (0.464)	-1.104 (0.920)	0.797 (0.573)	0.437 (0.517)	-0.847 (0.981)	0.512 (0.595)
Motive * bonus win	-8.904* (5.008)	-16.41** (8.266)	-0.921 (5.491)	-8.929* (5.010)	-18.85** (8.551)	0.820 (5.359)
Age				-0.139 (0.118)	0.161 (0.248)	-0.296** (0.117)
Female				-1.044 (2.573)	3.393 (4.689)	-4.192 (2.561)
Employed full time				1.838 (2.727)	-4.050 (4.562)	4.941* (2.660)
Student				3.957 (3.213)	7.075 (6.143)	5.219 (3.220)
Risk aversion				1.129* (0.579)	1.596* (0.939)	0.323 (0.604)
Patience				-0.261 (0.584)	0.212 (0.990)	0.183 (0.596)
Right wing				-3.666 (3.122)	-4.059 (5.404)	-2.932 (3.410)
Constant	39.39*** (3.810)	37.39*** (7.204)	51.75*** (4.286)	41.52*** (8.274)	20.78 (13.82)	61.56*** (8.844)
Education FE	No	No	No	Yes	Yes	Yes
Signal	All	Disincentivizing	Incentivizing	All	Disincentivizing	Incentivizing
Observations	413	154	259	409	153	256

**Note.** This table reports OLS regressions that regress beliefs on the motive treatment variable, winning a bonus dummy, and the interaction of the motive treatment and winning the bonus. Columns (1) and (4) reports regressions for the whole sample; columns (2) and (5) report regressions for subjects that received a disincentivizing signal; and columns (3) and (6) report regressions for subjects that received an incentivizing signal. This table replicates Table B9 with subjects that passed a basic comprehension test at the end of the experiment. Standard errors are robust to heteroskedasticity and shown in parentheses. *Significance levels:* \*10%, \*\*5%, \*\*\*1%.

Table B23: Regression beliefs on the interaction of being in the motive group and doing well in the task

	(1)	(2)	(3)	(4)	(5)	(6)
	Belief	Belief	Belief	Belief	Belief	Belief
Motive group	13.94** (6.094)	48.86*** (14.21)	-0.470 (6.719)	14.65** (5.989)	44.69*** (14.59)	2.654 (6.258)
Bonus win	12.83*** (3.516)	11.10* (5.793)	8.573* (5.001)	11.69*** (3.525)	11.20* (6.465)	3.057 (4.805)
Task 1 perf. decile	0.956 (0.657)	1.019 (1.068)	0.160 (0.932)	0.852 (0.697)	0.974 (1.064)	0.578 (0.912)
Motive * bonus win	-8.507* (5.124)	-32.50*** (10.22)	-4.368 (6.533)	-8.400 (5.141)	-32.86*** (10.65)	1.172 (6.486)
Motive * task 1 perf. decile	-0.582 (0.924)	-4.637** (1.797)	1.122 (1.170)	-0.765 (0.921)	-4.049** (1.803)	-0.113 (1.124)
Age				-0.142 (0.118)	0.122 (0.252)	-0.295** (0.117)
Female				-0.982 (2.583)	3.871 (4.710)	-4.175 (2.565)
Employed full time				1.974 (2.735)	-4.263 (4.672)	4.974* (2.688)
Student				3.890 (3.186)	4.946 (6.169)	5.259 (3.247)
Risk aversion				1.110* (0.578)	1.357 (0.943)	0.326 (0.604)
Patience				-0.285 (0.586)	0.335 (1.001)	0.177 (0.602)
Right wing				-3.808 (3.104)	-3.680 (5.474)	-2.973 (3.389)
Constant	37.69*** (4.653)	23.59*** (7.977)	54.43*** (5.323)	39.61*** (8.586)	10.40 (13.74)	61.28*** (9.272)
Education FE	Yes	Yes	Yes	Yes	Yes	Yes
Signal	All	Disincentivizing	Incentivizing	All	Disincentivizing	Incentivizing
Observations	413	154	259	409	153	256

**Note.** This table reports OLS regressions that regress beliefs on the motive treatment variable, winning a bonus dummy, performance decile, and the interaction of the motive treatment with the latter two variables. Columns (1) and (4) reports regressions for the whole sample; columns (2) and (5) report regressions for subjects that received a disincentivizing signal; and columns (3) and (6) report regressions for subjects that received an incentivizing signal. This table replicates Table B10 with subjects that passed a basic comprehension test at the end of the experiment. Standard errors are robust to heteroskedasticity and shown in parentheses. *Significance levels:* \*10%, \*\*5%, \*\*\*1%.

Table B24: Regression beliefs on the interaction of being in the motive group and demanding committing to exert future effort (subjects passed comprehension test at the end of experiment)

	(1)	(2)	(3)	(4)	(5)	(6)
	Belief	Belief	Belief	Belief	Belief	Belief
Motive group	10.11*** (3.492)	12.99** (5.851)	4.177 (3.700)	10.88*** (3.456)	12.92** (6.107)	3.479 (3.564)
Demands commitment	5.862* (3.521)	5.847 (5.107)	2.721 (3.731)	7.339** (3.545)	8.954 (5.988)	1.615 (3.661)
Motive * demands commitment	-9.164* (5.024)	-8.184 (8.412)	-3.939 (5.216)	-11.21** (5.081)	-12.13 (9.074)	-3.094 (5.198)
Age				-0.177 (0.110)	0.154 (0.223)	-0.381*** (0.108)
Female				-0.720 (2.629)	3.314 (4.767)	-3.816 (2.677)
Employed full time				1.493 (2.786)	-3.804 (4.415)	4.448* (2.677)
Student				4.379 (3.267)	6.620 (5.849)	4.996 (3.191)
Risk aversion				1.142* (0.583)	1.699* (0.927)	0.258 (0.616)
Patience				-0.173 (0.596)	0.114 (1.014)	0.264 (0.602)
Right wing				-4.085 (3.203)	-5.444 (5.521)	-2.927 (3.438)
Constant	47.43*** (2.382)	31.25*** (3.265)	59.38*** (2.499)	47.58*** (6.976)	15.82 (11.66)	68.44*** (7.180)
Education FE	Yes	Yes	Yes	Yes	Yes	Yes
Signal	All	Disincentivizing	Incentivizing	All	Disincentivizing	Incentivizing
Observations	413	154	259	409	153	256

**Note.** This table reports OLS regressions that regress beliefs on the motive treatment variable, enjoying the commitment dummy, and their interaction. Columns (1) and (4) reports regressions for the whole sample; columns (2) and (5) report regressions for subjects that received a disincentivizing signal; and columns (3) and (6) report regressions for subjects that received an incentivizing signal. This table replicates Table B11 with subjects that passed a basic comprehension test at the end of the experiment. Standard errors are robust to heteroskedasticity and shown in parentheses. *Significance levels:* \*10%, \*\*5%, \*\*\*1%.

Table B25: Regression beliefs on the interaction of being in the motive group and enjoying the task (subjects passed comprehension test at the end of experiment)

	(1)	(2)	(3)	(4)	(5)	(6)
	Belief	Belief	Belief	Belief	Belief	Belief
Motive group	2.657 (6.224)	-1.917 (11.19)	0.702 (7.121)	3.740 (6.273)	-1.121 (11.57)	-0.331 (6.933)
Bonus win	13.76*** (3.548)	6.566 (5.758)	6.570 (4.509)	12.75*** (3.556)	7.278 (6.324)	2.865 (4.304)
Motive * bonus win	-10.85** (5.052)	-18.56** (8.167)	-1.943 (5.434)	-10.66** (5.046)	-20.75** (8.578)	0.0984 (5.258)
Enjoys task	-4.421 (4.269)	-7.924 (7.410)	2.354 (4.906)	-4.077 (4.372)	-8.121 (8.128)	3.146 (4.629)
Motive * enjoys task	11.67* (6.358)	23.62** (11.45)	4.450 (6.699)	9.996 (6.466)	22.31* (12.28)	3.530 (6.494)
Task 1 perf. decile	0.635 (0.466)	-0.892 (0.905)	0.706 (0.577)	0.434 (0.520)	-0.655 (0.967)	0.391 (0.612)
Age				-0.159 (0.119)	0.137 (0.246)	-0.308*** (0.116)
Female				-0.941 (2.567)	4.288 (4.562)	-3.612 (2.535)
Employed full time				1.690 (2.728)	-4.803 (4.456)	4.984* (2.718)
Student				3.089 (3.205)	5.505 (6.118)	4.533 (3.213)
Risk aversion				1.153** (0.580)	1.464 (0.938)	0.206 (0.618)
Patience				-0.151 (0.598)	0.346 (1.023)	0.212 (0.595)
Right wing				-3.471 (3.120)	-3.353 (5.355)	-3.369 (3.430)
Constant	42.43*** (4.838)	41.88*** (8.928)	50.64*** (5.743)	44.38*** (8.586)	25.45* (14.10)	60.96*** (9.244)
Education FE	No	No	No	Yes	Yes	Yes
Signal	All	Disincentivizing	Incentivizing	All	Disincentivizing	Incentivizing
Observations	411	154	257	407	153	254

**Note.** This table reports OLS regressions that regress beliefs on the motive treatment variable, winning a bonus dummy, enjoying the task dummy, and the interaction of the motive treatment with the latter two dummy variables. Columns (1) and (4) reports regressions for the whole sample; columns (2) and (5) report regressions for subjects that received a disincentivizing signal; and columns (3) and (6) report regressions for subjects that received an incentivizing signal. This table replicates Table B25 with subjects that passed a basic comprehension test at the end of the experiment. Standard errors are robust to heteroskedasticity and shown in parentheses. *Significance levels:* \*10%, \*\*5%, \*\*\*1%.

Table B26: Redistribution regression interacting the treatment variable with merit dummy (subjects passed comprehension test at the end of experiment)

	(1)	(2)	(3)	(4)	(5)	(6)
	Redistr.	Redistr.	Redistr.	Redistr.	Redistr.	Redistr.
Motive group	0.0552 (0.244)	0.264 (0.299)	-0.0673 (0.340)	0.0835 (0.236)	0.180 (0.304)	0.0509 (0.340)
Meritocr.	-0.217 (0.163)	-0.126 (0.250)	-0.281 (0.215)	-0.180 (0.147)	-0.183 (0.227)	-0.206 (0.211)
Motive * meritocr-	-0.00600 (0.259)	-0.137 (0.330)	0.0778 (0.358)	-0.0366 (0.251)	-0.0984 (0.333)	-0.0333 (0.359)
Age				0.00636 (0.00409)	0.00923 (0.00679)	0.00427 (0.00513)
Female				-0.205** (0.0885)	-0.217 (0.150)	-0.189* (0.112)
Employed full time				-0.0927 (0.0936)	-0.0921 (0.147)	-0.0602 (0.120)
Student				0.0800 (0.113)	-0.00264 (0.167)	0.126 (0.151)
Risk aversion				-0.00963 (0.0203)	-0.0127 (0.0303)	-0.0119 (0.0273)
Patience				-0.00667 (0.0175)	-0.0571** (0.0256)	0.0222 (0.0236)
Right wing				0.0102 (0.109)	-0.107 (0.169)	0.0798 (0.149)
Constant	1.345*** (0.149)	1.308*** (0.231)	1.375*** (0.197)	1.319*** (0.239)	1.568*** (0.371)	1.220*** (0.320)
Signal	All	Disincentivizing	Incentivizing	All	Disincentivizing	Incentivizing
Education FE	No	No	No	Yes	Yes	Yes
Observations	413	154	259	409	153	256

**Note.** This table reports OLS regressions that interact the motive treatment variable with a dummy variable that indicates whether the subject self-identifies as a meritocrat (agrees with 7 out of 10 that subjects who transcribed more images should receive more money). Columns (1) and (4) report results for the whole sample; column (2) and (5) report results for a sample that only include subjects who received an incentivizing signal and column (3) and (6) reports results for a sample that only include subjects who received a disincentivizing signal. This table replicates Table B15 with subjects that passed a basic comprehension test at the end of the experiment. Standard errors are robust to heteroskedasticity and shown in parentheses. *Significance levels:* \*10%, \*\*5%, \*\*\*1%.

### 3.C Questionnaire

- Age
- Gender
- Country of living
- Nationality
- Level of Education
- Field of study
- Political orientation
- Risk aversion (Dohmen et al., 2011)
- Patience (Vischer et al., 2013)
- Question to characterize self-reported procrastination behavior: “I tend to put things off until later, although it would be better to do them right away.” Subjects then select one of the following items: *Absolutely not like me, Very little like me, Not really like me, A little like me, Very similar to me, Absolutely like me, I do not know*
- Question to characterize self-reported tendency to take measures to overcome self-control problems (situational): “I sometimes choose to not go to a gathering in the evening because if I go, I may stay longer than initially planned and end up being tired the next day.” Subjects then select one of the following items: *Absolutely not like me, Very little like me, Not really like me, A little like me, Very similar to me, Absolutely like me, I do not know*
- Question to characterize self-reported tendency to take measures to overcome self-control problems (general): “Generally speaking, to avoid putting things off, I often take measures that prevent me from doing so.” Subjects then select one of the following items: *Absolutely not like me, Very little like me, Not really like me, A little like me, Very similar to me, Absolutely like me, I do not know*
- Self-reported enjoyment of the task: “Did you enjoy transcribing images of blurry Greek letters?” (0 to 10 scale)
- Instruction recall: “The likelihood of receiving the bonus in Part 1 of the survey was independent of your performance if you were in the” (a) *Performance-Environment*, (b) *Chance-Environment*, (c) *None of the above*

### 3.D Attention Checks

The experiment administered two attention checks.



- **Attention Check 1:** In order to facilitate our research, we are interested in knowing certain factors about you. We are interested in whether you actually took the time to read the instructions; if not, then the data we collect based on your answers will be invalid. So, in order to demonstrate that you have read the instructions, please *ignore* the following question, and then write "I read the instructions" in the box labeled "other". Thank you very much.

The question read "What do you think is the right interval?" and subjects could choose between multiple intervals or (as instructed) enter something into the "other" box.

- **Attention Check 2:** Please write "the survey ended" into the box on the right:

Subjects could then either click on one of 10 buttons or, as instructed, write the text into the box.

## 3.E Pre-analysis plan

### Pre-Analysis Plan

#### Main Research Questions

This research project aims to answer the following research questions:

1. Do individuals distort beliefs towards believing that their effort is rewarded because they want to enhance their motivation to exert effort in the future (provision of a motive)? Are individuals that receive a signal that indicates that their effort is not rewarded more likely to distort beliefs than individuals that receives a signal that indicates that their effort is rewarded?
2. Are individuals that demand a commitment device more likely to distort beliefs? Are individuals that are bad at the task or dislike the task more likely to distort beliefs?
3. Does motivated belief distortion affect the willingness to redistribute income between two other individuals? Is this effect on redistributive behavior explained through the effect of the motive provision on the belief of being in a state of the world that rewards effort?

#### Experimental Design

The experiment is separated into four parts. Each part is defined by an activity that may have an effect on either the subjects' own payoff or the payoff of somebody else. At the end of the experiment, one part is randomly drawn on the individual level to become payoff relevant.

**Part 1** Subjects are informed that they will perform a real effort task where they transcribe a series of images containing 11 blurry Greek letters for 5 minutes. They are informed that they are matched with one randomly drawn subject (the competitor) that previously participated in the same study (they participated in a pilot study with 102 individuals). The way they are rewarded for this task is uncertain. With 50% chance the computer randomly assigned them either to the *Performance-Environment* or to the *Chance-Environment*. If they are in the *Performance-Environment*, they win a bonus of €7.50 with 80% (20%) chance if they transcribed more (fewer) images than their competitor. If they are in the *Chance-Environment*, they win the same bonus with 50% probability no matter whether they performed better or worse than their competitor. This uncertainty is not resolved. After a comprehension test, subjects carry on with the task.

**Part 2** If subjects are in the MOTIVE-GROUP (main treatment variation, 50% chance of being assigned) they are informed that they may redo the *same* transcription task in a later part of the survey. If they were previously assigned to the *Performance-Environment*, they are rewarded with a €5 bonus in the next task

if they transcribe more than 20 images in at most 10 minutes. If they are in the *Chance-Environment* they always get the additional bonus. Note that the environment agents face is the same as in Part 1. Subjects in the NO-MOTIVE-GROUP receive the identical information right before Part 4.

All subjects then receive two pieces of information: (1) whether they performed better than their competitor in Part 1 and (2) whether they won the bonus in Part 1 or not. The signal subjects get is my second treatment variation. They are subsequently asked to state their probabilistic belief (in %) about being in the *Performance-Environment*. Beliefs are incentivized using Karni (2009) with a possibility to win €2.50 and subjects can answer in intervals of 10.

**Part 3** All subjects make a redistributive decision. They are informed that they are paired with two subjects that previously participated in the same effort task as in Part 1. One of them received a bonus of €4, while the other one did not receive a bonus. The two individuals that they are paired with are in the same environment as themselves and the same rule (conditional on the environment) as in Part 1 applied to the two individuals. They are then asked whether they want to redistribute the bonus between the two individuals.

**Part 4** NO-MOTIVE-GROUP subjects are informed about the following task. All subjects then engage in doing the task.

**Questionnaire** All subjects finish the experiment with a questionnaire. The questionnaire first asks them to state how many images they plan to transcribe if they were re-invited to redo the task under a piece-rate scheme where they earn €0.25 per correctly transcribed image. It then asks them to specify a “minimum production level” where they will only be paid if they transcribed more images than specified by their “minimum production level”. They will actually be reinvited with 5% probability. Before the study concludes with a socio-demographic questionnaire, subjects are asked to recall their signal from Part 2.

### Relation between Research Questions and Design

The main treatment variation is whether a subject knows about the future task in Part 4 at the point of belief elicitation in Part 2. This provides the subject with the motive to distort beliefs in order to motivate future effort. Within treatment groups I exploit variation in the signal subjects observe. This allows me to test whether the effect differs by the type of signal subjects receive. Question 2 can be answered by testing for heterogeneity in the main treatment effect based on their propensity to choose a commitment device (Questionnaire), their performance in Part 1, or their self-declared taste for the task (Questionnaire). Finally, the third question is answered by comparing redistribution behavior in Part 3 across treatments.

## Sample

500 subjects are recruited via the online platform Prolific for the main experiment. 250 subjects are recruited to serve as recipients of the dictator game. Subjects of the main study receive €3 if they completed the experiment. A subject completed the experiment if they correctly answered two attention checks and if they completed a comprehension test at the beginning of the experiment in at most two tries. Subjects that are not correctly answering all questions after their second try are *not* allowed to complete the study. They receive €0.5. Subjects that are exceptionally fast (completed the study 3 standard deviations below the mean) are also excluded.

## Preliminaries

- Main outcomes of interest are (1) the beliefs elicited in Part 2 and (2) the amount redistributed in Part 3
- Main explanatory variables are (1) the treatment group (MOTIVE-GROUP is abbreviated with MG and NO-MOTIVE-GROUP is abbreviated with NG) and (2) whether the subject received a positive (S+) or a negative (S-) signal. A subject received a positive signal if she performed better than her competitor and won the bonus or transcribed fewer images and lost the bonus. A subject received a negative signal if she performed better than her competitor and lost the bonus or transcribed more images and won the bonus.
- Other explanatory variables that are used for the heterogeneity analysis include (1) whether the subject is more likely to self-identify as a meritocrat (dichotomized); (2) Part 1 performance decile of the subject; (3) whether she liked the task (dichotomized); (4) whether she commits herself to transcribe more than half of the images she plans to transcribe if she is reinvited.
- Variables used as control variables are (1) individual characteristics (age, gender, education, nationality, political orientation, risk aversion (Dohmen et al. (2011)), patience (Vischer et al. (2013))); (2) Part 1 performance decile of the subject; (3) winning the bonus; (4) performing better or worse than the competitor (5) false memory (dummy that a subject does not correctly recall her signal)

## Main Analysis

The main analysis of the paper is preceded by a descriptive section that plots histograms of beliefs stratified by the signal subjects receive.

The following part lists the analysis I want to undertake to answer the main research questions. Regressions will always use robust standard errors. Each bullet point commences with the outcome and treatment variable in bold.

- **Belief / MG** I first compare beliefs across MG and NG using a two-sided t-test. I then run a two-sided t-test on S+ and a one-sided t-test ( $H_0: \text{Belief} | \text{MG} \leq \text{Belief} | \text{NG}$ ) on S- subjects. I replicate these tests using regression analysis, where I regress elicited beliefs on MG and individual characteristics. Again, this regression is run separately for S- and S+.
- **Logit-Belief/MG,S+** I next compare whether subjects are more likely to distort beliefs relative to a Bayesian benchmark if they receive a negative rather than a positive signal. I regress logit-beliefs on the log odds-ratio of receiving a signal in the *Performance-* relative to the *Chance-Environment*. This variable is interacted with S+ and S-. I *additionally* interact the previously mentioned variables (log odds interacted with S+ and S-) with MG. The regression does not include an intercept. I test the  $H_0$  that the interaction with S+ and S- is equal to 1. I test the  $H_0$  that each interaction with MG is equal to 0. This coefficient informs how updating behavior differs across MG and NG subjects that received the same signal. Further, I test (t-test) the hypothesis that the magnitude of the coefficient that interacts the log odds ratio of a negative signal with MG is larger than the coefficient that interacts the log odds ratio of a positive signal with MG. The null hypothesis states the opposite.
- **Belief/MG** To test who distorts beliefs, I regress the belief on the treatment variable and on the interaction of the treatment variable with Part 1 performance decile, whether they liked the task and whether the subjects committed herself to work on the future task. I run this regression separately for subjects that received a positive or negative signal.
- **Redistribution/MG** To test whether subjects redistribute less if they are in MG, I test (t-test) whether the amount given to Worker B is different across MG and NG. I do this test separately for S+ and S-. To test whether the treatment effect differs across S+ and S- I run a regression where I interact the MG dummy with the S+ dummy. To assess the role of beliefs in explaining this variation, I conduct a mediation analysis. To that end, I regress redistribution on MG controlling for beliefs to compare this coefficient to the coefficient of a regression that omits the belief variable (inference is made using bootstrapped confidence-intervals). At last, I will interact MG with the variable that indicates whether a subject is a meritocrat. The last two analyses will also be run separately for S+ and S-.

## Exploratory Analysis

- **Memory** T-test that checks whether MG subjects are more likely to forget the signal they received. The test is done separately for S+ and S-.
- **Belief / Winning and Rank** I will ask whether there is an effect of winning the bonus on beliefs. I regress beliefs on a dummy that indicates winning controlling for effort decile fixed effects. I will run the same regression with rank as the main explanatory variable. The regression is replicated controlling for the informational content (S+) and including rank as well as winning as explanatory variables. Additionally I run the same regression interacting the explanatory variable of interest (rank or winning) with MG.
- **Redistribution / Winning and Rank** I will ask whether there is an effect of winning on distributive behavior. I regress redistribution on a dummy that indicates winning, controlling for effort decile fixed effects. I will run the same regression with rank as an explanatory variable. I will again perform mediation analysis to test whether the effects are explained by variation in beliefs.
- **Redistribution / heterogeneity** In case there is heterogeneity in belief distortion (see bullet point 3 in the last section), I will test for similar heterogeneity in the treatment effect on redistributive behavior.
- **Beliefs and redistributive behavior** To test for the role of beliefs in redistributive behavior I regress redistribution on beliefs, instrumenting beliefs by S+

## Robustness

To assess the robustness of my results, I replicate all regressions using the full set of controls (gender, political, orientation, age, level of education, country of residence, risk aversion and impatience). Furthermore, I will replicate all regressions where I control for Part 1 performance that uses performance decile fixed effects by controlling for Part 1 performance linearly.

Furthermore, I will test to what extent the signal, rank, and winning the bonus can be seen as exogenous. To that end I will run regressions where S+, rank, and winning the bonus is an outcome variable and the above-mentioned individual controls are the main explanatory variable. Any significance should disappear once controlling for performance either by decile fixed effects or linearly. Note that each regression has two explanatory variables (the characteristic and Part 1 performance)

At last I repeat the main analysis with a sample that answered correctly to a basic question about the treatment variation in the last questionnaire.

## 3.F Instructions

Figure B1: Screen 1 (all)

### Introduction

You are invited to participate in this online study. This is a project conducted by Max Lobeck, a PhD student at Paris School of Economics.

The results of this study will be used in a research project. It is therefore important that you **carefully read and follow the instructions**.

If you choose to participate in this study, you will be asked to do a simple task and answer questions about your opinions on social and economic issues. **The study should take around 30 minutes.**

In this study you have the chance to earn a bonus **in addition to the €3.00 (£ 2.7) you receive for completing** this study. **This bonus can be as high as €7.50.**

The survey is divided into parts. **At the end of the survey, one of these parts will be randomly selected by the computer.** Each part is equally likely to be selected.

**The earnings you made in this randomly selected part will be paid out at the end of the survey.** This means that each part is equally likely to determine how much you will earn in this study. It also means **that the amount you earn depends only on your decision in this randomly selected part of the survey.**

**This study does not use deception.** Everything we say in the instructions is implemented in that way. The study got ethical approval by the institutional review board of the Paris School of Economics (Number 2020-015).

The data gathered in this study is subject to EU GDPR regulation. It will only be used for research purposes.

All payments in the instructions are in Euro (€), the final amount will be paid out in British Pounds (£). We use the following exchange rate: €1 = £0.9.

If you have any questions or concerns, please contact Max Lobeck at max.lobeck@psemail.eu.

If you wish to participate in this study, please consent that you have read the above information and that you agree with participating in this study. Please state your Prolific ID so we can process your payment. On the last page, you will be given the completion code.

My Prolific ID is:

Consent:

I have read the above instructions and would like to participate in the study

Next

Figure B2: Screen 2 (all)

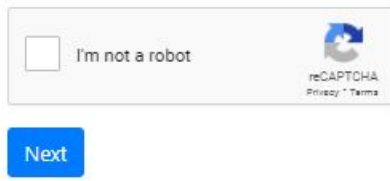




Figure B3: Screen 3 (all)

## Part 1 | What is the Task about?

The screen below describes the task you will perform during this part.

After you became familiar with the task, we will inform you how you are rewarded for the task.

### Content of the Task


In this task you will be presented with a series of images that contain 11 characters. The characters are either blurry Greek letters or points.

Your job is to transcribe as many of these images as possible in 5 minutes by clicking on the corresponding Greek letters in the bottom row.

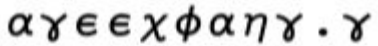
**This an example of a successful transcription task**

Time left to transcribe as many images as possible: 4:43

**Image:**



**Transcription:**



Recall that you have to transcribe every letter and every point. If you made a mistake you can correct it by clicking on the backwards pointing arrow.

---

**Successfully transcribed images: 0**

To submit your transcription and continue with the next image, please click on the button "Next image" below.

Next image

Even though the letters are blurry, they are depicted in the same way in all images you will face in the course of the task.

If you do not want to transcribe an image, you can always skip that image and continue with the **Next Image** button.

We will now familiarize you with the task by showing you two more examples and giving you the opportunity to transcribe them yourself. Please note that your performance during this practice phase will not affect any of your earnings.

Please proceed with the instructions by clicking on the "Next" button.

Next

Figure B4: Screen 4 (all)

Part 1 | Example of the task

You now have the opportunity to transcribe two images yourself to familiarize yourself with the task.

For the first example we give you the solution up front. By clicking on the letters in the bottom row, you can transcribe the image so that it matches the solution given to you. Note that you can correct your input by clicking on the left-pointing arrow.

Image :

Transcription:

ΧΦ

α β χ δ ε φ γ η λ . ✕

Please proceed with the instructions by clicking on the "Next" button.

Next

Figure B5: Screen 5 (all)

Part 1 | Another Example of the Task

In the second example you will transcribe another image. Here, we will not give you the solution up-front but you can verify your transcription after you tried it yourself.

Image:

Transcription:

βγαγ

α β χ δ ε φ γ η λ . ✕

[Click here to see the solution](#)

Please proceed with the instructions by clicking on the "Next" button.

Next

Figure B6: Screen 6 - first part (all)

## Part 1 | How are you rewarded for the Task?

### Competitor

After you completed the task, the computer will randomly choose your **competitor**.

This competitor is chosen from a pool of participants that just participated in this study on Prolific. These participants completed the same task for 5 minutes after receiving the same information about the task that you just received.

Your performance will be compared against the performance of your competitor. In case you transcribed the same number of images, you rank higher than your competitor if you completed the last correct transcription earlier.

### Environment

Before you start the task, the computer will randomly decide in which *Environment* you are performing the task:

- *Performance-Environment*
- *Chance-Environment*

The probability that the computer draws either *Environment* is equal to 50%. This means that it is equally likely that you are in the *Chance-Environment* or the *Performance-Environment*.

Thanks to your participation in the task you will have the chance to win a **bonus of €7.50**. The way you win depends on the environment in which you perform the task:

#### *Performance-Environment*

- The probability that you win **the bonus increases** with the number of images you transcribed correctly
- At the end of the task, the number of images that you transcribed will be compared **to the number of images your competitor transcribed correctly**. The probability that you win the bonus is:
  - Equal to **80%** (8 out of 10) if you transcribed **more images** than your competitor
  - Equal to **20%** (2 out of 10) if you transcribed **fewer images** than your competitor

Figure B7: Screen 6 - second part (all)

**Chance-Environment**

- The probability that you win **the bonus is independent** in the number of images you transcribed correctly
- The probability that you win the bonus is **always 50%** (5 out of 10)

After you completed the task, we will inform you whether you won the bonus and whether you correctly transcribed more or less images than your competitor.

Note that this information is informative about the *Environment* in which you actually completed the task.

For example: It is more likely that you completed the task in the *Performance-* rather than the *Chance-Environment* if you did not win the bonus and you transcribed fewer images than your competitor.

---

[Click here to continue with the comprehension questions](#)

Figure B8: Screen 6 - third part (all)

## Comprehension Questions

To make sure the instructions are clear enough, please answer the following questions. You may scroll up and re-read the instructions if you do not remember the answer.

**Note that you will not be able to proceed with the survey if you fail to answer all questions correctly after your *second* try.**

### Question 1

What is the initial probability in percent (%) that the computer randomly assigns you to the *Performance-Environment*?

### Question 2

Suppose that you are in the *Chance-Environment* and you transcribed *more* images than your competitor:  
Is the probability that you win the bonus *higher*, *lower* or *the same* than if you had transcribed *fewer* images than your competitor?

- The probability to win is higher
- The probability to win is lower
- The probability to win is the same

### Question 3

Suppose that you are in the *Chance-Environment*:

What is the probability in percent (%) that you win the bonus no matter how you performed relative to your competitor?

Figure B9: Screen 8 (all)

## Part 1 | Getting Ready

---

Now that you are familiar with the task and the payment rule, you will proceed with the actual task.

**Remember you have 5 minutes to transcribe as many images as possible.** On the top of your screen you will see how much time you have left.

---

Please click on the "Start" button to start the task.

Start

Figure B10: Screen 9 (all)

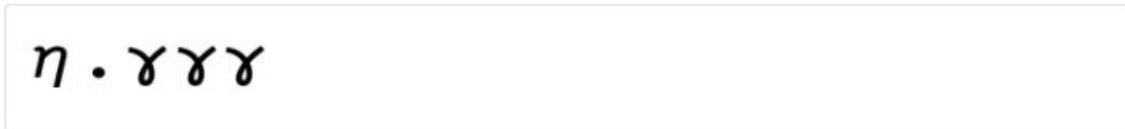
## Part 1 | Task

Time left to transcribe as many images as possible: 4:34

**Image:**



**Transcription:**



Recall that you have to transcribe every letter and every point. If you made a mistake you can correct it by clicking on the backwards pointing arrow.

Successfully transcribed images: 1

Your last transcription was correct

To submit your transcription and continue with the next image. Please click on the button "Next image" below.

Next image

Figure B11: Screen 10 (all)

## Part 1 | Task Summary

### Your performance in the task

Problem #	Your entry	Solution	Entry was correct
1	ieifihcgcb.	ieifihcgcb.	Yes
2	h.gggab.cbc	h.gggab.cbc	Yes
3	bfcg.cb.chh	bfcg.cb.chh	Yes
4	NA	giaeiaicec.	No

You have transcribed 3 problem(s) correctly.

Please click on the "Next" button to continue with the survey.

[Next](#)

Figure B12: Screen 11 (all)

## Part 1 | The End

You completed Part 1 of the survey. Please click on the "Next" button to proceed with the second part of the survey.

[Next](#)



Figure B13: Screen 21 (motive-group only)

## Part 4 | Task

In this part, you will participate to the **same transcription task** as in Part 1. **The environment in which you will perform this task is the one that was drawn at the beginning of Part 1.**

If you performed the task in the *Chance-Environment* in Part 1, then you will perform the next task in the *Chance-Environment*. If you performed the task in the *Performance-Environment* in Part 1, then you will perform the next task in the *Performance-Environment*.

The maximum duration of this part will be 10 minutes. Thanks to your participation in the task you will have the chance to win a **bonus of €5**. The way you win the bonus depends on the environment in which you perform the task:

### *Chance-Environment*

- You ***always*** win the bonus, no matter how many images you transcribed
- Your performance in the task does *not* affect whether you get the bonus or not

### *Performance-Environment*

- You **win the bonus** if you transcribe at least **20 images**
- You **do not win the bonus** if you transcribe fewer than **20 images**
- Your performance in the task does affect whether you get the bonus or not

Figure B14: Screen 22 (motive-group only)

### Comprehension Questions

To make sure the instructions are clear enough, please answer the following questions. You may scroll up and re-read the instructions if you do not remember the answer.

#### Question 1

What is the activity in the next task?

- Adding up numbers
- Solving CAPTCHAs
- Transcribing images of Greek letters
- Counting zeros

#### Question 2

Is the following statement true or false?

Suppose you are in the *Chance-Environment*:

You *always* win the second bonus even if you did not transcribe any images in the next task.

- True, even if I do not transcribe any images I win the bonus if I am in the *Chance-Environment*
- False, I have to transcribe 20 images to win the bonus if I am in the *Chance-Environment*

#### Question 3

Suppose you are in the *Performance-Environment*:

How many images do you have to transcribe to win the bonus?

#### Question 4

Assume that the *Performance-Environment* was drawn at the beginning of Part 1: What bonus-payment rule would apply to you for the next task?

- Performance-Environment's* bonus payment rule
- Chance-Environment's* bonus payment rule
- Neither

To submit your answers, please click on the "Next" button. The next page will inform you about the results of Task #1.

Next

Figure B15: Screen 12 (all)

## Part 2 | Results of the Task

---

This page informs you about the outcome of the Task you previously completed. Below you find whether you performed better or worse than your competitor and whether you won the bonus.

---

You **won** the bonus

---

You transcribed **fewer** images than your competitor

---

**Reminder:**

The first line indicates whether you won the bonus or not. If you performed the task in the *Performance-Environment*, the probability that you win the bonus is equal to 20% (2 out of 10) if you transcribed *fewer* images. If you performed the task in the *Chance-Environment*, the probability that you win the bonus is *always* equal to 50%.

The second line indicates whether you transcribed fewer or more images than your competitor.

---

To continue with the survey, please click on the "Next" button.

Next

Figure B16: Screen 13 (all)

## Part 2 | Your estimation of the probability that you are in the *Performance-Environment*

---

At the beginning of the survey, the computer decided in which environment you perform the task.

**In the next screen, you will be asked to report your estimation of the probability that you performed the task in the *Performance-Environment*.**

**You have the chance to win a bonus equal to €2.50 based on your estimation. Your chance to win this bonus is highest if your estimate is as accurate as possible.** It is in your best interest to state what you truly believe is the probability that you are in the *Performance-Environment*. If you want to learn more about the payment mechanism you can click [here](#) .

---

You can proceed by clicking on the "Next" button to provide your answer.

Next

Figure B17: Screen 14 (all)

## Part 2 | Your estimation

---

You can now enter your estimate by choosing one of the options below.

**I think the probability that I performed the task in the *Performance-Environment* is:**

0%  10%  20%  30%  40%  50%  60%  70%  80%  90%  100%

---

To submit your answer and proceed with the survey, please click on the "Next" button.

Next

Figure B18: Screen 15 (all)

## Part 2 | The End

---

You completed Part 2 of the survey. Please click on the "Next" button to proceed with the third part of the survey.

---

Next

Figure B19: Screen 16 (all)

## Part 3 | Distribution Decision

A few days ago two individuals, subsequently labelled Worker A and Worker B, were recruited via Prolific to conduct the same task as you (transcribing images of blurry Greek letters for 5 minutes).

**Worker A received a bonus of €4.00. Worker B did not receive the bonus.** Both workers receive a €1.00 completion reward.

**Both workers performed the task in the *same* environment (either *Chance* or *Performance*) as yourself.** As a reminder, you just stated that you completed the task in the *Performance-Environment* with 70% probability.

**The environment workers faced affects how their bonus was initially distributed.**

**The rule used to initially distribute this bonus is the same as the rule you faced in Part 1.** This means that the probability that Worker A wins the bonus is *independent* of his performance if the task is performed in the *Chance-Environment* and that it is *very likely* that Worker A performed better than Worker B if the task was performed in the *Performance-Environment*.

Please note that they are each other's competitors (Worker B is Worker A's competitor and Worker A is Worker B's competitor).

### Description of your decision

- In the next screen, you will be asked to decide about the final payoff of Workers A and B
- The workers were not told who won the bonus nor their environment.
- The workers know that a third person may redistribute this initial allocation of earnings. Your decision is completely anonymous.
- The workers will receive the payment that you choose on the next screen, but they will not receive any further information.

Figure B20: Screen 17 (all)

## Comprehension Questions

To make sure the instructions are clear enough, please answer the following questions. You may scroll up and re-read the instructions if you do not remember the answer.

### Question 1

Suppose you performed the task of Part 1 in the *Performance-Environment*:  
What is the environment of the two workers?

- Performance-Environment
- Chance-Environment
- Neither

### Question 2

Is the following statement True or False?

Suppose you performed the task of Part 1 in the *Performance-Environment*: In that case it is very likely that Worker A transcribed more images than Worker B

- True, it is very likely that Worker A transcribed more images than Worker B
- False, it is equally likely that Worker A transcribed more or less images than Worker B

Please submit your answers by clicking on the "Next" button. Note, that you will not be able to continue with this survey if you did not all answer the questions correctly after two attempts.

Next

Figure B21: Screen 18 (all)

## Part 3 | Distribution Decision

How do you want to distribute the bonus?

- Worker A is paid €4.00 and Worker B is paid €0.00
- Worker A is paid €3.00 and Worker B is paid €1.00
- Worker A is paid €2.00 and Worker B is paid €2.00
- Worker A is paid €1.00 and Worker B is paid €3.00
- Worker A is paid €0.00 and Worker B is paid €4.00

To review the instructions click [here](#)

To submit your answer and proceed with the survey, please click on the "Next" button.

Next

Figure B22: Screen 19 (all)

## Part 3 | Your Opinion

Before you move on to the next part, we want to ask for your opinion on the following question:

To what degree do you **agree** or **disagree** with the following statement:

**If two equally talented workers engage in an *identical* production task, the worker that produced *more units* should be rewarded with a *higher wage* than the worker that produced *fewer units*.**

Using this card, please place yourself on this scale, where **0** means '**totally disagree**' and **10** means '**totally agree**'. You can also use the values in-between to indicate where you fall on the scale.

totally disagree
totally agree  
 0  1  2  3  4  5  6  7  8  9  10

Please click on the "Next" button to continue with the survey.

Next

Figure B23: Screen 21 (no-motive-group only)

## Part 4 | Task

In this part, you will participate to the **same transcription task** as in Part 1. **The environment in which you will perform this task is the one that was drawn at the beginning of Part 1.**

If you performed the task in the *Chance-Environment* in Part 1, then you will perform the next task in the *Chance-Environment*. If you performed the task in the *Performance-Environment* in Part 1, then you will perform the next task in the *Performance-Environment*.

The maximum duration of this part will be 10 minutes. Thanks to your participation in the task you will have the chance to win a **bonus of €5**. The way you win the bonus depends on the environment in which you perform the task:

### *Chance-Environment*

- You **always** win the bonus, no matter how many images you transcribed
- Your performance in the task does *not* affect whether you get the bonus or not

### *Performance-Environment*

- You **win the bonus** if you transcribe at least **20 images**
- You **do not win the bonus** if you transcribe fewer than **20 images**
- Your performance in the task does affect whether you get the bonus or not



Figure B24: Screen 22 (no-motive-group only)

### Comprehension Questions

To make sure the instructions are clear enough, please answer the following questions. You may scroll up and re-read the instructions if you do not remember the answer.

#### Question 1

What is the activity in the next task?

- Adding up numbers
- Solving CAPTCHAs
- Transcribing images of Greek letters
- Counting zeros

#### Question 2

Is the following statement true or false?

Suppose you are in the *Chance-Environment*:

You *always* win the second bonus even if you did not transcribe any images in the next task.

- True, even if I do not transcribe any images I win the bonus if I am in the *Chance-Environment*
- False, I have to transcribe 20 images to win the bonus if I am in the *Chance-Environment*

#### Question 3

Suppose you are in the *Performance-Environment*:

How many images do you have to transcribe to win the bonus?

#### Question 4

Assume that the *Performance-Environment* was drawn at the beginning of Part 1: What bonus-payment rule would apply to you for the next task?

- Performance-Environment's* bonus payment rule
- Chance-Environment's* bonus payment rule
- Neither

To submit your answers, please click on the "Next" button. The next page will inform you about the results of Task #1.

Next

Figure B25: Screen 23 (all)

Part 4 | Getting Ready

Now that you are familiar with the payment rule, you will now proceed with the actual task.

Please click on the "Start" button to start the task.

Start

Figure B26: Screen 24 (all)

Part 4 | Task

Time left: 9:14

Image:



Transcription:

βαχάααεαε.β



Recall that you have to transcribe every letter and every point. If you made a mistake you can correct it by clicking on the backwards pointing arrow.

Successfully transcribed images: 0

Your last transcription was not correct

To submit your transcription and continue with the next image. Please click on the button "Next image" below.

To end the task and proceed with the survey, you can click on the button "End task" below. This decision is final.

Next image

End Task

Figure B27: Screen 25 (all)

Part 4 | Task Summary

Your performance in the task

Problem #	Your entry	Solution	Entry was correct
1	fhfbhhfiah	fhfbhhfiah.	No
2	bacaaaeae.b	bacaaaeae.b	Yes

You have transcribed 1 problem(s) correctly.

Please click on the "Next" button to continue with the survey.

Next

Figure B28: Screen 26 (all)

### Part 4 | The End

You completed Part 4 of the survey. Please click on the "Next" button to proceed with the questionnaire part of the survey.

Next

Figure B29: Screen 27 (all)

### Questionnaire 1 | Future Task

There is a 5% chance that you will be re-invited to do the same transcription task in 2 weeks.

Specifically, for each image you transcribe you will earn €0.25. This rule will apply to all participants and is independent of the current survey.

In this future task you will be able to transcribe up to 40 images. Hence, you can earn at most €10.00. If you would transcribe 20 images, you will earn €5.00, if you would transcribe 10 images, you will earn €2.50, etc.

In the next screen we will ask you to report how many images you *plan* to transcribe if you were reinvited.

- Note that your answer is not binding
- You will be able to transcribe more or less images in two weeks

By clicking on the "Next" button you can go to the next page and tell us how many images you *plan* to transcribe.

Next

Figure B30: Screen 28 (all)

### Questionnaire 1 | Future Task

You can now enter how many images you plan to transcribe if you were reinvited. You can only enter a whole number. The lowest number is 0 (I do not plan to transcribe any images if I were reinvited). The highest possible number is 40.

I plan to transcribe  images

To submit, please click on the "Next" button. To go back to the instructions, click on the "Back" button.

Back

Next

Figure B31: Screen 29 (all)

### Questionnaire 1 | Minimum Production Level

---

In the next screen we will ask you to **define a *minimum production level*** that applies to the task if you are reinvited.

- If your performance in the future task is **lower** than the minimum production level, your earnings will be **zero**
- If your performance in the future task is **higher** than the minimum production level, you will earn **€0.25** per image you correctly transcribed

**Example:** Assume you define a minimum production level of 7. If you do the task in two weeks and you transcribe:

- 4 images, then your performance is *lower* than the minimum production level, and **you earn €0.00**
- 8 images, then your performance is *higher* than the minimum production level, and **you earn  $8 \cdot €0.25 = €2.00$**
- 12 images, then your performance is *higher* than the minimum production level, and **you earn  $12 \cdot €0.25 = €3.00$**

---

By clicking on the "Next" button you can go to the next page.

[Next](#)

Figure B32: Screen 30 (all)

### Questionnaire 1 | Minimum Production Level

---

You can now enter your *minimum production level*. You can only enter a whole number. The lowest number is 0 (no minimum production level). The highest possible number is 40.

I choose a minimum production level of:

images

---

To submit your *minimum production level* click on the "Next" button. To go back to the instructions, click on the "Back" button.

[Back](#) [Next](#)

Figure B33: Screen 31 (all)

## Questionnaire 2 | Part 1 Outcome

In Part 1, we asked you to do the transcription task for 5 minutes. After the task you received two pieces of information: (1) whether you transcribed more images than your competitor and (2) whether you won a bonus or not.

Please repeat this information to us.

Did you transcribe more or fewer images than your competitor?

- More** (My performance was higher)
- Fewer** (My performance was lower)
- I do not recall**

Please indicate whether you won the bonus for the task of Part 1 or not

- Yes**, I won the bonus after Task #1
- No**, I did not win the the bonus after Task #1
- I do not recall**

Please submit your answer by clicking on the "Next" button to proceed with the survey.

Next

Figure B34: Screen 32 (all)

## Questionnaire 3

In order to facilitate our research, we are interested in knowing certain factors about you. We are interested in whether you actually took the time to read the instructions; if not, then the data we collect based on your answers will be invalid.

So, in order to demonstrate that you have read the instructions, please *ignore* the following question, and then write "I read the instructions" in the box labeled "other". Thank you very much.

What do you think is the right interval?

- 0-20
- 21-40
- 41-60
- 61-80
- 81-100

Other:

Please submit your answer by clicking on the "Next" button to proceed with the survey.

Next

Figure B35: Screen 33 (all)

Part 2 | Questionnaire 4

**Question 1**  
What is your age?

**Question 2**  
Please select your gender:  
 Female  
 Male  
 Other

**Question 3**  
In public settings, people sometimes talk of "the left" and "the right".  
Using this card, where would you place yourself on this scale, where 0 means "left" and 10 means "right"?  
0 1 2 3 4 5 6 7 8 9 10  
left right  
 (Default value: 0)

**Question 4**  
Please state your Nationality (if it applies, list multiple):

**Question 5**  
Please state the country you are currently living in:

**Question 6**  
Please state the highest degree you hold:  
 High School  
 Undergraduate Degree  
 Graduate Degree  
 Vocational Degree

**Question 7**  
Please state your field of studies:

Please click on the "Next" button to proceed with the survey.

[Next](#)

Figure B36: Screen 34 (all)

**Questionnaire 5**

---

**Question 1**  
 How do you see yourself?  
 Are you generally an impatient person, or someone who always shows great patience?  
 Using this scale (shown below) on this scale where 1 means 'very impatient' and 10 means 'very patient'. Please circle the values in brackets to indicate where you fall on the scale.

very impatient very patient  
 1 2 3 4 5 6 7 8 9 10 (10) (1) (2) (3) (4) (5) (6) (7) (8) (9)

---

**Question 2**  
 How do you see yourself?  
 Are you a person who is generally willing to take risks, or do you try to avoid taking risks?  
 Using this scale (shown below) on this scale where 1 means 'completely unwilling to take risks' and 10 means 'very willing to take risks'. You can also use the values in brackets to indicate where you fall on the scale.

completely unwilling to take risk very willing to take risk  
 1 2 3 4 5 6 7 8 9 10 (1) (2) (3) (4) (5) (6) (7) (8) (9) (10)

---

**Question 3**  
 Please select the statement that describes you best.  
 I tend to put things off until later, although it would be better to do them right away.

- Absolutely not like me
- Very little like me
- Not really like me
- A little like me
- Very similar to me
- Absolutely like me
- I do not know

---

**Question 4**  
 Please select the statement that describes you best.  
 I sometimes choose to not go to a gathering in the evening because if I go, I may stay longer than I initially planned and end up being tired the next day.

- Absolutely not like me
- Very little like me
- Not really like me
- A little like me
- Very similar to me
- Absolutely like me
- I do not know

---

**Question 5**  
 Please select the statement that describes you best.  
 Generally speaking, to avoid putting things off, I often take measures that prevent me from doing so.

- Absolutely not like me
- Very little like me
- Not really like me
- A little like me
- Very similar to me
- Absolutely like me
- I do not know

---

**Question 6**  
 Did you enjoy viewing images of Mary Givok Island?  
 Using this scale (shown below) on this scale where 1 means 'did not enjoy at all' and 10 means 'enjoyed a lot'. You can also use the values in brackets to indicate where you fall on the scale.

did not enjoy at all enjoyed a lot  
 1 2 3 4 5 6 7 8 9 10 (1) (2) (3) (4) (5) (6) (7) (8) (9) (10)

---

**Question 7**  
 The likelihood of receiving the bonus in Part 1 of the survey was *indeterminate* of your performance if you were in the

- Performance Performance
- Chance Performance
- None of the above

---

**Question 8**  
 Please write the "the survey ended" into the box on the right.

1 2 3 4 5 6 7 8 9 10 Done

---

Please click on the "Next" button to submit the answers and proceed to the last page of this survey to answer the completion code.

Next

Figure B37: Screen 35 (all)

## Final Page

---

You have finished the survey. Thank you for your participation.

This survey has four parts. As mentioned at the start of the survey, we will draw one of these parts randomly and this decision will determine the bonus that is actually paid out.

Below we list the outcome for each of the four parts:

- You got the bonus for the task in Part 1
- You did not win the bonus for your estimation of the likelihood in Part 2
- You decided that Worker A receives €1 and Worker B receives €3 in Part 3
- You got the bonus for the task in Part 4

The computer randomly selected the bonus for the estimation of the likelihood in Part 2 to be the part that becomes payoff relevant.

The amount of €0.0 will be transferred to you in addition to the fixed completion fee of €3.00 in the coming days.

The computer randomly decided that you will not be reinvited to do the task within the next two weeks.

**The prolific completion link is: <https://app.prolific.co/submissions/complete?cc=8C8F626A>**

If you have any comments, feel free to contact Max Lobeck at [max.lobeck@psemail.eu](mailto:max.lobeck@psemail.eu).

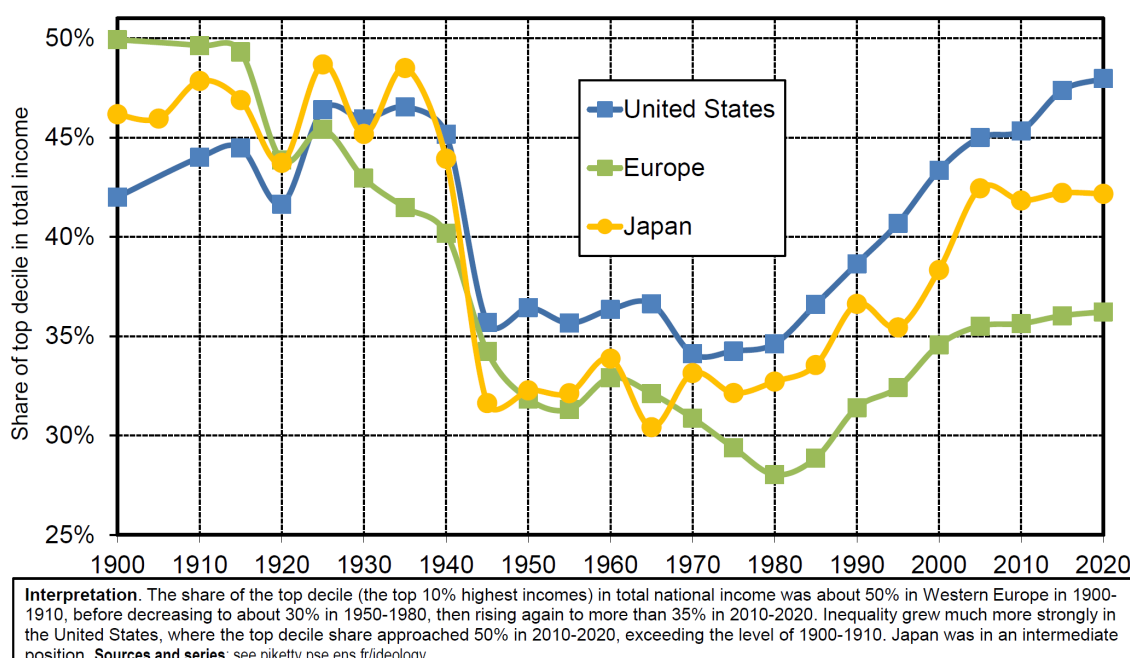
---



# Résumé détaillé en français

## 1 Motivation

FIGURE 4.1 : Inégalité de 1900 à 2020 en Europe Ouest, Japon et des États-Unis. Source : Piketty (2020)



Les inégalités économiques ont augmenté de façon spectaculaire dans le monde entier : alors que la part des revenus des 10 % les plus aisés aux États-Unis était d'environ 35 % en 1980, elle atteindra près de 50 % en 2020. Comme le montre la figure 4.1, l'inégalité de revenus a atteint une dimension qui n'avait pas été vue depuis la Seconde guerre mondiale. Le regain d'inégalité illustré n'est pas confiné aux États-Unis mais caractérise une dynamique mondiale. L'inégalité n'a cessé d'augmenter en Europe, au Japon, en Chine, en Russie et, en particulier, en Inde, où les 10 % les plus aisés s'approprient désormais plus de 55 % du revenu total.

La révélation de cette dynamique frappante de l'inégalité économique a placé l'inégalité au centre du débat universitaire et politique.<sup>1</sup> Le président Barack Obama a qualifié l'inégalité économique de "défi majeur de notre époque" (Obama, 2013). Ce débat avait, et a toujours, de

<sup>1</sup>Cela a été possible grâce à la diffusion de données nouvellement disponibles sur les parts des revenus des plus aisés dans des études universitaires. Ainsi, telles que Atkinson et al. (2011), Alvaredo et al. (2013) et Piketty (2014).

multiples facettes : il s'interroge sur les origines et les conséquences des inégalités économiques, et pose la question de savoir si ces inégalités sont justifiées en premier lieu. Les attitudes des individus envers l'inégalité sont fondamentales pour répondre à chacune de ces questions et l'objectif de cette thèse est donc de faire progresser notre compréhension de ces attitudes.

**Pourquoi devrions-nous étudier les attitudes à l'égard de l'inégalité ?** L'inégalité est fondamentalement un *choix* de société. Les sociétés ont les moyens de réduire l'inégalité en redistribuant les revenus, par exemple en prélevant des impôts qui augmentent progressivement en fonction du revenu.<sup>2</sup> Les attitudes envers l'inégalité sont essentielles pour comprendre pourquoi elles font ce choix. Lorsque les gouvernements définissent ces politiques, ils sont influencés—au moins dans une certaine mesure—par les préférences et les demandes de leurs citoyens. Il est donc indispensable de comprendre comment les individus se forgent des attitudes à l'égard des inégalités s'il on veut mieux comprendre la dynamique de l'inégalité économique elle-même. À titre d'exemple, la révolution conservatrice des années 1980 a popularisé le sentiment selon lequel des niveaux élevés de redistribution entravent la croissance économique. Cela a renforcé le discours méritocratique selon lequel les riches gagnent leurs hauts revenus grâce à un travail acharné, justifiant ainsi l'inégalité avant impôt (Piketty, 2020; Sandel, 2020). Ces récits et ces croyances ont façonné la politique économique des dernières décennies et ont conduit au démantèlement de l'État providence dans les économies occidentales, même par les gouvernements de centre-gauche.

Si cette dynamique est en partie due à l'offre d'idées dans les années 1990 et au début des années 2000, elle est également le produit d'une demande populaire en faveur des niveaux de redistribution plus faibles.<sup>3</sup> Une caractéristique frappante de cette période est que, du moins aux États-Unis, l'augmentation des inégalités de revenus ne s'est pas accompagnée d'une demande accrue de redistribution (Ashok et al., 2015); ce qui peut expliquer en partie pourquoi nous n'avons pas assisté à des efforts plus importants pour réduire les inégalités. Cela illustre donc l'importance de comprendre les *attitudes* envers l'inégalité afin de mieux comprendre l'*évolution* de celle-ci.

L'étude des attitudes à l'égard l'inégalité est, en outre, essentielle pour comprendre les conséquences de l'inégalité. Le fait de supposer que les individus se soucient de l'inégalité implique que celle-ci affecte leur utilité et leur comportement. Les attitudes envers l'inégalité peuvent également affecter les résultats et le comportement sur le marché du travail : si les individus n'aiment pas travailler dans une organisation où les niveaux d'inégalité sont élevés, l'inégalité a une incidence sur le fonctionnement des contrats d'incitation, le bien-être au travail et l'offre de main-d'œuvre.

---

<sup>2</sup>L'importance des politiques pour le niveau d'inégalité est étayée par l'observation selon laquelle la dynamique de l'inégalité est corrélée au taux marginal supérieur d'imposition, une mesure de la progressivité de l'impôt (Piketty, 2014; Saez and Zucman, 2020).

<sup>3</sup>L'évaluation de l'effet causal des préférences des citoyens en matière de redistribution et des niveaux de redistribution mis en œuvre est difficile. Cela est dû en partie au fait que les attitudes envers l'inégalité ne sont pas les seuls éléments qui comptent lors des élections. Il peut donc être difficile de dissocier l'effet des opinions sur l'inégalité des autres objectifs politiques qui peuvent figurer sur le bulletin de vote. Néanmoins, des enquêtes ont montré que les électeurs se soucient de l'équité distributive lors des élections (voir Almås et al., 2020, pour les États-Unis et la Norvège).

Indépendamment du fait que les attitudes envers l'inégalité sont importantes pour mieux comprendre la dynamique et les conséquences de l'inégalité, leur nature en fait un sujet d'étude fascinant en soi. Cela s'explique principalement par le fait que les attitudes à l'égard de l'inégalité (i) sont loin d'être homogènes, (ii) dépendent de l'importance de motifs concurrents qui peuvent ou non entrer en conflit les uns avec les autres, (iii) sont sujettes à des biais comportementaux et (iv) dépendent du contexte.

**Principales questions de recherche** Bien que la thèse comprenne trois documents de recherche autonomes, ils répondent tous à une problématique globale : *Comment l'environnement économique façonne-t-il les attitudes envers l'inégalité ?* Comme je le montrerai dans la section 2, la littérature antérieure a fait de grands progrès dans l'identification des raisons pour lesquelles les individus peuvent s'opposer l'inégalité. Cependant, notre compréhension de la manière dont ces motifs interagissent avec l'environnement économique et institutionnel est encore limitée. Il est crucial de faire progresser nos connaissances dans cette direction, car elles sont essentielles pour mieux comprendre la dynamique de l'acceptation de l'inégalité et la manière dont les politiques façonnent les attitudes envers celle-ci.

Le chapitre 1 value systématiquement la manière dont les préférences en matière de distribution des revenus dépendent du contexte dans lequel elles sont révélées. À cette fin, il étudiera conjointement comment les attitudes à l'égard de l'inégalité dépendent (1) de la *forme* relative que prennent les inégalités, (2) de la *certitude* d'occuper une position donnée au sein d'une distribution des revenus, et (3) du fait que cette position est déterminée par des différences de *chance* ou des différences de *effort*. Les résultats permettent de mieux comprendre dans quels contextes les individus peuvent accepter ou rejeter l'inégalité. Par exemple, les grandes inégalités qui ne favorisent que les plus hauts placés obtiennent un suffrage unanime si tous les individus bénéficient faiblement de ces inégalités et si tous les individus ont une chance de bénéficier de ces inégalités. Cette acceptation de l'inégalité est indépendante de la manière dont on a obtenu ce rang. Cependant, l'unanimité se brise dès que les positions dans la distribution sont fixes et que certains peuvent ne pas bénéficier des inégalités. Alors, une minorité non négligeable de ces sujets rejette ces inégalités en diminuant les revenus au sommet de la distribution *sans* redistribuer cet argent. Cela illustre l'importance des perspectives de revenu dans la formation des attitudes à l'égard des inégalités.

Le chapitre 2 étudie les attitudes des managers envers l'inégalité dans le contexte de l'entreprise. La problématique est de savoir si les attitudes des dirigeants à l'égard de l'inégalité influent sur l'offre d'incitations dans les entreprises et si elle est influencée par le système d'incitations auquel un dirigeant est lui-même confronté. Dans cet article, nous constatons que l'attitude des dirigeants à l'égard de l'inégalité affecte le choix des incitations. Nous montrons également que les préoccupations distributives sont partiellement, mais pas totalement, évincées par les incitations auxquelles un manager est lui-même confronté. Cela implique que l'effet des préférences distributives des managers sur les systèmes d'incitation persiste, même si le manager est incité à agir contre sa vision d'équité. Cette étude est liée à la problématique primaire de cette thèse

en appliquant la pertinence des attitudes d'inégalité au contexte de l'entreprise et en testant comment elles dépendent de l'environnement incitatif auquel la décideuse est confrontée.

Le chapitre 3 étudie un déterminant crucial des attitudes envers l'inégalité : les croyances concernant la source de l'inégalité. Comme je le montrerai plus loin, les individus sont plus enclins à accepter l'inégalité s'ils croient qu'elle reflète des différences dans l'effort plutôt que la chance. Dans ce chapitre, je me demande si ces croyances sont déformées pour motiver l'effort. Par ce biais, je teste si les croyances chance-effort sont endogènes à la structure d'incitation à laquelle les individus sont confrontés. L'analyse révèle que les croyances relatives à l'effort et à la chance sont effectivement déformées à des fins de motivation et sont donc affectées par les structures d'incitation auxquelles les individus s'attendent à être confrontés. Ce type de distorsion des croyances conduit à une surestimation de l'importance de l'effort par rapport à la chance pour avoir de la réussite.

Le reste de l'introduction est structuré comme suit : la section 2 présente une étude de la littérature existante sur les attitudes à l'égard de l'inégalité ; la section 3 décrit les méthodes empiriques que j'utilise pour répondre aux problématiques ; et la section 4 présente un résumé détaillé de chaque chapitre.

## 2 Revue de la littérature

Cette section fournit une revue de la littérature sur les attitudes envers l'inégalité qui aidera le lecteur à placer les contributions de cette thèse dans le contexte plus large de la littérature existante. La revue commence par la question de savoir pourquoi et si les individus ont des préférences en matière de distribution de revenus. Je résumerai ensuite les preuves existantes sur la manière dont ces attitudes à l'égard de l'inégalité sont façonnées par les expériences et l'environnement économique, avant de conclure la revue par une présentation des *conséquences* des attitudes envers l'inégalité.

### 2.1 Raisons des attitudes envers l'inégalité

Tout au long de cette thèse, je soutiendrai que les individus se soucient de l'inégalité économique. On peut donc se demander pourquoi les individus devraient d'abord avoir des préférences en matière de distribution de revenus. Dans la section suivante, je décrirai les concepts développés ces dernières années et j'indiquerai les preuves en faveur ou à l'encontre de ces concepts théoriques.

#### Raisons égocentriques envers l'inégalités

Les ouvrages qui s'interrogent sur les raisons de nos attitudes à l'égard de l'inégalité se concentrent principalement sur les préférences non-égoïste. Néanmoins, il existe plusieurs situations dans lesquelles les individus peuvent préférer certaines distributions de revenus à d'autres pour des raisons totalement égoïstes.

L'une des raisons pour lesquelles les individus veulent réduire les inégalités est qu'ils s'attendent à bénéficier de la réduction de ces inégalités par la redistribution des revenus. Cette

idée est particulièrement présente dans les modèles d'économie politique qui s'appuient sur le cadre introduit par Romer (1975) et Meltzer and Richard (1981). Les études empiriques qui s'appuient sur des enquêtes mettent souvent en évidence une corrélation négative entre la demande de redistribution et le revenu actuel ou attendu par des ménages (voir par exemple Alesina and La Ferrara, 2005; Alesina and Giuliano, 2011; Owens and Pedulla, 2014). Les expériences de laboratoire qui attribuent de manière exogène des niveaux initiaux de revenu constatent souvent qu'en moyenne, les individus attribuent un poids positif à la partie égoïste de leur fonction d'utilité (par exemple Durante et al., 2014). En outre, une minorité non négligeable (16,75 %) de la population adulte des États-Unis s'est avérée être complètement égoïste, ce qui signifie qu'elle choisit toujours l'option qui maximise son revenu (Fisman et al., 2017).<sup>4</sup> Le lien entre la maximisation de son revenu et les attitudes d'inégalité peut sembler évident à première vue. Cependant, l'environnement économique influence la manière dont les motifs égoïstes se traduisent en attitudes d'inégalité, ce qui donne souvent des résultats surprenants à première vue. Je reviendrai sur cette littérature dans la section 2.2, lorsque j'examinerai comment l'incertitude façonne les attitudes vis-à-vis l'inégalité.

### Préférences non-égoïstes sur la distribution des revenus

Si les raisons égoïstes de redistribution peuvent constituer une motivation puissante pour expliquer pourquoi les individus rejettent ou acceptent les inégalités, il existe de nombreux cas où cela ne suffit pas à rationaliser le comportement. Une façon de conceptualiser cette divergence consiste à supposer que les individus se soucient intrinsèquement de la distribution dominante des revenus.

**Conceptualiser la justice distributive** Pour placer les attitudes vis-à-vis l'inégalité dans un cadre commun, je suppose, dans l'esprit de Cappelen et al. (2013), que si la distribution dominante des revenus s'écarte de la distribution des revenus qu'un individu considère juste, il subit un coût moral. Celui-ci peut être caractérisé par la fonction de valeur suivante d'un individu donné  $i$  :

$$V_i(y_i, \mathbf{y}) = y_i - \alpha_i M(|D(\mathbf{y}) - D_i^F(\mathbf{y})|) \quad (4.1)$$

où  $y_i$  est le revenu propre de l'individu  $i$ ,  $\mathbf{y}$  est un vecteur de revenus qui peut inclure ou non  $y_i$ ,  $D(\mathbf{y})$  est la distribution dominante des revenus inclus dans  $\mathbf{y}$ ,  $D^F(\mathbf{y})$  est la distribution des revenus que  $i$  considère comme la distribution *fair* dans l'ensemble des distributions de revenus réalisables, désignée par  $\Omega$ , et  $M(\cdot)$  est le coût moral qui se produit si  $D(\mathbf{y}) \neq D_i^F(\mathbf{y})$  et qui est croissant et faiblement convexe dans la différence des deux. Enfin,  $\alpha_i$  caractérise l'importance relative de la maximisation de son propre revenu et de la minimisation de  $M(\cdot)$ .

Avant de continuer, je tiens à préciser comment interpréter cette fonction de valeur très

<sup>4</sup>L'expérience utilisée par Fisman et al. (2017) utilise un jeu de dictateur modifié qui fait varier le coût du don, ainsi que le compromis efficacité-égalité entre les choix faits par les sujets. Des décisions répétées permettent aux auteurs d'identifier l'égoïsme des dictateurs et leur volonté d'arbitrer entre égalité et efficacité (voir ci-dessous).

générale. Premièrement,  $D_i^F(\mathbf{y})$  dépend des alternatives et des ressources disponibles (c'est-à-dire  $\Omega$ ). Deuxièmement,  $D_i^F(\mathbf{y})$  peut dépendre du contexte. Par exemple, les individus peuvent avoir un regard différent selon qu'il s'agisse de ce qui est juste au lieu de travail ou bien de la redistribution des revenus au niveau de la société. Cela implique que  $\mathbf{y}$  peut inclure les gains de tous les individus de la société ou d'un sous-ensemble d'individus, par exemple ceux sur le lieu de travail. Troisièmement,  $D_i^F(\mathbf{y})$  peut être différent si votre revenu est en jeu lorsque vous décidez parmi les distributions de revenus. Quatrièmement,  $D_i^F(\mathbf{y})$  peut différer d'un individu à l'autre, un sujet que j'aborderai ci-dessous lorsque je discuterai de l'hétérogénéité des idéaux d'équité.

**Qu'est-ce qui entre dans  $D^F(\mathbf{y})$  ?** Les chapitres de ma thèse s'appuient sur des approches théoriques qui renseignent sur la forme de  $D^F(\mathbf{y})$ . Il s'agit notamment des théories normatives de la justice distributive développées par les philosophes moraux (voir Konow, 2003, pour une revue extensive) ainsi que des approches caractérisées par les modèles de préférences sociales développés par les économistes comportementaux (voir Cooper and Kagel, 2016, pour une revue extensive).<sup>5</sup>

L'une des raisons pour lesquelles les individus peuvent préférer certaines distributions de revenus à d'autres est qu'ils sont intéressés par la maximisation du bien-être social. L'*Utilitarisme* est la théorie la plus importante dans cet esprit. Introduite par Bentham (1789) et développée plus tard par Sen (1979) dans l'approche "welfariste", elle soutient que le revenu doit être distribué de manière à maximiser la somme des utilités. En d'autres termes, l'utilitarisme stipule que la distribution des revenus doit viser à maximiser la somme du bien-être ou du bonheur individuel.

$$D^F(\mathbf{y}) = \max_{D(\mathbf{y}) \in \Omega} \left\{ \sum_{k=1}^N U_i(y_k) \right\} \quad (4.2)$$

Au niveau sociétal, les arguments utilitaires peuvent conduire à une demande de réduction des fortes inégalités si l'on part du principe que l'utilité marginale diminue avec le revenu (Dalton, 1920). Étant donné la nature subjective de l'utilité et l'absence d'une métrique quantifiable pour l'utilité cardinale, il est relativement difficile d'identifier des préférences distributives qui suivent le principe utilitaire. Cela s'explique par deux raisons : premièrement, ceux qui doivent prendre ces décisions, doivent se forger des croyances sur les préférences de ceux qui sont affectés par la décision de distribution ; deuxièmement, même si les individus pouvaient soumettre leurs préférences cardinales, ils sont fortement incités à tromper l'organisme qui décide des politiques de distribution.<sup>6</sup>

Dans la littérature sur les préférences sociales, le motif d'efficacité a reçu plus d'attention

<sup>5</sup>Dans cette revue de la littérature, je ne discuterai pas des approches basées sur la réciprocité, qui sont très influentes dans la littérature sur les préférences sociales. J'ai pris cette décision car les chapitres de cette thèse ne parlent pas de cette littérature. Néanmoins, je pense que la réciprocité est un puissant facteur de motivation pour accepter les inégalités.

<sup>6</sup>Il existe toutefois des études plus anciennes qui utilisent des scénarios hypothétiques et qui trouvent un certain soutien aux points de vue d'équité basés sur la comparaison des utilités plutôt que des revenus (Yaari and Bar-Hillel, 1984; Konow, 1996, 2003).

qu'un motif véritablement utilitaire. Alors que l'utilitarisme stipule qu'il faut maximiser la somme des utilités, le motif d'efficacité s'intéresse à la maximisation de la somme des revenus.<sup>7</sup>

$$D^F(\mathbf{y}) = \max_{D(\mathbf{y}) \in \Omega} \left\{ \sum_{k=1}^N y_k \right\}. \quad (4.3)$$

L'efficacité en tant que motif qui informe les préférences en matière de répartition des revenus est intéressante à étudier, car il a des conséquences directes sur la politique. En définitive, l'impôt progressif est l'une des politiques les plus efficaces pour réduire les inégalités. Étant donné que les impôts ont un effet de distorsion sur l'offre de travail, une partie de la production est perdue et le gâteau total se rétrécit. Cette tension crée un conflit entre l'efficacité et le motif égalitaire, qui a été au centre de nombreuses études sur les attitudes à l'égard de l'inégalité (voir, par exemple, Okun, 1975).

Alors que les motifs ci-dessus soutiennent que l'on doit mettre en œuvre la distribution des revenus qui est la moins "gaspilleuse", le concept opposé de justice est l'*égalité*. Tout comme les préoccupations relatives à l'efficacité, le principe d'équité égalitaire considère les résultats plutôt que les processus : il rejette les distributions de revenus qui sont inégales, quelle que soit la manière dont elles se sont produites (Nielsen, 1985). Dans la littérature, deux approches théoriques conduisent à des motifs égalitaires pour la redistribution. La première est une approche désintéressée ou normative qui rejette l'inégalité sans tenir compte de la position de chacun dans la distribution des revenus. Ici, la fonction objective caractérise un désir de minimiser les différences de revenus :

$$D^F(\mathbf{y}) = \min_{D(\mathbf{y}) \in \Omega} \left\{ \sum_{k=1}^N |y_k - \tilde{y}| \right\}, \quad (4.4)$$

où  $\tilde{y}$  est le revenu médian.

L'autre approche, qui a été particulièrement importante dans la littérature sur la préférence sociale, utilise une approche comparative et suppose que les individus comparent leur revenu ou leur consommation à ceux de leurs pairs (Veblen, 1899; Duesenberry, 1949). L'article majeur de Fehr and Schmidt (1999) développe un cadre dans lequel les individus souffrent d'une perte d'utilité en ayant un revenu plus ou moins important que d'autres individus. La fonction d'utilité qu'ils proposent distingue les préoccupations liées au fait d'être derrière d'autres individus (envie) et d'être devant d'autres individus (aversion pour l'avance (Bruhin et al., 2018)) dans la distribution des revenus :

$$U_i(\mathbf{y}) = y_i - \alpha_i \frac{1}{N-1} \sum_{i \neq j} \max\{y_j - y_i, 0\} - \beta_i \frac{1}{N-1} \sum_{i \neq j} \max\{y_i - y_j, 0\}. \quad (4.5)$$

Une hypothèse clé des auteurs est que les individus sont plus envieux qu'aversion à l'avance,

<sup>7</sup>On peut faire valoir que ces deux concepts vont de pair pour de faibles quantités de revenus, comme c'est souvent le cas dans les expériences en laboratoire, où l'on peut supposer que l'utilité est linéaire en fonction du revenu. Mon point de vue est que la distinction entre l'utilitarisme et les motifs d'efficacité n'est pas toujours clairement communiquée dans la littérature sur les préférences sociales.

c'est-à-dire  $\alpha_i \geq \beta_i$ . Comment intégrer l'aversion pour l'inégalité dans l'esprit de Fehr and Schmidt dans le cadre esquissé dans (4.1) ? Une façon de procéder serait de supposer que  $D^F(\mathbf{y})$  suit (4.4). Cependant, le poids que l'agent accorde au terme moral dans (4.1),  $\alpha$ , diffère selon qu'il se situe au-dessus ou au-dessous de la médiane de la distribution des revenus.<sup>8</sup>

Alors que l'égalitarisme considère que toutes les inégalités sont injustes, le *principe de différence*, formulé par Rawls (1971), soutient que l'on devrait rejeter les inégalités à moins qu'elles n'aident les plus démunis. Les préférences en matière de répartition des revenus suivent alors une règle de maximin.

$$D^F(\mathbf{y}) = \max_{D(\mathbf{y}) \in \Omega} \left\{ \min\{y_1, \dots, y_N\} \right\} \quad (4.6)$$

L'identification des préférences rawlsiennes s'avère difficile. La théorie soutient que les individus décideront d'une telle distribution des revenus lorsqu'ils choisiront derrière le voile de l'ignorance, et tente de recréer les conditions caractérisées par la faible demande de justice rawlsienne démontrée par Rawls (Michelbach et al., 2003; Frohlich et al., 1987).

Néanmoins, l'idée que les inégalités sont acceptables si elles profitent au membre le plus démuné du groupe est reprise par le modèle développé par Charness and Rabin (2002). Dans ce modèle, les individus se soucient du revenu du plus démuné en même temps que du gain total. Cela se traduirait par la distribution préférée des revenus suivante, avec  $\gamma$  caractérisant le poids accordé au revenu du plus démuné par rapport à l'efficacité.

$$D^F(\mathbf{y}) = \max_{D(\mathbf{y}) \in \Omega} \left\{ \gamma \min\{y_1, \dots, y_N\} + (1 - \gamma) \sum_{k=1}^N y_k \right\}. \quad (4.7)$$

Tout comme Charness and Rabin (2002), plusieurs approches théoriques visent à combiner les concepts énumérés ci-dessus dans un cadre unique. Cette approche est particulièrement utile pour la caractérisation empirique des attitudes envers l'inégalité, car elle permet d'identifier avec parcimonie les préoccupations en matière d'inégalité et d'efficacité. Par exemple, Bruhin et al. (2018) propose et estime un modèle qui tient compte de l'envie, de l'aversion pour l'avance, du dépit, de l'égoïsme et des utilités maximisant les gains. Le modèle d'Andreoni and Miller (2002) s'écarte des approches étudiées ci-dessus en modélisant les préférences d'autrui à l'aide d'une fonction d'utilité CES. Ainsi, il intègre les préférences complètement égoïstes, les préférences maximales rawlsiennes et les préférences utilitaires comme un cas particulier. Il est intéressant de noter qu'elle permet de mesurer le degré d'aversion à la différence ou de préoccupation pour la maximisation du gain total. Fisman et al. (2007) font progresser cette approche pour distinguer la manière dont les individus veulent distribuer le revenu entre eux et les autres ainsi qu'au sein d'un groupe avec autres personnes, en établissant une distinction claire entre l'altruisme, d'une part, et les préférences distributives d'autre part.

La littérature citée ci-dessus soutient que les individus ont une préférence intrinsèque pour

---

<sup>8</sup>Un autre modèle d'aversion pour l'inégalité célèbre, bien que moins influent, est Bolton and Ockenfels (2000), où un individu se soucie de la mesure dans laquelle son revenu diverge du revenu moyen du groupe ou de la société.



l'égalité ou l'efficacité. Toutefois, on peut également supposer que les individus rejettent toujours une égalisation des revenus parce qu'ils considèrent que la répartition actuelle des revenus est équitable.

$$D^F(\mathbf{y}) = D(\mathbf{y}) \quad (4.8)$$

Dans la littérature, ce point de vue est généralement appelé le point de vue de l'équité libertaire (voir par exemple Cappelen et al., 2007) et s'appuie sur la tradition des penseurs libertaires tels que Friedrich A. Hayek, qui s'opposent à toute redistribution des inégalités à partir des transactions du marché.<sup>9</sup> Cette vision de l'équité est fondée sur les résultats, car la répartition des revenus qui prévaut est considérée comme équitable, quelle que soit la manière dont elle est apparue.

Ce que les concepts énumérés ci-dessus ont en commun, c'est qu'ils sont purement *basés sur le résultat* et donc *consequentialistes*. En d'autres termes, les individus ne se préoccupent pas de savoir *pourquoi* une répartition de revenus atteint la forme qu'elle atteint, mais la rejettent ou l'acceptent en fonction de sa forme, par rapport aux alternatives réalisables. Cela change si l'on suppose que le *processus* qui génère la distribution des revenus influence les attitudes d'inégalité qui ont été conceptualisées par les théories du désert et de l'équité (Konow, 2003), que je désigne comme le principe d'équité méritocratique ou libéral égalitaire tout au long de cette thèse.

Le point de départ est la reconnaissance du fait que les différences de revenus peuvent être dues à différentes raisons telles que la chance, la naissance, le choix et l'effort. Le critère essentiel est qu'il faut égaliser les différences de revenus qui échappent au contrôle d'un individu. En revanche, celles dont un individu est responsable (Fleurbaey, 2008) ou peut être tenu responsable (Konow, 1996) en raison de ses choix individuels ne devraient pas être égalisées (voir Fleurbaey, 2008, pour une introduction au fondement philosophique des théories libérales égalitaires de la justice). Cette notion peut être appliquée à la prise de risque (Dworkin, 2002) et à la mesure dans laquelle l'effort devrait être récompensé (Cappelen and Tungodden, 2009). Elle peut être capturée (avec beaucoup de simplification) comme suit dans  $D^F(\mathbf{y})$  :

$$D^F(\mathbf{y}) = \mathbf{x} + \min_{D(\mathbf{y}) \in \Omega} \left\{ \sum_{k=1}^N |w_k - \tilde{w}| \right\}, \quad (4.9)$$

où l'on suppose que le revenu individuel  $y_i$  est composé de  $x_i$  qui représente la partie du revenu total dont un individu peut être tenu responsable et de  $w_i$  qui est la partie du revenu dont un agent ne peut être tenu responsable. Dans (4.9),  $\mathbf{x}$  est donc un vecteur caractérisant les revenus pour lequel les individus peuvent être tenus responsables. En revanche,  $w_k$  est la partie du revenu pour laquelle l'individu  $k$  ne peut être tenu responsable,  $\tilde{w}$  est le revenu médian de  $\mathbf{w}$ .

La question qui en découle est la suivante : de quoi devons-nous être tenus responsables ? Qu'est-ce qui entre dans  $w_i$  et qu'est-ce qui entre dans  $x_i$  ? Cette question est dans l'œil de l'observateur et dépend de la définition individuelle de ce qui est sous le contrôle d'un individu

<sup>9</sup>Il est intéressant de noter que l'argument de Hayek en faveur de la préservation des inégalités n'était *pas* fondé sur l'équité ou la méritocratie (see Sandel, 2020, for a discussion).

et de l'importance que cela revêt pour l'équité distributive. Certains peuvent, par exemple, affirmer que la volonté d'exercer un effort est un trait individuel, donné par la naissance, et donc hors du contrôle individuel. D'autres peuvent affirmer qu'il ne faut pas égaliser les différences de revenus dues à des différences de talent. Et d'autres encore soutiennent qu'il faut le faire.<sup>10</sup> Si ces distinctions dépendent de ce que les individus considèrent comme des dimensions pertinentes de la responsabilité, ce qui peut être caractérisé comme une préférence ou une conviction, il y a aussi la question sur comment les individus perçoivent la réalité. Les individus peuvent avoir des idéaux d'équité similaires mais avoir des convictions différentes dès lors qu'il s'agit de savoir si les individus sont réellement responsables de leurs résultats parce que leurs croyances sur l'environnement économique divergent (Alesina et al., 2020). Par exemple, une recommandation politique clé de la pensée méritocratique est d'appeler à l'égalité des chances (Arneson, 1989; Roemer, 1998; Sandel, 2020). La mesure dans laquelle cela est réalisé est une conviction qui varie d'un individu à l'autre (Alesina et al., 2018), et cette conviction est au centre du troisième chapitre.

**Quels sont les motifs pertinents ?** La multiplicité des motifs examinés ci-dessus soulève la question sur les motifs non-égoïstes pertinents qui sous-tendent les attitudes d'inégalité. Il s'agit, avant tout, d'une question empirique. La littérature empirique en économie expérimentale a une grande tradition d'identification et de démêlage des différents motifs de considération des individus. Une partie de la littérature utilise des jeux classiques tels que le jeu de l'ultimatum (Güth et al., 1982) ou le jeu du dictateur (Kahneman et al., 1986b; Forsythe et al., 1994) pour mesurer l'aversion pour l'inégalité. Un grand nombre des modèles cités ci-dessus ont été inspirés par les régularités comportementales trouvées dans ces jeux, qui ont montré que 40 à 50 % des offres basses sont généralement rejetées dans les jeux de l'ultimatum (Ledyard, 1995; Cooper and Kagel, 2016) et que les sujets des jeux du dictateur donnent en moyenne 28 % de leur gain initial (Engel, 2011). Bellemare et al. (2008) par exemple, adaptent une version augmentée du modèle Fehr and Schmidt au comportement dans un jeu d'ultimatum et de dictateur et trouvent des preuves considérables d'envie dans la population néerlandaise mais moins d'aversion pour l'avance. Un problème auquel on est confronté lorsque l'on identifie les attitudes d'inégalité en utilisant ces jeux est qu'ils ne peuvent pas contrôler les motifs réciproques et/ou distinguer les différents motifs distributifs. Pour contrôler les motifs réciproques et distinguer les différents motifs distributifs les uns des autres, une vaste littérature expérimentale (y compris les chapitres de cette thèse) utilise des jeux de dictateur augmentés, dans lesquels le dictateur peut choisir entre un ensemble de distributions de revenus. En faisant varier l'ensemble des distributions de revenus ( $\Omega$ ), les expérimentateurs peuvent alors démêler les différents motifs énumérés ci-dessus : égoïsme, envie, aversion pour l'avance, efficacité et maximin.

Des travaux antérieurs dans cette littérature ont trouvé des preuves que les individus sont moins préoccupés par l'inégalité que par le gain des moins plus démunis et l'efficacité (Charness and Rabin, 2002; Kritikos and Bolle, 2001; Engelmann and Strobel, 2004, 2007). Cela a déclenché

<sup>10</sup>Ce dernier point est probablement la manière la plus fréquente dont le mérite est défini dans le discours public, dans le sens où un système économique devrait donner à chacun la chance de "s'élever aussi loin que son *talent* le lui permet" (Sandel, 2020).

un débat sur la question de savoir si les préoccupations en matière d'inégalité ne sont pas aussi importantes qu'on le pensait auparavant, ou si les preuves sont un artefact de l'élicitation des préférences d'étudiants de premier cycle en économie (voir Fehr et al., 2006). L'argument repose sur le fait que les préférences distributives peuvent être hétérogènes dans différentes sous-populations. Par conséquent, les conclusions d'une population (les étudiants) peuvent ne pas s'appliquer à l'ensemble de la population. Pour obtenir une réponse à cette question, il faut caractériser l'hétérogénéité des préférences distributives non-égoïstes à l'aide d'échantillons représentatifs de la population qui nous intéresse.

En outre, la caractérisation de l'hétérogénéité des préférences distributives est importante pour évaluer les conséquences des préférences non-égoïstes, car même les préférences minoritaires peuvent parfois avoir une forte influence sur les résultats. Par exemple, même les proposant égoïstes dans un jeu de l'ultimatum proposent un montant positif dès lors qu'une *minorité* suffisante de récepteurs ont une aversion pour l'inégalité (Fehr and Schmidt, 1999). De même, les politiciens peuvent cibler les politiques de distribution de manière à ce qu'elles correspondent aux préférences d'un groupe important d'électeurs (par exemple, les personnes âgées). Les expériences de laboratoire standard (y compris le chapitre 2 de cette thèse) ont trouvé une hétérogénéité substantielle dans les préférences distributives, même si la population qu'elles étudient est relativement homogène (Andreoni and Miller, 2002; Fisman et al., 2007; Cappelen et al., 2007; Bruhin et al., 2018). Cela indique que l'hétérogénéité est susceptible d'avoir de l'importance, même si la population d'intérêt est relativement petite et homogène (par exemple, sur le lieu de travail), et que les résultats basés sur les moyennes de la population peuvent être susceptibles de cacher des types de préférences pertinents.

Ces dernières années, l'utilisation d'échantillons représentatifs nous a permis de caractériser l'hétérogénéité par des caractéristiques observables. Fisman et al. (2015b), par exemple, comparent les préférences non-égoïstes d'un échantillon représentatif d'Américains à celles d'étudiants d'universités d'élite. Leur modèle, basé sur Andreoni and Miller (2002), permet d'identifier la position des individus sur le compromis efficacité/égalité, ainsi que le degré d'égoïsme. Ils montrent qu'une part plus importante parmi les étudiants d'élite est égoïste et qu'ils sont plus susceptibles d'être axés sur l'efficacité que sur l'égalité, par rapport à l'échantillon représentatif d'Américains. Ce résultat est particulièrement intéressant si l'on considère que les étudiants d'élite sont susceptibles d'occuper des postes influents plus tard dans la vie active.<sup>11</sup> D'autres études utilisant des échantillons représentatifs ont révélé qu'une grande partie de la population globale est plus sensible aux préoccupations d'inégalité qu'à celles d'efficacité (Almås et al., 2020; Müller and Renes, 2021; Epper et al., 2020). Ces études révèlent également qu'une partie non négligeable de la population a des opinions libertaires en matière d'équité (Almås et al., 2020).

Dans l'ensemble, cette littérature montre que les préoccupations égalitaires et d'efficacité sont des motifs pertinents pour les attitudes envers l'inégalité. Les préoccupations liées à l'efficacité semblent être particulièrement importantes dans les échantillons d'étudiants et moins dans les échantillons représentatifs. Néanmoins, il existe encore des différences dans les préférences

---

<sup>11</sup>Il donne, en outre, du preuve empirique à l'argument de Fehr et al. (2006) selon lequel les étudiants sont plus axés sur l'efficacité que la population globale.

obtenues entre les études qui utilisent des groupes de sujets similaires. Les recherches futures devraient déterminer dans quelle mesure ces différences sont dues à l'utilisation de méthodes d'élicitation différentes.

Les preuves citées ci-dessus proviennent d'expériences (ou de traitements) où la source de l'inégalité est la chance. Les méritocrates, en revanche, se soucient de la source de l'inégalité. Ainsi, une grande partie de la littérature étudie si l'acceptation de l'inégalité est sensible à la source de l'inégalité. Des enquêtes ont étudié la pertinence des croyances relatives à la chance et à l'effort sur les préférences en matière de redistribution. Elles révèlent généralement que les croyances relatives au rôle de la chance et de l'effort dans la réussite sont fortement corrélées à l'acceptation de l'inégalité et, plus généralement, aux préférences en matière de redistribution (Alesina and La Ferrara, 2005; Fong, 2001).

Les premières approches expérimentales ont étudié des expériences de marchandage ou de dictateur où le rôle du premier arrivant est gagné en étant meilleur à un quiz. Ces expériences ont généralement montré que le marchandage final favorise le sujet qui a été meilleur au quiz par rapport au cas où les rôles ont été déterminés de façon aléatoire (Hoffman et al., 1994; Ruffle, 1998) ou que les dictateurs se comportent de façon plus égoïste après avoir gagné leur rôle par un concours (Cherry et al., 2002).

Des approches plus récentes utilisent des jeux de dictateur où les sujets génèrent une distribution initiale du revenu par leurs décisions. Konow (2000) a été le pionnier de cette approche en demandant à des paires de sujets de préparer des lettres. La production individuelle était ensuite multipliée par un prix, ce qui conduisait à une distribution initiale de la valeur de la production. Les dictateurs (soit l'un des sujets de la paire, soit un spectateur neutre) ont ensuite la possibilité de redistribuer la valeur de la production *commune* tout en étant informés de la contribution de chaque récepteur à la valeur de la production commune. Selon les traitements, l'expérimentateur fait varier si les différences d'effort ou les différences de prix (exogènes) expliquent les différences de contributions individuelles à la valeur de la production conjointe. Konow montre que les spectateurs préfèrent généralement les distributions qui reflètent les différences d'effort mais pas celles qui reflètent les différences de chance, ce qui est en accord avec le principe de responsabilité décrit ci-dessus. Frohlich et al. (2004) rapportent les résultats d'une expérience de jeu de dictateur avec une étape de production préalable et montrent qu'une grande partie des dictateurs tiennent compte des différences de production. Cappelen et al. (2007) étudient le comportement distributif où les sujets sont invités à prendre une décision d'investissement. Le rendement de l'investissement est soit élevé, soit faible. Ils sont ensuite jumelés avec une autre personne et il leur est demandé de distribuer le résultat commun de la phase d'investissement. Il est essentiel que les sujets sachent combien l'autre individu a investi et quel était son taux de rendement. Ainsi, les sujets peuvent faire la distinction entre les facteurs qui contribuent à la production totale et ceux qui échappent à celui du destinataire ou au leur. Les auteurs montrent que les sujets sont, en moyenne, sensibles à la source de l'inégalité. Ils sont plus susceptibles de choisir une distribution inégale si la différence de production résulte de différences dans les décisions d'investissement plutôt que de différences dans le rendement de l'investissement. Almås et al. (2020) utilisent une expérience dans laquelle un spectateur impartial est jumelé à deux travailleurs. Le

spectateur sait si les différences initiales de revenu sont dues à des différences d'effort ou simplement à la chance. Ils ont mené cette expérience sur des échantillons représentatifs au niveau national en Norvège et aux États-Unis et ont montré que la source de l'inégalité initiale a une incidence sur les décisions de redistribution dans les deux pays. Cette expérience a été reproduite dans le monde entier, et ils identifient un fort effet chance-mérite dans presque tous les pays (Almås et al., 2021). Lefgren et al. (2016) étudient l'effet chance-mérite dans le contexte du vote pour la redistribution. Ils trouvent un effet de mérite important et significatif sur le comportement de vote et montrent que les sujets sont prêts à voter pour des taxes qui vont à l'encontre de leur intérêt personnel pour récompenser l'effort.<sup>12</sup>

Si les études ci-dessus révèlent toutes des preuves solides de l'importance de la source de l'inégalité, ce n'est pas toujours le cas. Par exemple, des études récentes comparent les décisions de redistribution prises derrière le voile de l'ignorance et celles prises une fois qu'un individu connaît sa position. Ces études, y compris le chapitre 1 de cette thèse, trouvent un effet chance-mérite lorsque les individus choisissent derrière le voile de l'ignorance, mais cette différence disparaît une fois que les individus connaissent leur position dans la distribution (Bjerk, 2016; Durante et al., 2014).

Comme dans la littérature empirique qui étudie les préférences distributives lorsque la source de l'inégalité est la chance, il faut se demander si *tous* les individus sont sensibles à la source de l'inégalité. Des expériences récentes ont posé cette question et ont révélé une hétérogénéité considérable dans les opinions sur l'équité, classant les individus dans différents types d'équité. Cappelen et al. (2007), par exemple, identifient trois types d'équité distincts et proéminents : (1) les libertaires (acceptant toute inégalité initiale), (2) les méritocrates (acceptant les inégalités dues aux différences d'effort mais pas à la chance), et (3) les égalitaristes dans un groupe relativement homogène d'étudiants en commerce. L'étude de Almås et al. (2020) identifie la distribution de ces opinions sur l'équité en utilisant des échantillons représentatifs de la population américaine et norvégienne. Comme le montre la figure 4.2, les auteurs constatent que chaque vision de l'équité est représentée dans les deux pays.

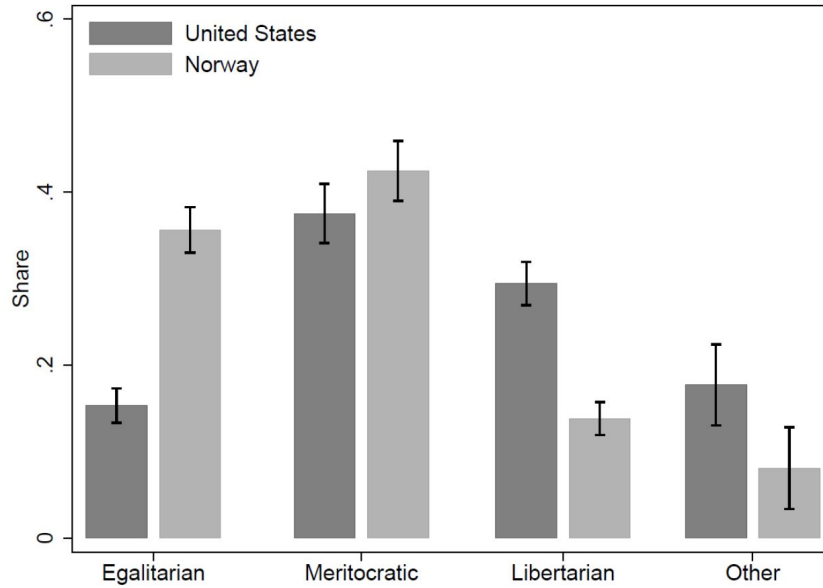
Même si l'étude révèle également une importante hétérogénéité au sein du pays,<sup>13</sup> c'est l'hétérogénéité entre les pays qui est, sans doute, la plus frappante. Comme le montre la figure 4.2, l'option "méritocratique" est le principe de justice qui attire le plus grand nombre d'individus dans les deux pays. Cependant, les Américains sont significativement plus susceptibles d'être libertaires que les Norvégiens, et ils sont significativement moins susceptibles d'être

---

<sup>12</sup>Notez que les différences de revenus peuvent non seulement résulter de différences d'effort ou de chance, mais elles peuvent aussi refléter des différences de choix plus généralement. Plusieurs expériences étudient si les différences dans les choix risqués affectent l'acceptation de l'inégalité. Cappelen et al. (2013) trouvent que les sujets redistribuent en moyenne moins, si *ex-post* l'inégalité reflète plutôt des différences dans les choix risqués. Des résultats similaires sont trouvés dans Akbaş et al. (2019). Un article connexe de Mollerstrom et al. (2015) étudie comment les individus font la distinction entre la chance brute et la chance optionnelle (Dworkin, 2002). Leur principale conclusion est que les individus peuvent être caractérisés comme des "compensateurs de choix" qui tiennent compte du fait qu'un individu s'est exposé à la chance d'option, même si cela n'a pas influencé son revenu final. Il s'agit d'un résultat intéressant car il indique que les intentions semblent plus pertinentes pour la décision distributive que les résultats.

<sup>13</sup>dans les deux pays, les femmes sont plus susceptibles de rejeter l'inégalité que les hommes, et les conservateurs ont tendance à mieux accepter l'inégalité. Le statut socio-économique est un facteur prédictif de l'acceptation de l'inégalité aux États-Unis, mais pas en Norvège.

FIGURE 4.2 : Distribution des types d'équité en Norvège et aux États-Unis constatée dans l'étude de Almås et al. (2020). La figure vient directement de Almås et al. (2020).



égalitaires. Cette différence dans la composition des types contribue à expliquer les différences entre les pays en matière de demande de redistribution. Cela est à première vue surprenant, parce que la littérature précédente a souvent souligné que cette différence peut plutôt provenir de l'hétérogénéité entre les pays dans le poids que les citoyens accordent à l'égalité par rapport à l'efficacité, ou des croyances sur la source de l'inégalité (les deux sont contrôlés par la conception dans cette expérience).

## 2.2 Qu'est-ce qui influence les attitudes envers l'inégalité ?

La première partie de l'étude de la littérature a présenté les différentes motivations des attitudes à l'égard de l'inégalité et de celles-ci. Cette évidence montre que les individus se soucient de la répartition des revenus et qu'ils le font pour différentes raisons : certains sont motivés par la maximisation de leur gain, d'autres veulent récompenser les personnes qui réussissent mieux une tâche, tandis que d'autres encore veulent minimiser les inégalités ou aider les plus démunis. L'évidence montre que la fonction de valeur présentée dans (4.1) est une caractérisation pertinente des attitudes envers l'inégalité et établit les bases nécessaires à la problématique centrale de cette thèse.

Dans cette partie, je présenterai l'évidence empirique de l'interaction de ces motifs avec l'environnement économique auquel nous sommes confrontés et les expériences que nous vivons. Je m'intéresse donc à la manière dont les éléments de la fonction de valeur (4.1) dépendent du contexte et sont influencés par l'environnement économique et les expériences. Ils affectent les attitudes en matière d'inégalité en influençant les perspectives de revenu  $y_i$ , en façonnant les principes d'équité relatifs aux autres  $D^F(\mathbf{y})$ , et en affectant ce qui entre dans  $D^F(\mathbf{y})$  (par exemple, les croyances sur la source de l'inégalité), en maintenant le principe d'équité constant.

J'aborderai ce sujet sous trois angles différents qui me semblent particulièrement pertinents pour le propos général des chapitres de ma thèse. Premièrement, je montrerai si les idéaux d'équité  $D^F(\mathbf{y})$  sont malléables et influencés par notre environnement et nos expériences. Deuxièmement, j'aborderai de la manière dont l'incertitude économique influence les attitudes à l'égard de l'inégalité à travers les perspectives de revenus. Troisièmement, je présenterai comment la littérature s'intéresse à comment l'environnement et les expériences économiques influencent la façon dont les individus apprennent l'importance relative de la chance et de l'effort.<sup>14</sup>

### Malléabilité des principes d'équité

L'une des façons dont l'environnement et les expériences économiques affectent les attitudes à l'égard de l'inégalité est de façonner le principe d'équité pertinent lui-même, c'est-à-dire  $D^F(\mathbf{y})$ . Ainsi, cette littérature s'interroge sur la manière dont les principes d'équité changent, tout en maintenant constantes les croyances sur d'autres paramètres pertinents tels que la variation du revenu propre et les croyances sur l'importance relative de la chance et de l'effort.

**Points de vue égocentriques sur l'équité** Plus haut, j'ai montré que les individus se soucient des inégalités non seulement parce qu'ils cherchent à tirer profit d'une réduction potentielle de l'inégalité, mais aussi pour des raisons normatives, car ils trouvent certaines distributions de revenus plus justes que d'autres. L'une des questions ciblées par la littérature précédente est de se demander s'il existe une interaction entre des motifs égoïste et non-égoïste. Les opinions d'équité égocentrique ont été introduites en économie par Babcock et al. (1995) dans le contexte du marchandage. Ces opinions d'équité égocentrique peut être rationalisée en supposant que les individus cherchent à justifier moralement leur demande égoïste d'inégalités plus ou moins élevées en exploitant une marge de manœuvre morale (Dana et al., 2007). Cela se produit lorsque deux motifs souhaités (agir équitablement et maximiser le revenu) entrent en conflit, ce qui entraîne une dissonance cognitive (Festinger, 1957). Cette tension peut être soulagée par un comportement intéressé. Un biais égocentrique dans les opinions d'équité ou les attitudes d'inégalité peut provenir de deux sources : premièrement, il peut y avoir un biais égoïste dans les principes d'équité, ce qui signifie que les individus adoptent le principe d'équité qui leur procure le revenu le plus élevé. Deuxièmement, il peut y avoir un biais égoïste dans les croyances. Dans ce cas, les individus peuvent exploiter l'incertitude concernant le coût d'efficacité de la redistribution ou l'importance relative de la chance et de l'effort pour justifier leur opposition ou leur soutien à l'inégalité.<sup>15</sup>

Les principes d'équité égoïstes supposent que les individus modifient leur principe d'équité normative si leur propre revenu est en jeu. Il convient de préciser que cela n'est pas pris en compte par  $\alpha$  dans (4.1), qui caractérise le compromis entre la maximisation de son propre revenu et le paiement d'un coût moral si la distribution des revenus mise en œuvre n'est pas la distribution

<sup>14</sup>Il va sans dire qu'il existe d'autres canaux qui façonnent les attitudes envers l'inégalité que je n'aborde pas dans la revue car ils ne sont pas directement liés aux chapitres de ma thèse. Il s'agit notamment de la manière dont l'identité sociale façonne les préférences en matière de redistribution et les comportements de respect des autres. (Luttmer, 2001; Kranton et al., 2020; Shayo, 2009).

<sup>15</sup>J'aborderai ce dernier point plus en détail à la fin de cette section.

des revenus normativement préférée. L'idée derrière les principes d'équité égoïstes est que  $D^F(\mathbf{y})$  lui-même peut changer. Par exemple, un individu peut avoir des principes d'équité libertaires s'il se trouve en haut de la distribution initiale de revenus, mais adopter des principes d'équité égalitaires s'il se trouve en bas de la distribution de revenus. Théoriquement et empiriquement, cela a été étudié par Konow (2000) que j'ai présenté ci-dessus. Konow trouve des preuves de l'existence de points de vue équitables égoïstes car il montre que les dictateurs bienveillants qui se sont précédemment engagés dans une décision de distribution qui a affecté leur propre revenu acceptent davantage les inégalités dues à la chance.<sup>16</sup> Cependant, la littérature qui a suivi n'a pas trouvé de preuves solides d'un tel biais d'intérêt. Cappelen et al. (2007) exploitent la richesse de leur modèle pour retirer  $\alpha$  et ensuite comparer si les sujets s'en tiennent à leur principe d'équité dans des situations où un biais d'intérêt personnel prédit une décision de distribution différente de celle prédite par le principe d'équité identifié (en maintenant  $\alpha$  constant) mais ils n'ont trouvé aucune preuve qui confirme l'hypothèse. Ce résultat a été reproduit dans l'étude de Cappelen et al. (2013) sur l'équité et la prise de risque. Les auteurs y comparent le comportement de distribution des spectateurs et des parties prenantes. Alors que les décisions de distribution des parties prenantes sont biaisées pour se rendre plus avantageuses (capturées par  $\alpha$ ), la distribution des principes d'équité reste stable. Les preuves tirées de la littérature empirique récente, qui utilise des modèles plus complexes et génère des données plus riches, montrent donc que les principes d'équité sont remarquablement stables et relativement peu susceptibles d'être soumis à ce type de biais égocentrique.<sup>17</sup>

Une question empirique qui reste à résoudre est de savoir si les points de vue égocentriques sur l'équité sont égocentriques pour la situation de vie globale d'une personne, comme le souligne Cappelen et al. (2020b). Des preuves corrélationnelles confirment cette hypothèse. Almås et al. (2017) montrent que les enfants issus d'un milieu socio-économique élevé (ci-après, MSE) acceptent mieux l'inégalité que ceux issus d'un milieu socio-économique faible et Fisman et al. (2020) montrent que les individus qui ont connu une augmentation du revenu de leur ménage sont devenus plus égoïstes.<sup>18</sup> Il est toutefois difficile d'identifier une relation causale, car il faudrait une variation exogène des antécédents, des aspirations et de la situation actuelle de l'individu. Une autre question empirique qui reste ouverte est de savoir dans quelle mesure ces opinions d'équité sont stables dans différents contextes et si le comportement intéressé est pertinent dans cette dimension. On pourrait, par exemple, supposer que les individus ont des opinions différentes en matière d'équité dans le contexte d'un lieu de travail ou d'une entreprise et au niveau de la société.

<sup>16</sup>Une expérience très récente de Amasino et al. (2021) réplique Konow (2000). Ils étudient également le mécanisme d'offre à l'origine de ce résultat et trouvent des preuves en faveur de l'hypothèse selon laquelle le biais d'intérêt est dû au fait que les sujets accordent moins d'attention aux informations sur le véritable degré de mérite.

<sup>17</sup>Cela ne signifie pas que les biais égocentriques dans les jugements moraux ne sont jamais pertinents. Comme je le montrerai ci-dessous, les croyances relatives à l'effort et à la chance sont beaucoup plus malléables à ce type de biais intéressés et les travaux récents de Barron et al. (2020) fournissent des preuves expérimentales qui sont cohérentes avec la notion selon laquelle les individus agissent conformément à un motif moral (vérité ou équité distributive) qui produit un revenu plus élevé.

<sup>18</sup>Il convient de souligner que le message général de cet article est que les préférences pour le respect des autres sont remarquablement stables au sein d'un même individu et dans le temps.



**Les opinions sur l'équité sont-elles façonnées par les expériences et les institutions ?**

Le principe d'équité peut également être façonné par l'environnement institutionnel et les expériences que nous faisons. Encore une fois, je me concentrerai ici sur la façon dont  $D^F(\mathbf{y})$  est façonné par les institutions et je me concentrerai plus loin sur les croyances en la chance et l'effort.

Un volet de la littérature étudie comment l'éducation affecte les opinions sur l'équité. Jakiela (2015), par exemple, étudie les préférences sociales dans des villages ruraux du Kenya et dans une université américaine. Son expérience utilise un jeu de dictateur et varie selon si le revenu initial est gagné ou aléatoire. Elle ne constate aucun effet du revenu gagné au Kenya, mais un effet important aux États-Unis. Il est intéressant de noter qu'elle constate une hétérogénéité au sein de l'échantillon kenyan : ceux qui vivaient dans des communautés ayant un accès à la route principale et un niveau d'éducation plus élevé étaient plus susceptibles de récompenser l'effort. Cela fournit des preuves laissant penser que l'exposition aux marchés et à l'éducation (de style occidental) peut conduire à une plus grande prévalence des idéaux d'équité méritocratique. Cappelen et al. (2020a) fournissent des preuves causales montrant que l'éducation façonne les opinions d'équité, en exploitant la variation aléatoire d'un essai contrôlé randomisé (ECR) sur l'éducation maternelle. Ils montrent que la fréquentation d'une école maternelle rend les enfants plus égalitaires, tandis que ceux qui ont été scolarisés à la maison accordent une plus grande importance à l'efficacité. En revanche, cela n'affecte pas leur degré d'égoïsme. Kosse et al. (2020) montrent que la pro-socialité peut être façonnée par des interventions dans la petite enfance. Ils réalisent un ECR en plaçant les enfants dans un programme de mentorat. Ils montrent que le programme a un effet important sur la pro-socialité des enfants — qui est représentée par  $\alpha$  dans (4.1) — et comble même un écart initial de pro-socialité entre les enfants issus de familles de statut socio-économique faible et élevé. Cet effet est médié par le fait d'avoir un modèle pro-social et des interactions sociales. Ces études montrent que les attitudes d'inégalité, au sens large, sont façonnées par l'éducation et par les personnes avec lesquelles nous interagissons pendant l'enfance.

On peut s'interroger également sur comment les expériences personnelles façonnent les principes d'équité. Bauer et al. (2014) étudient comment l'expérience du conflit façonne le comportement égalitaire en Géorgie et en Sierra Leone. Ils montrent que le fait de subir la guerre en tant qu'enfant ou jeune adulte rend les individus plus égalitaires envers les membres du groupe interne, mais pas envers les membres du groupe externe. Fisman et al. (2015a) étudient comment les préférences distributives ont changé lors de la Grande Récession dans un cadre où la source de revenu est la chance. Ils constatent que la récession a rendu les sujets plus égoïstes et plus préoccupés par l'efficacité. Cappelen et al. (2021) se concentrent sur la façon dont la pandémie de COVID-19 façonne les opinions sur l'équité en utilisant une expérience d'amorçage. Ils constatent que la pandémie a rendu les Américains plus enclins à accepter les inégalités dues à la chance, mais aussi plus disposés à donner la priorité à la société sur leurs propres problèmes. Barr et al. (2016) étudie si une période de chômage affecte les opinions d'équité. Elle compare le comportement distributif en laboratoire à l'aide d'un modèle longitudinal. En exploitant la variation intra-sujet des périodes de chômage, elle constate que les individus deviennent plus égalitaires

(moins susceptibles de récompenser les différences d'effort) après avoir connu une période de chômage. Cette constatation est solide pour contrôler les croyances méritocratiques et suggère que les points de vue sur l'équité ont changé.

Les preuves contrôlées provenant du laboratoire qui démontrent comment les expériences façonnent les attitudes envers l'inégalité sont rares. Cassar and Klein (2019) montrent que l'expérience de la réussite nous fait mieux accepter l'inégalité que l'expérience de l'échec. Dans leur expérience, les sujets reçoivent un prix et sont ensuite jumelés avec deux *autres* individus dont l'un a gagné le prix mais pas l'autre. Ils prennent ensuite une décision de distribution. Les auteurs constatent que ceux qui ont gagné un prix sont plus susceptibles de récompenser le sujet qui a également gagné le prix, même si la répartition initiale était complètement aléatoire. Ils montrent que ce résultat peut être attribué au fait que les sujets se sentent plus proches de leur compagnon perdant ou gagnant. Des recherches plus approfondies sur l'effet causal des expériences sur les opinions en matière d'équité sont toutefois encore nécessaires.

Que nous apprend cette littérature sur la manière dont l'environnement économique et les expériences façonnent  $D^F(\mathbf{y})$ ? D'une part, la littérature citée ci-dessus montre que les principes d'équité sont remarquablement stables aux manipulations expérimentales. D'autre part, les principes d'équité sont, dans une certaine mesure, malléables aux crises macroéconomiques, au chômage, aux programmes éducatifs et sont corrélés au milieu familial. En combinant ces deux éléments, on peut conclure que la source de l'hétérogénéité peut être trouvée dans l'exposition aux cultures (Luttmer and Singhal, 2011), aux idées et aux opinions sur l'équité pendant l'enfance ou les années impressionnables (Krosnick and Alwin, 1989) et à travers les grands événements de la vie qui peuvent changer la vision du monde. Une question intrigante pour les recherches futures est de savoir dans quelle mesure les institutions de redistribution affectent les opinions sur l'équité, par exemple en affectant les normes. Cela pourrait aider à mieux comprendre pourquoi les opinions sur l'équité sont si hétérogènes entre les pays.

### **Attitudes envers l'inégalité et l'incertitude**

Les individus sont confrontés à une plus ou moins grande incertitude quant à leur position future dans la distribution des revenus. Cette variation peut façonner les attitudes à l'égard de l'inégalité d'un motif égoïste en façonnant les perspectives de revenus. L'étude de l'incidence de l'incertitude sur les attitudes égoïstes envers l'inégalité a montré qu'elle pouvait entraîner une demande de redistribution plus ou moins importante. La direction de l'effet est susceptible d'être déterminée par les attentes quant à l'endroit où l'on se retrouve dans la distribution.

L'incertitude quant à la position future d'une personne peut l'amener à se retrouver au bas de l'échelle des revenus. Pour se prémunir contre d'éventuels chocs de revenus négatifs, les individus peuvent rejeter l'inégalité et demander une redistribution des revenus. Cette idée, développée à l'origine par Harsanyi (1955), découle de la courbure de la fonction d'utilité et crée une demande de redistribution car elle fonctionne comme une assurance si l'on ne parvient pas à atteindre la moitié supérieure de la distribution des revenus. Cela devient particulièrement pertinent si l'on ne sait pas encore où l'on se situe dans la distribution des revenus. C'est également un argument en faveur du principe de différence de Rawls (1971) qui est choisi *avant*

que la position dans la distribution des revenus soit révélée. Empiriquement, il existe une relation entre l'aversion au risque et les préférences pour la redistribution. Gärtner et al. (2017) étudient un échantillon représentatif de la Suède, établissant une relation significative au niveau de la population. Plusieurs autres études ont tenté de quantifier dans quelle mesure l'aversion pour le risque peut expliquer une aversion pour l'inégalité en comparant les décisions prises dans un certain environnement avec celles prises dans un environnement incertain ou derrière le voile de l'ignorance. La conclusion générale de cette littérature est que le motif de l'assurance a son importance, mais qu'il ne suffit pas à expliquer entièrement l'opposition à l'inégalité (par exemple Johansson-Stenman et al., 2002; Carlsson et al., 2005; Schildberg-Hörisch, 2010). Le premier chapitre de ce document vient s'ajouter à cet ensemble de preuves. Nous y montrons que les motifs liés à l'assurance sont importants, mais qu'ils ne sont pas en mesure d'expliquer entièrement toute opposition à l'inégalité, reproduisant ainsi les preuves citées ci-dessus.

Si les individus peuvent rejeter les inégalités parce qu'ils veulent s'assurer contre les chocs négatifs sur les revenus, la littérature a également proposé que les individus rejettent les inégalités pour un motif que l'on pourrait qualifier de contrasté : l'aspiration à bénéficier de revenus élevés au sommet de la distribution. Ce mécanisme a été introduit pour la première fois par Hirschman and Rothschild (1973) sous le nom « d'effet tunnel ». Bénabou and Ok (2001) ont ensuite développé l'hypothèse connexe de la perspective de mobilité ascendante (POUM). Cette hypothèse repose sur l'idée qu'une forte inégalité peut être perçue comme bénéfique à l'avenir, car elle est le signe de son propre revenu futur. Ainsi, les croyances sur la mobilité ascendante sont essentielles pour l'acceptation de l'inégalité dans ce contexte. Des preuves expérimentales de ce canal sont fournies par Checchi and Filippin (2004) et Agranov and Palfrey (2020). Sur le terrain, les éléments qui vont dans le sens de l'hypothèse POUM se sont révélés particulièrement pertinents dans les économies en transition (Senik, 2004; Grosfeld and Senik, 2010) et dans les entreprises, où les salaires des cadres peuvent fournir des informations sur leurs propres perspectives d'avenir (Clark et al., 2009; Cullen and Perez-Truglia, 2018; Godechot and Senik, 2015). Alesina and La Ferrara (2005) analysent des données d'enquête et montrent que les personnes qui s'attendent à une croissance de leurs revenus sont plus susceptibles de s'opposer à la redistribution. Comme le POUM dépend des croyances individuelles sur la mobilité ascendante, il est étroitement lié aux croyances sur la capacité relative de chacun. Buser et al. (2020) montrent que la confiance dans la capacité relative est un prédicteur important des préférences en matière de redistribution et l'écart entre les sexes en matière de *sur*-confiance peut en outre expliquer une partie de l'écart entre les sexes dans la demande de redistribution. Dans cette thèse, deux de mes chapitres abordent ces idées. Le chapitre 1 étudie comment les perspectives de revenus peuvent effacer toute opposition à l'inégalité et le chapitre 3 étudie comment les croyances méritocratiques, liées à la mobilité, sont malléables aux incitations économiques auxquelles les individus sont confrontés. Ensuite, j'aborderai plus en détail les croyances méritocratiques.

### **Formation de croyances sur l'importance de la chance et de l'effort**

La croyance en la source de l'inégalité est un facteur important qui détermine si les individus pensent qu'une répartition de revenus est juste ou injuste. En particulier pour les individus

méritocratiques, l'évaluation d'une répartition de revenus dépend de la source de l'inégalité. Néanmoins, la source de l'inégalité a également de l'importance pour les motifs de la redistribution liés à l'intérêt personnel, car elle détermine les perspectives de revenus évoquées plus haut.<sup>19</sup>

L'importance réelle de l'effort par rapport à la chance n'est pas observée et est incertaine. L'incertitude quant à la source de l'inégalité implique que nous ayons des croyances sur le véritable état du monde, et ces croyances déterminent la façon dont on divise  $y$  en  $w$  et  $x$  (voir (4.9)); ce qui nous amène à nous interroger sur comment nous formons ces croyances et sur ce qui les influence.

**Mobilité économique et croyance en la chance et l'effort** Une question très pertinente sur le plan empirique et pratique consiste à se demander quel type d'information influence les croyances en matière de chance et d'effort. La réponse la plus courante à cette question est sans doute la mobilité économique. La mobilité peut être considérée comme une mesure de l'égalité des chances, car elle caractérise la mesure dans laquelle le statut socio-économique des parents détermine l'avenir d'une personne (Alesina et al., 2001). Comme il s'agit d'un facteur qui échappe au contrôle de l'individu, de faibles niveaux de mobilité économique peuvent indiquer que le milieu socio-économique est plus important pour avancer que le travail acharné. Par exemple, si la probabilité qu'un individu moyen issu d'un ménage de n'importe quel quintile de la distribution des revenus a 20 % de chances d'atteindre le quintile supérieur, alors la société devient complètement mobile, car tout le monde a les mêmes chances d'atteindre le sommet. En revanche, si les individus issus du quintile inférieur n'ont que 1 % de chances d'atteindre le quintile supérieur, alors que ceux du quintile supérieur ont 90 % de chances d'y rester, la mobilité est faible car la corrélation entre l'origine socio-économique et les résultats est beaucoup plus forte.

Les croyances en matière de mobilité influent sur les attitudes en matière d'inégalité de deux manières : Premièrement, par le biais canal POUM décrit ci-dessus et deuxièmement, par le biais du canal non-égoïste discuté dans cette partie de l'enquête. Le canal non-égoïste soutient que le niveau de mobilité reflète le degré d'inégalité des chances, qui est un critère essentiel pour que les différences de résultats reflètent des différences de choix individuels plutôt que d'héritage. Par ce biais, elle a une implication intéressante pour les décideurs politiques. Si les citoyens sont plus sensibles à l'égalité des chances plutôt qu'à l'égalité des résultats, ils exigeraient également des politiques favorisant la mobilité économique.<sup>20</sup>

Les fondements théoriques de l'effet de la mobilité sur les croyances en matière de chance et d'effort ont été exposés pour la première fois par Piketty (1995). Piketty soutient que les individus (ou dynasties) apprennent à connaître la chance et l'effort à travers des expériences de mobilité. Comme l'apprentissage est coûteux, différents individus peuvent finir par converger

<sup>19</sup>Les individus qui pensent être plus talentueux et plus travailleurs que les autres devraient avoir des perspectives de revenus plus élevées s'ils pensent que ces éléments sont importants pour déterminer leur rang de revenu par rapport à la chance (voir Buser et al., 2020, pour une analyse de l'excès de confiance en ses capacités et de l'acceptation de l'inégalité).

<sup>20</sup>En effet, le discours politique de ces dernières années a évolué dans ce sens et a été attribué à un changement de paradigme méritocratique au cours des dernières décennies (Sandel, 2020).

vers des croyances différentes sur la chance et l'effort, même s'ils ont commencé au même point dans le même système économique. Cette dynamique est due au fait que deux individus peuvent commencer par fournir le même effort, mais l'un des deux a subi un choc positif tandis que l'autre a subi un choc négatif. L'individu qui a subi le choc négatif peut alors réviser ses croyances à la baisse et fournir moins d'efforts, diminuant ainsi ses chances de connaître une mobilité ascendante, alors que l'inverse est vrai pour l'autre individu qui avait initialement de la chance. Cette dynamique divergente conduit à des équilibres multiples dans lesquels certains individus croient que l'effort est important et demandent de faibles niveaux de redistribution, tandis que le contraire est vrai pour d'autres qui ont connu de faibles niveaux de mobilité.

Ces dernières années, la littérature empirique sur la mobilité économique et les croyances effort–chance s'est considérablement développée en procédant à des évaluations causales entre les croyances en la mobilité et les croyances effort–chance, ainsi que les préférences pour la redistribution. Il est important de rassembler des preuves causales sur cette relation car le lien entre les croyances en la chance et l'effort et la mobilité est moins évident qu'il n'y paraît à première vue. Tout d'abord, les individus peuvent avoir des antécédents très forts en ce qui concerne les croyances relatives à la chance et à l'effort en raison de leur milieu culturel ou parce qu'ils voient une valeur affective dans le fait de croire que l'effort est important pour progresser (Lerner, 1980) ou bien le contraire. Deuxièmement, même si la mobilité est totale, cela ne signifie pas automatiquement que l'effort compte pour progresser. On pourrait, par exemple, avoir une mobilité totale dans un monde où (a) la chance est également répartie dans la société et la chance domine l'effort dans le processus de génération de revenus et (b) tout le monde travaille de manière égale et la chance est également répartie ; dans ce cas, la chance devient le facteur déterminant.<sup>21</sup>

La première preuve corrélacionnelle du lien entre les expériences de mobilité et les croyances effort–chance a été documentée par Alesina et al. (2001), en montrant que l'expérience de la mobilité professionnelle diminue la demande de redistribution. Giuliano and Spilimbergo (2014), en outre, montrent que l'expérience d'une récession façonne les croyances des individus en matière de chance et d'effort et, par conséquent, leurs préférences pour la redistribution. Cela peut être considéré comme une preuve en faveur du mécanisme proposé Piketty (1995). L'un des tests les plus explicites de Piketty (1995) est celui de Gärtner et al. (2019). Ils utilisent une expérience dans laquelle ils confrontent les Suédois à leurs propres expériences de mobilité. Ils montrent d'abord que les Suédois sous-estiment généralement le degré de mobilité au cours de leur vie. Par conséquent, les croyances en matière de mobilité ne sont pas nécessairement correctes. En outre, ils identifient des corrélations significatives entre les perceptions de la mobilité et les croyances en la chance et l'effort, ainsi qu'avec les préférences pour la redistribution. Le deuxième résultat exploite la composante expérimentale de leur enquête : Conformément à Piketty (1995), les individus sont plus susceptibles de croire à l'importance de l'effort pour avancer, lorsqu'ils apprennent que la mobilité était plus élevée qu'ils ne le pensaient initialement. Ainsi, l'article établit un lien de causalité entre le fait d'apprendre sa propre mobilité et les croyances de chance et d'effort des individus. Le changement exogène des croyances se traduit par un changement

---

<sup>21</sup>Ce dernier point devient particulièrement pertinent dans les marchés où le gagnant prend tout (Frank, 2017).

des préférences redistributives chez les individus de droite qui sous-estimaient auparavant leur mobilité ascendante.

Alesina et al. (2018) se concentrent sur la mobilité intergénérationnelle, plutôt que sur leurs propres expériences de mobilité.<sup>22</sup> À l'aide d'une enquête expérimentale menée en France, en Italie, en Suède, au Royaume-Uni et aux États-Unis, les auteurs montrent d'abord que les Européens sous-estiment les niveaux de mobilité intergénérationnelle, tandis que les Américains surestiment la mobilité. Ce résultat est frappant, car ils affirment que la mobilité réelle n'est pas si différente d'un continent à l'autre. Ils font ensuite varier de manière exogène les croyances des répondants sur la mobilité intergénérationnelle en montrant aux sujets traités des informations pessimistes sur la mobilité. Ces informations entraînent une diminution de la croyance selon laquelle l'effort est important pour progresser et une augmentation du soutien aux politiques de redistribution qui favorisent l'égalité des chances.<sup>23</sup>

**Formation motivée des croyances chance-effort** L'une des contributions de Alesina et al. (2018) et Gärtner et al. (2019) est de montrer que les individus ont tendance à mal percevoir la mobilité économique et que différents individus peuvent avoir des perceptions différentes de la réalité. Cela soulève la question du mécanisme qui conduit à cette hétérogénéité et à cette perception erronée des croyances. Bien qu'une telle perception erronée puisse se produire dans le cadre du modèle développé par Piketty (1995), il existe également des théories comportementales de distorsion de croyance motivée qui peuvent rationaliser ces résultats empiriques.

Ces dernières années, un grand nombre d'articles ont été consacrés à l'étude du biais d'attribution égoïste dans les croyances relatives à l'effort de la chance. Le comportement égoïste, tel que discuté ci-dessus, peut être appliqué à la formation de croyances liées à l'effort et à la chance dans un environnement incertain : Pour maintenir une image positive de soi, les individus peuvent délibérément vouloir attribuer l'échec à des éléments hors de leur contrôle (la chance) et le succès à des éléments sous leur contrôle (l'effort) (Frank, 2017). Ce mécanisme conduit à une polarisation des croyances relatives à la chance et à l'effort, où ceux qui réussissent dans la vie finissent par accorder trop d'importance à l'effort pour avancer, ignorant les bonnes fortunes qui ont pu les amener là en premier lieu ; ceux qui ont des difficultés dans la vie accordent trop d'importance à la chance par rapport à l'effort, ignorant peut-être qu'ils auraient pu faire plus dans le passé. Ce type de biais égocentrique peut également conduire à une polarisation des préférences distributives, car il détermine ce que les individus considèrent comme juste ou injuste.

Des contributions récentes de la littérature expérimentale ont apporté des résultats qui vont dans le sens du mécanisme susmentionné. Deffains et al. (2016) montrent que (de manière

<sup>22</sup>Piketty (1995) se concentrent sur les *expériences* de mobilité au niveau personnel ou familial, plutôt que sur l'apprentissage de la mobilité au niveau sociétal. Cependant, on peut appliquer une lecture intergénérationnelle ou globale du modèle. Dans ce cas, les prédictions seraient similaires dans le sens où apprendre que la mobilité est élevée est un signal que l'effort est plus susceptible d'être récompensé.

<sup>23</sup>Ce dernier effet n'est présent que pour les répondants de gauche. Ils affirment que cette hétérogénéité dans l'effet du traitement peut être due à une méfiance plus générale des individus de droite à l'égard du gouvernement, comme cela a été noté dans la littérature précédente (e.g. Kuziemko et al., 2015). Cela montre que cela ne doit pas nécessairement se traduire par une volonté de réduire les inégalités par le biais de la politique, car cette volonté peut dépendre d'autres facteurs qui ne sont pas directement liés aux croyances en matière de chance et d'effort.

exogène) avoir une performance relativement mauvaise dans une tâche est attribué à des facteurs hors de son contrôle, alors que le contraire est vrai pour ceux qui ont relativement bien réussi. Cela se traduit par une demande de redistribution respectivement plus élevée et plus faible, ainsi que par des attitudes polarisées à l'égard de la redistribution. Cassar and Klein (2019) trouvent un effet similaire en utilisant un jeu de dictateur pour montrer que ce biais d'attribution affecte l'acceptation de l'inégalité. Une contribution très récente de Fehr and Vollmann (2020) documente un effet similaire, tout en montrant en outre que cet effet n'est pas hétérogène en fonction de l'affiliation politique, ce qui indique que ce biais est relativement ancré dans la psyché humaine. Di Tella et al. (2007) présente des preuves sur le terrain qui vont dans le sens d'un biais d'attribution intéressé. Ils montrent que les squatteurs de terres sont plus susceptibles de croire que l'effort est important pour avancer s'ils ont reçu des titres fonciers de manière quasi-aléatoire. Enfin, on peut se demander si ce biais égocentrique est motivé par le désir de maintenir un niveau élevé de confiance en soi ou si les individus veulent exploiter une marge de manœuvre morale qui leur permet de justifier moralement le paiement de taxes moins élevées. Valero (2020) teste cette dernière hypothèse en utilisant une expérience de laboratoire et ne trouve aucune preuve en sa faveur.

Alors que la littérature sur le biais d'attribution soutient généralement que les individus déforment les croyances relatives à la chance et à l'effort parce qu'ils veulent conserver une image positive d'eux-mêmes, on peut également supposer que les individus déforment les croyances relatives à la chance et à l'effort parce qu'ils obtiennent une valeur de la croyance selon laquelle l'effort ou la chance sont importants pour avancer. La théorie la plus influente en économie qui intègre cette idée a été développée par Bénabou and Tirole (2006). Les auteurs soutiennent que les individus forment des croyances chance-effort motivées parce qu'ils veulent croire que l'effort est important pour avancer. Cette demande pour maintenir la croyance que l'effort est important peut être soit affective, soit instrumentale.

La filière affective a été proposée pour la première fois par la littérature psychologique et soutient que les individus ont le désir de croire que l'effort est récompensé et que chacun obtient ce qu'il mérite. L'argument sous-jacent est que croire en l'importance de l'effort pour avancer (croyances du monde juste) réduit l'anxiété car les individus ont envie de croire qu'ils peuvent contrôler leur vie (Lerner, 1980). Il s'agit d'un motif affectif de distorsion des croyances, où les individus reçoivent de façon *directe* l'utilité de croire que leur effort est récompensé. Un individu peut maintenant choisir délibérément d'ignorer les informations qui indiquent que l'effort est *non* récompensé pour maintenir cette croyance en un monde juste. Il s'agit d'une maximisation de l'utilité tant que le gain lié au maintien de la croyance est supérieur au coût de la distorsion de la croyance.

Le canal instrumental suppose que la valeur des croyances en un monde juste est motivante. Bénabou and Tirole (2006) soutiennent que les agents biaisés par le présent déforment les croyances sur la chance et l'effort pour surmonter une mauvaise allocation de l'effort due à leur manque de motivation. Ce type de distorsion des croyances se produit si le gain de la distorsion des croyances (surmonter le manque de volonté en croyant que l'effort est plus important qu'il ne l'est en réalité) est inférieur à son coût (exercer potentiellement trop d'effort). Ce compromis

dépend des incitations auxquelles les individus sont confrontés. Cela implique que la demande de distorsion des croyances motivantes est endogène à la structure d'incitation à laquelle les agents s'attendent à être confrontés - un sujet sur lequel je reviendrai plus loin. La littérature empirique sur ce type de distorsion des croyances est rare car la littérature empirique sur les croyances motivées s'est principalement concentrée sur les croyances qui sont pertinentes pour l'estime de soi. Le chapitre 3 est l'une des premières contributions à vérifier si les croyances liées à l'effort et à la chance sont déformées à des fins de motivation.

**Les croyances en la chance et l'effort sont-elles endogènes aux politiques de redistribution ?** L'une des principales questions ouvertes dans cette littérature est de savoir si ces croyances sont endogènes aux institutions économiques. Une question qui a retenu l'attention dans la littérature antérieure est celle de savoir si les croyances effort-chance sont façonnées par les politiques de redistribution auxquelles nous sommes confrontés. Les contributions théoriques de l'économie politique montrent que les croyances effort-chance peuvent en effet être façonnées par les niveaux de redistribution auxquels nous sommes confrontés, ce qui aide à comprendre les variations entre pays en ce qui concerne les croyances effort-chance et des niveaux de redistribution.

Alesina and Angeletos (2005) saisissent cette idée en soutenant que les niveaux de redistribution affectent la capacité des personnes à distinguer le succès par l'effort et le succès par la chance. Leur argument établit que lorsque les niveaux de redistribution sont élevés, les individus travaillent et investissent moins et il devient donc plus difficile de distinguer le succès dû à l'effort de celui dû à la chance. Cela conduit à des croyances pessimistes quant à l'importance de l'effort et à une demande persistante de redistribution. L'inverse est vrai si les niveaux initiaux de redistribution sont faibles. La distribution des revenus est alors beaucoup plus corrélée à l'effort, ce qui maintient une faible demande de redistribution. Les auteurs concluent leur article en indiquant les possibles origines des différences transatlantiques dans les croyances chance-effort et l'équité. Tout comme Piketty (1995), ils identifient les expériences historiques comme le principal moteur : en Europe, l'État-providence moderne a évolué à partir d'une société féodale, où les circonstances indépendantes de la volonté de chacun, comme l'héritage, jouent un rôle important dans le destin d'une personne. Selon les auteurs, les États-Unis, en revanche, sont issus d'une société relativement égalitaire et mobile (pour les hommes blancs non-esclavés), où les colons blancs étaient moins limités par leur héritage pour gravir les échelons de la société.

Bénabou and Tirole (2006) utilisent un type de mécanisme très différent qui montre comment les niveaux de redistribution peuvent affecter les croyances effort-chance. Comme je l'ai indiqué plus haut, les auteurs soutiennent que les individus veulent croire que l'effort est important pour progresser si les avantages de la distorsion des croyances l'emportent sur le coût de la distorsion des croyances. En d'autres termes, il doit y avoir une demande pour motiver la distorsion des croyances. C'est le cas si le gain marginal de l'effort est suffisamment élevé, ce qui est plus susceptible d'être le cas pour de faibles niveaux de redistribution. Ainsi, le gain lié à la distorsion des croyances augmente à mesure que les niveaux de redistribution diminuent. Cela pourrait donc conduire à une situation dans laquelle la société a de faibles niveaux de redistribution et



s'engage donc dans une distorsion des croyances motivante qui l'amène à surestimer l'importance de l'effort pour progresser. Pour des niveaux élevés de redistribution, les incitations à s'engager dans une distorsion des croyances motivantes sont moindres et, par conséquent, les individus seront plus pessimistes quant à l'importance de l'effort pour avancer.<sup>24</sup>

La littérature empirique sur ce sujet est relativement rare et devrait faire l'objet de recherches futures. Cela s'explique en partie par un problème d'identification qui se pose si l'on suppose d'une part que les croyances en la chance et l'effort sont une fonction de la redistribution et de l'inégalité après impôt et d'autre part que l'on prédit la demande de réduction des inégalités. Néanmoins, la littérature économique a fait quelques progrès dans cette direction au cours des dernières années. Gärtner et al. (2019) trouvent des évidences suggestives pour une lecture intergénérationnelle de Bénabou and Tirole (2006) en montrant que les parents transmettent à leurs enfants des croyances trop optimistes sur la chance et l'effort. C'est particulièrement le cas si les parents s'attendent à de faibles niveaux de redistribution dans le futur, ce qui est une prédiction clé du modèle théorique. Le chapitre 3 de cette thèse complète l'article susmentionné en testant une autre prédiction clé du modèle Bénabou and Tirole (2006), à savoir que les croyances en matière de chance et d'effort sont endogènes aux incitations auxquelles les individus s'attendent à faire face. Roth and Wohlfart (2018) montrent que l'expérience de niveaux élevés d'inégalité pendant les années impressionnables conduit les individus à être plus convaincus que l'effort est important pour avancer. Cela pourrait être rationalisé par le cadre développé par Bénabou and Tirole (2006), car le niveau d'inégalité affecte la demande de distorsion des croyances motivantes en façonnant les incitations.

Que nous apprend la littérature résumée ci-dessus ? Comme nous l'avons montré, les croyances relatives à l'effort et à la chance ne sont pas fixes mais constituent une variable sur laquelle les individus se forgent des croyances. Ces croyances sont façonnées par la manière dont nous percevons la mobilité économique et elles sont déformées à des fins personnelles. Les évidences empiriques, en particulier en laboratoire, suggèrent qu'elles sont - au moins dans une certaine mesure - plus malléables à l'environnement économique que l'idéal d'équité  $D^F(\mathbf{y})$  lui-même. Cela les rend particulièrement intéressantes à étudier et constitue une voie fructueuse pour les recherches futures.

### 2.3 Quelles sont les conséquences des attitudes à l'égard de l'inégalité ?

La littérature résumée ci-dessus visait principalement à caractériser les attitudes individuelles envers l'inégalité et à demander comment elles sont façonnées par l'environnement économique. Les chapitres 1 et 3 sont consacrés à répondre à cette problématique. Le chapitre 2 améliore également notre compréhension des conséquences des attitudes envers l'inégalité en demandant si les préférences distributives normatives affectent la fourniture d'incitations dans les organisations. Dans cette section, je résumerai brièvement la littérature sur la manière dont les

---

<sup>24</sup>Ce modèle pourrait donc expliquer pourquoi les Américains sont trop optimistes en matière de mobilité sociale (Alesina et al., 2018) mais il n'est pas clair comment il peut être utilisé pour expliquer pourquoi les Européens sous-estiment la mobilité sociale dans une société.

préoccupations en matière d'équité et les préférences distributives affectent la demande de redistribution et les résultats et interactions sur le marché du travail. Je me concentre sur ces applications car les chapitres de cette thèse sont motivés par celles-ci. Néanmoins, je tiens à souligner que les opinions sur l'équité et les préférences distributives ont des conséquences au-delà de l'économie politique de la redistribution et des marchés du travail. En particulier, les considérations d'équité et les attitudes d'inégalité se sont révélées cruciales pour comprendre pourquoi les entreprises ne tirent pas toujours parti de leur pouvoir de marché en cas de demande excédentaire (voir, en particulier, Kahneman et al., 1986a), l'observation empirique du comportement de marchandage, ainsi que la confiance interpersonnelle et la coopération (voir, par exemple, Ashraf et al., 2006; De Bruyn and Bolton, 2008; Hedegaard et al., 2021).

**Demande de redistribution** L'une des conséquences (relativement évidente) des attitudes en matière d'inégalité est la demande de redistribution dans une économie, ainsi que le soutien aux partis qui soutiennent plus ou moins la redistribution (voir Alesina and Giuliano, 2011, pour un examen approfondi de la littérature). Il semble raisonnable de supposer que si les individus rejettent l'inégalité, ils devraient alors exiger une réduction de ces inégalités par le biais de la politique. En fait, un grand nombre des études recensées ci-dessus sont explicitement motivées par cette question (par exemple, le modèle théorique d' Alesina and Angeletos, 2005) et utilisent la demande de redistribution comme variable de résultat caractérisant l'opposition aux niveaux d'inégalité avant impôt (par exemple, Durante et al., 2014; Fong, 2001). Tyrann and Sausgruber (2006) étudient comment les opinions sur l'équité influent sur le vote en faveur de la redistribution, montrant que la prise en compte des préférences distributives est cruciale pour comprendre les résultats du vote en laboratoire. Le lien entre les attitudes à l'égard des inégalités et les préférences pour la redistribution peut être expliqué par le canal égoïste et celui de non-égoïste.<sup>25</sup>

Alors que les études empiriques démontrent souvent que les motifs égoïstes sont des prédicteurs importants de la demande de redistribution (voir par exemple Cruces et al., 2013; Alesina and Giuliano, 2011), les preuves sont moins évidentes pour les motifs non-égoïstes. D'une part, la littérature examinée ci-dessus a identifié une corrélation robuste entre les opinions d'équité et les préférences pour la redistribution. Almås et al. (2020), par exemple, montrent qu'une majorité d'échantillons de Norvégiens et d'Américains déclarent que l'équité est un facteur déterminant pour choisir leurs représentants politiques. Ils constatent également une forte corrélation entre les opinions en matière d'équité et la demande de redistribution. Cette corrélation a été reproduite dans une étude, encore inédite de Almås et al. (2021) qui porte sur des pays du monde entier. En outre, Epper et al. (2020), montrent comment les préférences de respect des autres se transforment en préférences politiques de redistribution.

Des études connexes ont établi un lien entre les préférences distributives et les préférences politiques. Les Américains soucieux d'égalité, par exemple, étaient plus susceptibles d'être affiliés

<sup>25</sup>Dans le canal égoïste, c'est précisément la conviction de gagner ou de perdre à la redistribution qui motive les attitudes à l'égard des inégalités existantes (ou attendues). Dans le canal non-égoïste, ce sont les préoccupations d'équité qui déterminent les préférences en matière de redistribution. Par exemple, on s'attend à ce que les individus ayant une aversion pour l'inégalité demandent davantage de redistribution qu'un méritocrate qui pense que l'effort est un prédicteur important du revenu avant impôt.

au parti démocrate et de voter pour Obama en 2012 (Fisman et al., 2017). Cette corrélation pourrait s'expliquer par les différentes approches de la politique de redistribution. Toutefois, les récentes enquêtes expérimentales qui exploitent la variation exogène de paramètres pertinents tels que la perception des inégalités (Kuziemko et al., 2015) ou les croyances en matière de mobilité (Alesina et al., 2018) ne parviennent pas à identifier qu'un changement d'attitude à l'égard des inégalités existantes se traduise par un changement des préférences politiques. Cela pourrait être dû soit à une méfiance plus générale à l'égard de la capacité du gouvernement à mettre en œuvre ces politiques, soit au fait que la variation exogène des attitudes à l'égard des inégalités n'est pas suffisamment forte pour modifier également les préférences politiques.<sup>26</sup>

**Attitudes en matière d'inégalité sur le lieu de travail** Dans la littérature empirique sur les attitudes à l'égard de l'inégalité, en particulier dans la littérature sur les préférences sociales, il existe une forte tradition consistant à se concentrer sur l'inégalité sur le lieu de travail. Depuis Akerlof and Yellen (1988), Frank (1984) et Kahneman et al. (1986b), les économistes se sont demandé comment les préoccupations d'équité pouvaient expliquer les rigidités que nous observons sur les marchés du travail. En outre, les modèles de préférences sociales trouvent souvent une application directe dans des contextes où les agents interagissent en petits groupes ; ce qui est particulièrement pertinent sur le lieu de travail. Les attitudes à l'égard de l'inégalité salariale sur le lieu de travail ont des implications de grande portée. Comme je le montrerai ci-dessous, l'inégalité salariale peut avoir des effets négatifs sur l'offre de travail, tant sur la marge intensive que sur la marge extensive. Cela peut avoir des conséquences financières importantes, sans compter les coûts psychologiques que les travailleurs peuvent subir en raison d'une perte de satisfaction au travail. Néanmoins, cela peut aussi signifier que les employeurs ou les gérants cherchent à éviter de grands écarts de salaires entre leurs employés dans le but de maximiser les profits, ce qui impliquerait une réduction globale des niveaux d'inégalité avant impôt en raison de la compression des salaires.

Les preuves issues d'expériences en laboratoire peuvent être classées en deux catégories. Une première série d'articles utilise généralement le jeu d'échange de cadeaux (Fehr et al., 1993) pour étudier comment les préoccupations d'équité affectent l'offre de travail. Bien que ce jeu a été conçu pour capturer des motifs réciproques, plutôt qu'une préoccupation pure pour l'inégalité, on peut soutenir que les travailleurs sont préoccupés par les inégalités dans le partage du loyer. En d'autres termes, un agent peut refuser de fournir un effort si le contrat profite à l'employeur de manière disproportionnée. La littérature expérimentale, étudiée par Charness and Kuhn (2011), a montré que les agents se soucient de l'équité des salaires, et qu'ils refusent d'exercer un effort s'ils pensent que l'employeur abuse de sa situation en offrant des salaires injustes. Dans cette situation, le motif de réciprocité est probablement plus fort que le motif d'inégalité pure, puisque la motivation de l'employeur est la clé du comportement des travailleurs. Une autre série d'articles utilisant des expériences en laboratoire étudie comment

<sup>26</sup>En outre, la volonté de réduire les inégalités sociétales par le biais de la redistribution peut dépendre des croyances concernant la perte d'efficacité de ces politiques (Okun, 1975). Cependant, la section 2.1 a montré que des études récentes ont trouvé un faible soutien pour les préoccupations d'efficacité dans des échantillons représentatifs. Cela semble se traduire par des contextes plus explicites dans l'évaluation de la demande de politiques redistributives (voir par exemple Stantcheva, 2020).

l'inégalité horizontale entre les travailleurs affecte l'offre de travail des agents. En théorie, les agents ayant une aversion pour l'inégalité suppriment l'offre de travail si les inégalités deviennent trop importantes (Englmaier and Wambach, 2010; Bartling and Von Siemens, 2010) et il existe certaines preuves empiriques qui le démontre, en particulier pour les inégalités désavantageuses (par exemple Clark et al., 2010; Bracha et al., 2015; Gächter and Thöni, 2010; Greiner et al., 2011; Ku and Salmon, 2012) mais d'autres trouvent que l'inégalité entre les travailleurs n'affecte pas les salaires (e.g. Hennig-Schmidt et al., 2010; Bartling and von Siemens, 2011; Charness and Kuhn, 2007; Gächter et al., 2012). La raison exacte de cette disparité dans la littérature expérimentale n'est toujours pas claire, mais on peut supposer que la motivation de l'employeur et la justification (par exemple la chance ou l'effort) de l'inégalité salariale sont essentielles pour expliquer tout effet négatif de l'inégalité sur l'offre de travail en laboratoire (Bracha et al., 2015; Gross et al., 2015).

Une littérature relativement importante en économie du travail a simultanément évolué pour étudier les effets de l'inégalité salariale à l'aide d'expériences sur le terrain. Ces études trouvent généralement des effets plus forts de l'inégalité salariale sur le comportement et la satisfaction au travail.<sup>27</sup> Card et al. (2012) utilisent une expérience en terrain, dans laquelle ils informent de manière aléatoire certains employés de l'Université de Californie sur un site web où ils peuvent consulter les revenus de leurs collègues. Les auteurs constatent que les travailleurs traités dont le revenu est inférieur à la médiane font état d'une plus faible satisfaction professionnelle et salariale et d'une plus forte propension à quitter leur emploi. Ces résultats sont cohérents avec un modèle dans lequel les travailleurs se soucient de l'inégalité défavorable. Une expérience de terrain récente menée par Cullen and Perez-Truglia (2018) informe de manière aléatoire les travailleurs d'une grande entreprise d'Asie du Sud-Est des salaires de leurs collègues, en faisant varier la hiérarchie. Ils constatent que l'inégalité salariale horizontale a des effets négatifs sur l'offre de travail à marge intensive s'ils apprennent que leurs pairs gagnent plus, alors qu'elle conduit les individus à travailler plus dur s'ils apprennent que leur manager gagne plus qu'ils ne le pensaient initialement. Cette dichotomie entre effets positifs de l'inégalité sur le comportement pour l'inégalité verticale et effets négatifs sur le comportement pour l'inégalité horizontale fait écho à des résultats similaires issus de la littérature sur le bien-être (Clark et al., 2009; Godechot and Senik, 2015). Par ailleurs, Cohn et al. (2014) mènent une expérience de terrain où les travailleurs sont appariés en équipes de deux. Si le salaire d'un seul des travailleurs est réduit, la réduction de l'effort est plus importante que lorsque le salaire des deux travailleurs est réduit. Ces résultats sont cohérents avec un modèle où les sujets ont une aversion pour les inégalités désavantageuses. Breza et al. (2018) dirigent l'expérience de terrain la plus complète qui étudie l'effet des écarts de salaire. Ils montrent que les réponses des travailleurs à l'inégalité salariale dépendent du fait que l'inégalité peut être justifiée par des différences de productivité (visible) en utilisant une expérience de terrain en Inde. Grâce à un riche dispositif expérimental, ils peuvent exploiter la variation de l'inégalité salariale au sein de l'équipe, la transparence de la productivité ainsi que la composition de l'équipe. En exploitant cette variation, ils identifient un fort effet

---

<sup>27</sup>Cela n'est peut-être pas très surprenant, étant donné que les préoccupations en matière d'équité peuvent être beaucoup plus fortes sur le lieu de travail réel, auquel on est émotionnellement attaché

négalif de l'inégalité sur la production et l'assiduité, en particulier des travailleurs qui souffrent d'une inégalité désavantageuse. Le résultat sur la marge extensive de l'offre de travail (assiduité) est particulièrement frappant car les travailleurs renoncent en effet à une quantité importante de revenus. En exploitant la variation exogène de la capacité à observer la productivité, ils montrent en outre que ces effets négatifs sont atténués si les écarts de salaire peuvent être justifiés par des différences de productivité. Enfin, dans leur enquête finale, ils montrent que l'inégalité salariale a des effets négatifs sur la cohésion sociale au sein du groupe.

Les résultats de ces expériences sur le terrain montrent clairement que l'inégalité sur le lieu de travail peut avoir des effets négatifs sur l'offre de travail et le bien-être. Une caractéristique clé qui détermine si ces effets négatifs se produisent semble être la justification des différences de salaire. Cela implique que nous pouvons observer davantage d'inégalité au sein de l'entreprise si les managers peuvent communiquer efficacement la raison des différences de revenus. D'ailleurs, dès lorsque ces effets négatifs sont particulièrement forts pour les personnes qui gagnent des salaires inférieurs.

Comme le lecteur a pu le remarquer, la plupart de ces travaux portent sur les attitudes des travailleurs en matière d'inégalité. Les préférences distributives des managers ont à peine retenu l'attention de cette littérature. L'étude des préférences distributives des dirigeants constitue la contribution du chapitre 2.

La revue précédente de la littérature sur les attitudes à l'égard de l'inégalité a montré que les individus ont des préférences en matière de distribution des revenus. Ces attitudes ont des implications très diverses sur les préférences en matière de redistribution et sur les résultats du marché du travail. Comme je l'ai montré dans la première partie de la revue, les attitudes à envers l'inégalité sont motivées par différents motifs de protection de soi et des autres. La plupart des motifs proposés ont été identifiés dans des environnements contrôlés au moyen d'expériences en laboratoire ou en ligne. Si ces résultats plaident en faveur de la pertinence de chaque motif, ils révèlent également une pluralité frappante d'attitudes d'inégalité, tant au sein d'une même expérience que par rapport à d'autres expériences. La question se pose alors de savoir ce qui motive cette pluralité. Comme je l'ai montré dans la deuxième partie de l'analyse documentaire, les attitudes à l'égard de l'inégalité sont façonnées par l'environnement économique et les expériences que nous vivons - que ce soit en modifiant les perspectives de revenus auxquelles nous sommes confrontés, en affectant les opinions sur l'équité, ou encore en modifiant les croyances sur la source de l'inégalité. Les chapitres de cette thèse font progresser cette littérature : (i) en montrant comment l'importance des différentes motivations qui forment les attitudes d'inégalité interagissent avec l'environnement de choix ; (ii) en dévoilant l'importance des préférences distributives pour les décisions d'incitation ; et (iii) en évaluant si les croyances sur la source de l'inégalité sont façonnées par les incitations auxquelles nous sommes confrontés.

### **3 Méthodes empiriques utilisées dans cette thèse**

Cette thèse s'appuie sur des *expériences de laboratoire* pour améliorer notre compréhension des attitudes à l'égard de l'inégalité. L'économie expérimentale est une méthode empirique qui étudie

le comportement dans un environnement *contrôlé*. L'expérimentation en tant que méthode empirique a gagné en popularité dans les années 1980 et s'est imposée comme un domaine relativement petit mais influent en économie. Les expériences en laboratoire sont également de plus en plus utilisées en complément ou en combinaison avec d'autres méthodes empiriques (Roth, 2015). La principale différence des expériences en laboratoire par rapport aux autres méthodes empiriques, c'est qu'elles permettent à l'expérimentateur d'exercer un contrôle total sur le processus de génération des données (Jacquemet and L'Haridon, 2018). Cela facilite l'identification des relations causales entre les variables explicatives et les résultats d'intérêt. Elle permet également une mise en correspondance plus directe des choix avec les préférences.

Les attitudes à l'égard de l'inégalité font depuis longtemps l'objet d'études en laboratoire. Cela s'explique en partie par le fait que les expériences ont contribué à briser le paradigme de l'intérêt personnel en économie (voir Güth et al., 1982; Kahneman et al., 1986b, pour les premiers exemples) et ont directement alimenté le développement de modèles de préférences non-égoïstes. En effet, les modèles fondamentaux de préférences sociales utilisent des données expérimentales existantes pour motiver et tester leur modèle (voir par exemple Fehr and Schmidt, 1999) ou fournissent des preuves expérimentales avec leur théorie pour calibrer les paramètres ou tester ses prédictions (voir par exemple Charness and Rabin, 2002).

En outre, comme nous l'avons vu dans la revue de la littérature, plusieurs motivations à l'origine des attitudes envers l'inégalité ont été mises en avant. Ces motivations sont souvent simultanément présentes et pertinentes sur le terrain. Par exemple, des données d'enquêtes non expérimentales peuvent montrer que les personnes ayant des revenus élevés s'opposent à une réduction des inégalités. Cela peut s'expliquer par le fait qu'elles pensent que les inégalités sont dues à des différences dans le travail, par leur refus de payer pour une réduction des inégalités par le biais de la redistribution ou par une véritable préférence pour le maintien de fortes inégalités. Le laboratoire nous permet de créer un cadre unique qui nous permet de démêler ces motivations. Une façon courante de procéder consiste à demander aux sujets de choisir entre différentes distributions de revenus tout en faisant varier l'ensemble des distributions de revenus possibles ( $\Omega$ ) afin de démêler les différentes motivations. C'est l'approche que j'utilise dans les chapitres 1 et 2, pour lesquels je crée des situations où les sujets expérimentaux doivent faire des choix consécutifs entre différentes distributions de revenus ou différents systèmes de rémunération à la pièce. Les choix sont conçus de telle sorte que certains motifs qui peuvent être pertinents pour l'acceptation ou le rejet de l'inégalité sont atténués dans certaines décisions mais deviennent pertinents dans d'autres. Ainsi, on peut distinguer les préoccupations égoïstes liées à l'inégalité des revenus des autres préoccupations et, en fin de compte, identifier des préférences en matière de répartition des revenus.

Les expériences en laboratoire sont, en outre, une méthode efficace pour étudier comment l'environnement économique façonne les attitudes en matière d'inégalité. L'identification est réalisée en observant comment le comportement change après avoir fait varier un aspect spécifique du cadre expérimental tout en conservant les autres caractéristiques constantes. Dans tous les chapitres, j'utilise cette approche en faisant varier des aspects cruciaux de l'environnement. Dans le chapitre 3, par exemple, je cherche à évaluer comment les incitations influencent les croyances

relatives à l'effort et à la chance. Pour identifier la relation entre les incitations et les croyances, je crée des situations contrefactuelles qui sont identiques, à l'exception du fait que dans une situation, un décideur s'attend à faire face à des incitations alors que ce n'est pas le cas dans l'autre situation. Une identification aussi claire est essentielle pour comprendre comment les attitudes envers l'inégalité peuvent changer d'une situation à l'autre.

Quel est l'objectif de l'utilisation des expériences ? Selon la définition désormais classique de Roth (1988), les expériences peuvent être utilisées pour “*parler aux théoriciens*”, “*rechercher des faits*”, et “*chuchoter à l'oreille des princes*”. La première approche utilise des expériences pour tester le pouvoir prédictif de la théorie en créant un environnement qui ressemble au contexte caractérisé dans la théorie, pour comparer ensuite le comportement observé avec la prédiction théorique. La deuxième approche utilise des expériences pour explorer les régularités empiriques qui génèrent de nouvelles connaissances pouvant ensuite servir de base à de nouvelles théories. Elle peut également être considérée comme la “*vue scientifique*” citée par PromiseSuccessLab2015, car les résultats des expériences en laboratoire nous aident à mieux comprendre le comportement en dehors du laboratoire. La troisième approche utilise les preuves expérimentales comme base de conseil, par exemple sur des questions politiques. Les chapitres de ma thèse s'inscrivent pour la plupart dans la deuxième catégorie : l'objectif de ces chapitres est de dévoiler des régularités comportementales qui peuvent ensuite être utilisées pour informer davantage de théorie ou bien mieux pour comprendre le comportement sur le terrain. Le chapitre 2, par exemple, montre que les préférences distributives sont des déterminants importants de l'utilisation des incitations. Ainsi, les futurs modèles du marché du travail devraient tenir compte de l'importance des préférences relatives à d'autres facteurs dans le choix du système d'incitation. Le chapitre 3 répond également au premier objectif en testant un mécanisme qui avait déjà été proposé par la théorie.

Quelles sont les limites des méthodes empiriques que j'utilise dans ma thèse ? L'une des principales limites des expériences de « recherche de faits » est leur validité externe. Il est pertinent de se demander si les résultats et les idées de mes expériences se reproduisent dans d'autres contextes. Étant donné la perte inhérente à la caractérisation de tous les détails du monde réel qui accompagne l'abstraction nécessaire d'une expérience de laboratoire, il est impossible de tester pleinement la validité externe de manière empirique (Jacquemet and L'Haridon, 2018). Néanmoins, cela mérite d'être débattu. L'une des préoccupations est que les effets de traitement que j'identifie en laboratoire ne se traduisent pas dans le monde réel. Cela peut être dû à l'hétérogénéité de l'effet du traitement dans les différentes populations. Bien qu'on ne puisse jamais l'exclure complètement, il existe désormais de nombreuses preuves que le signe des corrélations trouvées avec des bases de sujets standard en laboratoire se reproduit dans des bases de sujets plus diversifiés (Snowberg and Yariv, 2021). En outre, je tiens à souligner que certains des effets documentés dans ma thèse ont déjà été identifiés dans la littérature empirique non expérimentale mais n'ont pas fait l'objet d'une interprétation causale. En ce sens, mes expériences servent de complément puissant qui devrait être utilisé comme une preuve de concept alimentant les conclusions suggestives faites dans la littérature précédente. Une autre dimension de l'importance de la validité externe ne réside pas dans les différences de corrélations

mais dans les différences de préférences, c'est-à-dire les niveaux. Cet aspect est, sans doute, plus difficile à satisfaire et les conclusions de mon deuxième chapitre sont particulièrement sensibles à ce problème. Les managers étudiés en laboratoire peuvent être très différents de ceux des entreprises, en raison de l'auto-sélection des individus dans les postes et les professions. Dans la section finale du chapitre 2, nous discuterons donc de cette limite de manière plus détaillée.

## 4 Résumé des chapitres

Les pages suivantes présentent un résumé détaillé de chaque chapitre.

### **Preferences over income distribution : Evidence from a choice experiment**

Dans ce chapitre co-écrit avec Sophie Cètre, Claudia Senik et Thierry Verdier, nous contribuons à répondre à la question de recherche global en évaluant comment les préférences distributives, c'est-à-dire la façon dont une personne souhaite que les revenus soient distribués, se révèlent selon le contexte du choix. Nous nous concentrons sur trois aspects : i) le critère de dominance au sens de Pareto (si une distribution des revenus permet à chacun d'être faiblement mieux loti par rapport à l'autre distribution) ii) si les choix sont faits derrière le voile d'ignorance (sans connaître ses futures circonstances de vie) ou avec une position connue dans la distribution des revenus, et iii) si les rémunérations relatives sont basées sur le mérite ou la chance. Nous utilisons une expérience qui consiste en une série de choix entre deux projets qui aboutissent à des distributions de "bonus" différents. Plus précisément, nous demandons aux sujets de faire une série de choix binaires entre deux distributions de bonus pour un groupe de cinq personnes (le sujet et quatre autres participants anonymes du laboratoire). Nous faisons varier l'origine de la position dans la distribution (en fonction de la chance ou d'une tâche requérant un certain niveau d'effort). La distribution peut être dominante au sens de Pareto ou non par rapport à l'autre. Nous demandons également aux sujets de choisir successivement derrière le voile d'ignorance, donc sans connaître leur rang et leur gain futurs, puis en connaissance de leur position au sein de leur groupe.

La série de choix binaires que les sujets doivent faire peut être divisée en deux catégories. Dans la première catégorie de choix, le gain total est le même dans les deux projets proposés, mais une des répartitions est plus inégale en haut et en bas de la distribution. Dans la deuxième catégorie de choix, le projet le plus inégal domine au sens de Pareto le projet plus égalitaire, c'est-à-dire qu'il améliore faiblement la situation de tous les membres du groupe en termes absolus. Enfin, nous assignons aléatoirement les sujets à deux traitements : le groupe "Mérite" et le groupe "Chance". Dans le traitement "Mérite", la position des personnes au sein de leur groupe de cinq personnes est déterminée par leur performance à une tâche à accomplir après que les choix aient été faits derrière le voile d'ignorance. Dans le traitement "Chance", le classement est déterminé de manière aléatoire.

Notre principal résultat est que, derrière le voile d'ignorance, les sujets préfèrent unanimement le projet aux inégalités plus élevées lorsqu'il est dominant au sens de Pareto. Dans ce cas, il n'y a pas de différence entre les sujets appartenant au traitement "Chance" ou au traitement



“Mérite”. L’unanimité ne disparaît qu’une fois que les positions au sein des distributions de revenus sont fixées, c’est-à-dire lorsque les sujets connaissent leur propre classement avant de choisir entre les deux distributions. Dans ce cas, environ 75% des sujets préfèrent la distribution dominante au sens de Pareto à une distribution des revenus plus comprimée. Les 25% restants préfèrent saboter la situation des plus aisés en supprimant de l’argent en haut de la distribution via le choix du projet plus égalitaire, même si cela n’améliore pas le sort des bas salaires. De plus, lorsque les sujets choisissent entre deux distributions qui ont la même efficacité (même gain total), environ 65% d’entre eux préfèrent la distribution plus égalitaire. Lorsqu’ils choisissent derrière le voile d’ignorance, les sujets sont nettement plus susceptibles d’adopter la distribution inégalitaire s’ils sont dans le groupe “Mérite” plutôt que le groupe “Chance”. Cet effet du traitement disparaît dès que les sujets apprennent leur position dans la distribution, et 70% d’entre eux préfèrent des inégalités plus faibles si cela n’affecte pas leur propre gain. Tous les sujets qui sont mieux lotis dans la distribution plus égalitaire choisissent cette dernière, mais seulement 80% des sujets qui obtiendraient un gain plus avantageux dans la distribution plus inégalitaire choisissent cette dernière. Par conséquent, 20% des individus sont fortement opposés aux inégalités et agissent en conséquence, même si cela a un coût personnel.

### **Principals distributive preferences and the incentivization of agents**

Dans le deuxième chapitre, qui est un travail conjoint avec Sophie Cêtre, nous nous demandons si les préférences distributives des dirigeants ou des managers affectent l’allocation des incitations au sein des entreprises. Le point de départ de notre analyse est la conjecture selon laquelle les préférences distributives des managers peuvent interférer avec la mise en œuvre de systèmes d’incitation qu’un manager peut considérer comme maximisant le profit ou la production. Par exemple, un égalitariste peut être réticent à mettre en œuvre des incitations de type tournoi à forte puissance, car celles-ci impliquent de très grandes inégalités.

Nous montrons qu’il existe une corrélation solide entre les préférences distributives des cadres dirigeants et les structures incitatives de leurs entreprises. Nous utilisons une enquête française réalisée auprès de 4 000 employeurs et cadres dirigeants qui comprend un ensemble de questions relatives aux rémunérations des travailleurs. Nous montrons que lorsque les cadres pensent qu’une politique de salaires individualisés peut être injuste, ils sont moins enclins à mettre en place une rémunération basée sur la performance. Nous montrons que la relation perd de sa force mais reste importante et statistiquement significative lorsque nous incluons des motifs stratégiques pour utiliser ou éviter la rémunération à la performance tels que le fait de croire que ce type de rémunération motive les travailleurs ou qu’elle est au contraire susceptible de créer des tensions, la prévalence des syndicats, etc. Cette corrélation persiste également après l’inclusion d’un large éventail de contrôles spécifiques aux caractéristiques des cadres dirigeants et de leurs entreprises.

Il est compliqué d’établir un lien de causalité dans un tel contexte. Pour contourner ce problème, nous menons une expérience en laboratoire de type principal-agent, en randomisant les sujets pour qu’ils occupent des postes de managers (principal) ou de travailleurs (agent). Chaque principal est associé à deux travailleurs qui diffèrent en fonction de leurs niveaux de

compétence. Les deux travailleurs choisissent un niveau d'effort coûteux pour produire un bien, et le niveau d'effort ne peut être contractualisé. Les managers choisissent entre plusieurs contrats de rémunération à la pièce pour les deux travailleurs. Ces taux à la pièce génèrent une part variable de la rémunération basée sur la performance pour chaque travailleur. Nous attribuons aléatoirement le principal (le manager) soit à un groupe de *Stakeholders* (son revenu est proportionnel à la production des travailleurs), soit à un groupe de Spectateurs (son revenu est fixe). Les Spectateurs peuvent donc mettre en œuvre leur répartition des revenus préférée sans frais, ce qui donne une mesure de l'idéal normatif d'équité du principal. Dans le groupe *Stakeholder*, le principal est incité à tenir compte de la motivation des travailleurs s'il veut augmenter la production commune et ainsi maximiser son propre revenu. Cela donne une mesure de la propension à payer des managers pour mettre en place la répartition des revenus qu'ils préfèrent. La différence de comportement entre ces deux groupes permet d'isoler les préférences normatives en matière de distribution.

En outre, notre cadre nous permet de déterminer avec précision l'importance relative de trois idéaux d'équité (un résultat égalitaire, efficace ou un traitement équitable). Les contrats salariaux à la pièce constituent une innovation par rapport à la littérature existante, car la comparaison des taux à la pièce choisis pour chaque travailleur, en fonction de son niveau de compétence, conduit à une classification directe en trois types de préférences distributives. Le choix de récompenser le travailleur à haut niveau d'aptitude par un taux à la pièce plus élevé témoigne d'une volonté de privilégier l'efficacité puisque, dans notre contexte, cette approche maximise la production. Récompenser les deux travailleurs avec le même taux à la pièce implique de les payer proportionnellement à la production qu'ils ont réalisée. Cela conduit à une équité procédurale puisque les deux travailleurs sont traités de la même manière avec le même salaire à la pièce. Enfin, accorder une rémunération à la pièce plus élevée au travailleur à faible capacité témoigne d'un souci d'égalité, puisque les différences de productivité seront compensées. Nous calibrons ces contrats égalitaires de manière à ce que si les deux travailleurs exercent le même niveau d'effort, ils recevront le même salaire final. Cela revient à une situation plutôt commune dans les entreprises où les travailleurs reçoivent le même salaire car ils évoluent au même poste, même s'ils ne produisent pas les mêmes quantités.

Notre analyse prend en compte deux paramètres importants : (i) est-ce que les agents choisissent un niveau d'effort optimal par rapport à la rémunération à la pièce qui leur est proposée ? (ii) est-ce que le principal anticipe correctement ce comportement ? Avant de demander au principal de choisir les contrats salariaux qu'il souhaite proposer à ses travailleurs, nous lui demandons d'anticiper les réactions des travailleurs quand ceux-ci feront face aux différents niveaux de rémunérations à la pièce. Cela nous permet d'avoir un contrôle sur l'arbitrage efficacité-égalité auquel le principal pense faire face avant que les travailleurs ne se mettent à travailler.

Nous constatons que même dans un contexte d'entreprise très marqué dans cette expérience (possible effet d'identité) et une situation d'aléa moral, les managers ont des préoccupations égalitaires. Ils sont, en moyenne, prêts à faire des compromis pour privilégier une diminution des inégalités au sein de l'entreprise, au prix d'une production plus faible. Cette volonté est bien moindre s'ils sont dans le groupe des *Stakeholders* et c'est également moins le cas lorsque l'enjeu

de l'arbitrage entre efficacité et égalité augmente. Les *Stakeholders* sont aussi plus sensibles à ces incitations à la marge que les Spectateurs. Lorsque l'alternative au contrat qui maximise la production (fortes inégalités) est le contrat favorisant un traitement équitable (plutôt que le contrat égalitaire), les managers ne sont pas plus susceptibles de le choisir en moyenne.

Cela indique qu'une procédure équitable en tant que telle n'est pas considérée comme une caractéristique contractuelle exceptionnellement attrayante et que les managers sont plus intéressés par les résultats distributifs finaux.

Nous effectuons une analyse de l'hétérogénéité des profils-type des managers en ce qui concerne leurs préférences distributives, à l'aide d'un modèle structurel. Nous assignons les managers à l'un des trois types suivants : (1) ceux focalisés sur la production qui privilégient toujours le contrat qui maximise la production conjointe. Ce type de principal n'attache aucune importance au bien-être des agents ; (2) ceux favorables à une redistribution élevée, et qui donc vont attacher une grande importance au revenu de l'agent à faible capacité, et (3) un groupe intermédiaire qui attache une importance significative au revenu de l'agent à faible capacité seulement si la différence de taux à la pièce devient trop défavorable pour cet agent.

Les estimations structurelles nous permettent de faire des estimations contrefactuelles pour modéliser l'implication de ces trois types de préférences sur les performances de l'entreprise dans des contextes légèrement différents de ceux de l'expérience. Nous pouvons par exemple modéliser une situation où les agents détiennent des préférences sociales horizontales, alors que dans notre expérience, nous éliminons ce mécanisme. Les simulations contrefactuelles qui modifient les préférences des travailleurs montrent que les préoccupations égalitaires ne sont pas toujours associées à une perte de profit pour l'entreprise. Des principes sophistiqués de maximisation de la production imiteront le comportement des principes égalitaires parce qu'ils font en fin de compte les choix les plus efficaces si les travailleurs sont égalitaires. Mais lorsque les managers sont naïfs et n'actualisent pas leurs attentes en matière d'effort, ceux qui ont des préférences égalitaires obtiennent de meilleurs résultats pour des niveaux modérés d'aversion aux inégalités des agents.

#### 4.1 Motivating Beliefs in a Just World

Le troisième chapitre s'intéresse à la croyance en la chance et l'effort. Comme je l'ai montré dans la revue de la littérature, ces croyances sont des prédicteurs importants de l'acceptation de l'inégalité, mais notre connaissance de la manière dont elles se forment est encore limitée. Ce chapitre contribue à répondre à cette problématique plus large en étudiant si les individus déforment les croyances chance-effort pour motiver l'effort. J'appelle cette forme de croyances motivées des croyances *motivantes*. Bénabou and Tirole (2002) proposent que les individus déforment les croyances pour contrer une sous-provision d'effort due à des problèmes de contrôle de soi. Ces mêmes auteurs ont créé un modèle qui montre comment les croyances motivantes affectent les préférences pour la redistribution (Bénabou and Tirole, 2006). Plus précisément, ils ont montré que, si un agent économique ayant des problèmes de maîtrise de soi s'attend à de faibles niveaux de redistribution, le fait d'exercer de faibles niveaux d'effort peut devenir très coûteux. Cela crée une demande pour motiver l'effort futur en déformant les croyances.

Pour valider ce modèle, il est important de fournir des preuves empiriques de la distorsion des croyances motivantes, car ces preuves montreraient que les croyances relatives à la chance et à l'effort sont façonnées par les attentes concernant les niveaux futurs de redistribution. Ces preuves nous permettraient de mieux comprendre l'interaction dynamique entre l'inégalité, les préférences de redistribution et les croyances sur l'importance de la chance et de l'effort. De telles preuves impliqueraient également que la relation de cause à effet entre les croyances en la chance et l'effort et la redistribution va dans les deux sens : les croyances affectent la demande de redistribution et les niveaux attendus de redistribution affectent les croyances en façonnant les incitations.<sup>28</sup>

Pour tester la prédiction selon laquelle les incitations futures déforment les croyances sur l'importance de l'effort dans la réussite, j'utilise une expérience en ligne. Dans cette expérience, dont les étapes sont caractérisées par la figure 4.3, les sujets commencent par effectuer une tâche lourde d'effort réel. Cette tâche est réalisée dans un environnement incertain, où la règle de paiement dépend de l'état du monde (Environnement) qui a été fixé au début de l'expérience. Si le sujet est dans la condition Performance-Environnement, la probabilité de gagner un prix pour avoir accompli la tâche est une fonction croissante de sa performance dans la tâche. Plus précisément, le sujet participe à un tournoi contre un concurrent tiré au sort, où la probabilité de gagner le prix est égale à 80% si le sujet transcrit plus d'images que son concurrent, tandis que la probabilité est égale à 20% s'il transcrit moins d'images. Si le sujet est dans la condition Chance-Environnement, sa performance dans la tâche n'a aucun effet sur sa probabilité de gagner le prix ; le sujet gagne le prix avec une probabilité de 50% quel que soit le nombre de mots qu'il transcrit.

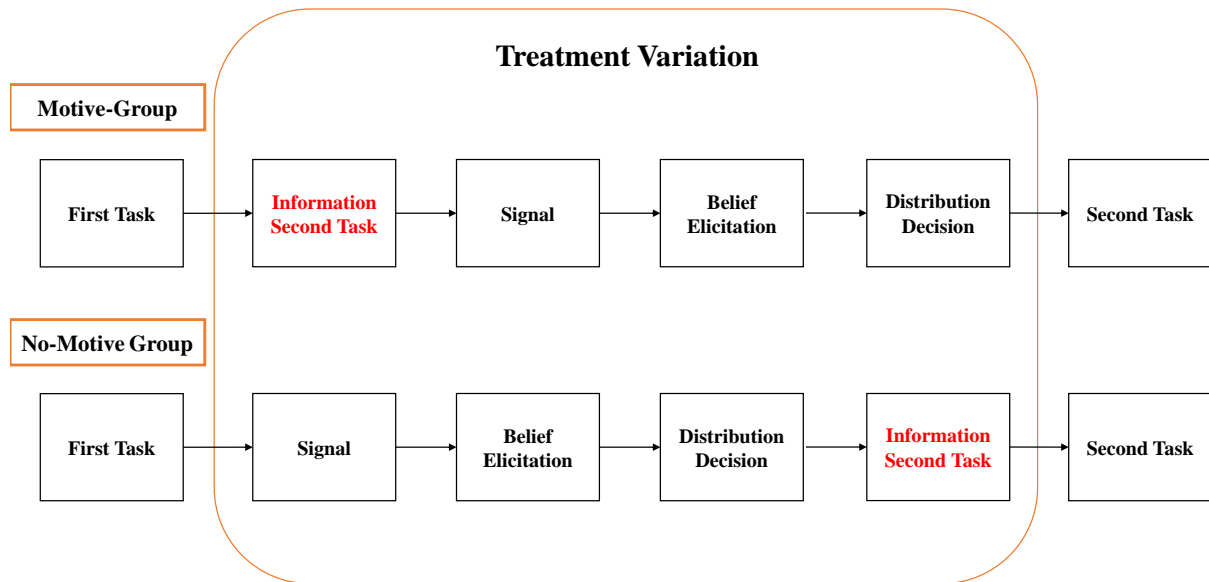
Une fois la tâche terminée, les sujets reçoivent un signal sonore qui les informe de la condition dans laquelle ils se trouvaient (environnement de performance ou environnement de chance). Le but de ce signal est d'induire une variation dans les croyances de base concernant l'état du monde. J'induis ces croyances en donnant aux sujets un feedback sur le résultat de la première tâche. Le feedback comprend deux éléments d'information : (1) si un sujet a gagné le prix et (2) s'il a transcrit plus ou moins d'images que son concurrent. À l'aide de ces informations, les sujets peuvent former des croyances postérieures sur la probabilité de se trouver dans l'environnement de chance ou de performance. Par exemple, un sujet qui apprend qu'il a transcrit plus d'images que son concurrent, mais qu'il n'a pas gagné le prix, est susceptible de se percevoir comme ayant une forte probabilité d'être dans l'environnement chance, c'est-à-dire la condition dans laquelle le succès n'est pas lié à l'effort. En revanche, si la même personne apprend qu'elle a gagné le prix, elle devrait se percevoir comme ayant une forte probabilité de se trouver dans l'environnement de

---

<sup>28</sup>Les croyances affectent les préférences en matière de redistribution de plusieurs manières : premièrement, il existe des raisons égoïstes de demander une redistribution moindre si l'on déforme les croyances pour motiver les efforts futurs. Un individu qui croit qu'il sera un contributeur net si son effort est reflété dans la distribution du revenu avant impôt serait moins susceptible de soutenir la redistribution après s'être engagé dans une distorsion de croyance motivante. Deuxièmement, la distorsion des croyances motivantes peut affecter les préférences pour la redistribution pour d'autres raisons. Les individus méritocratiques qui déforment leurs croyances de manière motivante sont moins susceptibles de croire qu'une distribution initialement inégale est due à la chance et, s'ils acceptent des inégalités qui reflètent des différences d'effort, ils devraient, par conséquent, opter pour une redistribution moindre par rapport à une situation où ils ne déforment pas leurs croyances pour motiver un effort futur.

la performance, où l'effort influe sur la probabilité de réussite.<sup>29</sup> Suite à la réception du signal, je demande aux sujets leurs croyances probabilistes sur l'environnement (chance ou performance) dans lequel ils se trouvent.

FIGURE 4.3 : Schéma caractérisant les étapes de l'expérience



Pour vérifier si les sujets déforment les croyances de base pour se motiver à faire des efforts, j'introduis une deuxième tâche que les sujets peuvent accomplir à la fin de l'expérience et qui sert d'incitation - et, par conséquent, de motif - pour déformer les croyances à des fins de motivation. Comme pour la première tâche, la règle de paiement dépend de l'environnement qui a été établi au début de l'expérience : si le sujet se trouve dans l'environnement Performance, l'effort détermine si l'on reçoit une récompense pour la performance à la deuxième tâche, tandis que pour les sujets dans l'environnement Chance, l'effort n'a aucun effet sur la probabilité de recevoir une récompense. Les sujets qui connaissent la deuxième tâche peuvent surestimer leur probabilité d'être dans l'environnement de performance afin de se motiver à travailler dur sur la deuxième tâche.

Pour identifier la distorsion des croyances motivantes, je fais varier le moment où j'informe les sujets de la deuxième tâche : les sujets qui sont assignés au groupe de traitement Motif sont informés de la seconde tâche *avant* l'élicitation des croyances et, par conséquent, sont incités à déformer les croyances pour motiver l'effort. Les sujets du groupe de traitement sans motivation reçoivent cette information *après* la sollicitation des croyances. Ces derniers ne sont pas incités à déformer les croyances pour motiver l'effort, car ils ne savent pas qu'ils devront accomplir une

<sup>29</sup>Le signal imite des expériences réelles que les individus peuvent utiliser pour déduire l'importance de l'effort pour réussir dans la vie : Un collègue peut obtenir une promotion même si l'on se considère comme plus talentueux et plus productif que la personne qui a obtenu cette promotion ; d'autres personnes peuvent obtenir un poste auquel elles ont postulé, tout en sachant qu'elles n'ont obtenu ce poste qu'en raison de leurs liens personnels avec le PDG de l'entreprise ; d'autres personnes encore peuvent remporter un prix pour leur travail, tout en sachant qu'elles ont travaillé plus dur et obtenu de meilleurs résultats que les autres personnes présélectionnées pour le prix.

deuxième tâche dans l'expérience. Cette variation me permet de tester l'hypothèse principale de l'expérience : les sujets du groupe Motif, qui savent qu'ils devront accomplir une deuxième tâche, sont, en moyenne, plus confiants de se trouver dans l'environnement de performance que les sujets du groupe sans motivation qui ne savent pas qu'ils devront fournir un effort à l'avenir.

Ce design me permet de tester la distorsion des croyances motivantes de manière non paramétrique en comparant les croyances postérieures des deux groupes de traitement. Néanmoins, je peux aller plus loin et demander quel type de signal entraîne une distorsion de croyance motivante. Premièrement, je peux demander si les sujets sont plus ou moins susceptibles de s'engager dans une distorsion de croyance motivante lorsqu'ils reçoivent un signal qui suggère qu'ils se trouvent dans l'environnement de performance plutôt que dans l'environnement de chance. Deuxièmement, le design me permet de tester si les événements qui ne sont pas informatifs sur l'état réel du monde affectent les croyances. Plus précisément, je demande si les individus sont plus susceptibles de croire à l'importance de l'effort s'ils ont gagné une récompense - plutôt que perdu -, en maintenant constant le contenu informationnel de l'événement. Cela me permet de déduire quels types d'événements induisent une distorsion motivante des croyances.

Par ailleurs, je me demande si la distorsion des croyances motivantes affecte les décisions de redistribution entre deux autres individus, car ces croyances sont de puissants prédicteurs de la demande de redistribution pour les individus méritocratiques. Après la première phase de l'expérience, au cours de laquelle les sujets reçoivent un signal concernant l'environnement auquel ils ont été affectés (c'est-à-dire Chance ou Performance) et au cours de laquelle les sujets du groupe Motivation sont informés de la seconde tâche, je donne aux sujets la possibilité de redistribuer une allocation de prime initialement inégale entre deux participants non impliqués. Ces participants ont été préalablement recrutés pour effectuer la même première tâche que la décideuse elle-même. J'explique ouvertement aux participants que l'allocation initiale a été déterminée par la même règle de paiement à laquelle ils viennent eux-mêmes d'être confrontés. Les sujets peuvent ensuite redistribuer cette allocation initiale. En exploitant la variation entre les groupes de traitement et les signaux, je peux tester (a) si la distorsion de la croyance motivante affecte l'acceptation de l'inégalité pour les motifs de considération des autres et (b) si les expériences passées affectent les décisions de redistribution au-delà du contenu informationnel des expériences.

Mes résultats montrent que les sujets déforment leurs croyances pour motiver leurs efforts futurs. Les sujets qui savent qu'ils vont effectuer une autre tâche dans le même environnement sont significativement plus convaincus (sept points de pourcentage) de se trouver dans l'environnement de performance. Cet effet moyen masque l'hétérogénéité par type de signal. Les sujets du groupe de motivation qui ont reçu un signal désincitatif indiquant que la récompense n'est pas liée à l'effort, c'est-à-dire qu'ils ont été affectés à l'environnement chance, sont significativement plus convaincus (neuf points de pourcentage) qu'ils se trouvent dans l'environnement performance par rapport aux sujets du groupe de contrôle qui ont reçu le même signal. Mes résultats ne montrent aucune différence dans les croyances entre les groupes de traitement pour les sujets qui ont reçu un signal incitatif indiquant que la récompense dépend de l'effort, c'est-à-dire pour ceux qui ont été affectés à l'environnement de performance. Cela montre que la distorsion des

croyanances motivantes est particulièrement fréquente si les personnes reçoivent des informations désincitatives, c'est-à-dire des informations qui indiquent que l'effort n'est pas important pour réussir. En exploitant la variation indépendante de l'événement qui conduit à un signal donné, je montre que la distorsion des croyances motivantes est particulièrement prononcée pour les individus qui savent (ou croient) qu'ils auraient bien réussi dans un monde qui récompense réellement l'effort, c'est-à-dire pour les personnes qui ont appris que le résultat de la tâche n'était pas justifié par leur performance relative. Dans l'ensemble, les résultats de mon expérience fournissent des preuves solides que les individus déforment leurs propres croyances en matière de chance et d'effort afin de se motiver pour accomplir la tâche à laquelle ils s'attendent à être confrontés à l'avenir.

En ce qui concerne les résultats relatifs à la décision de distribution, je montre que la distorsion des croyances motivantes n'affecte pas de manière significative le comportement de distribution. Cela suggère que les croyances peuvent jouer un rôle dans la motivation de l'effort futur, mais que ce changement dans les croyances n'est pas assez fort pour se refléter dans le comportement de distribution global. Même si les sujets qui sont sûrs d'être dans l'environnement de performance sont moins susceptibles de redistribuer, je constate que les expériences passées ont tendance à avoir une grande importance pour les décisions de redistribution dans ce contexte. Il est important de noter que les sujets qui n'ont pas gagné de prix et qui ont obtenu résultats de moins bons que leur concurrent redistribuent des montants plus importants que les autres sujets, même si les premiers ont reçu un signal indiquant une plus grande probabilité d'être dans l'environnement de performance. Ce résultat souligne l'importance de prendre en compte lors de l'analyse du comportement distributif les caractéristiques de l'événement qui ne sont pas informatives sur l'importance relative de la chance et de l'effort.

## 5 Références

- Agranov, Marina and Thomas R. Palfrey. The effects of income mobility and tax persistence on income redistribution and inequality. *European Economic Review*, 123 :103372, 2020.
- Akbaş, Merve, Dan Ariely, and Sevgi Yuksel. When is inequality fair? An experiment on the effect of procedural justice and agency. *Journal of Economic Behavior & Organization*, 161 : 114–127, 2019.
- Akerlof, George A and Janet L Yellen. Fairness and Unemployment. *American Economic Review*, 78(2) :44–49, 1988.
- Alesina, Alberto and George-Marios Angeletos. Fairness and redistribution : US vs. Europe. *American Economic Review*, 95(5) :960–980, 2005.
- Alesina, Alberto and Paola Giuliano. Chapter 4 - Preferences for Redistribution. In Benhabib, Jess, Alberto Bisin, and Matthew O. Jackson, editors, *Handbook of Social Economics*, volume 1, pages 93–131. North-Holland, 2011.
- Alesina, Alberto and Eliana La Ferrara. Preferences for redistribution in the land of opportunities. *Journal of Public Economics*, 89(5-6) :897–931, 2005.
- Alesina, Alberto, Edward L Glaeser, and Bruce Sacerdote. Why Doesn't the United States Have a European-Style Welfare State? *Brookings Papers on Economic Activity*, 2001(2) :187–277, 2001.
- Alesina, Alberto, Stefanie Stantcheva, and Edoardo Teso. Intergenerational Mobility and Preferences for Redistribution. *American Economic Review*, 108(2) :521–554, 2018.
- Alesina, Alberto, Armando Miano, and Stefanie Stantcheva. The Polarization of Reality. *AEA Papers and Proceedings*, 110 :324–328, 2020.
- Almås, Ingvild, Alexander W Cappelen, Kjell G Salvanes, Erik Ø Sørensen, and Bertil Tungodden. Fairness and family background. *Politics, Philosophy & Economics*, 16(2) :117–131, 2017.
- Almås, Ingvild, Alexander W. Cappelen, and Bertil Tungodden. Cutthroat Capitalism versus Cuddly Socialism : Are Americans More Meritocratic and Efficiency-Seeking than Scandinavians? *Journal of Political Economy*, 128(5) :1753–1788, 2020.
- Almås, Ingvild, Alexander W Cappelen, Bertil Tungodden, and Erik Ø. Sørensen. Fairness Across the World : Preferences and Beliefs. *Working Paper*, 2021.
- Alvaredo, Facundo, Anthony B Atkinson, Thomas Piketty, and Emmanuel Saez. The Top 1 Percent in International and Historical Perspective. *Journal of Economic Perspectives*, 27(3) : 3–20, 2013.



- Amasino, Dianna, Davide D Pace, and Joël J. van der Weele. Fair Shares and Selective Attention. *Working Paper*, 2021.
- Andreoni, James and John Miller. Giving According To Garp : An Experimental Test of the Consistency of Preferences for Altruism. *Econometrica*, 70(2) :737–753, 2002.
- Arneson, Richard J. Equality and Equal Opportunity for Welfare. *Philosophical Studies : An International Journal for Philosophy in the Analytic Tradition*, 56(1) :77–93, 1989.
- Ashok, Vevikin, Ilyana Kuziemko, and Ebonya Washington. Support for redistribution in an age of rising inequality : New stylized facts and some tentative explanations. *Brookings Papers on Economic Activity*, pages 367–405, 2015.
- Ashraf, Nava, Iris Bohnet, and Nikita Piankov. Decomposing trust and trustworthiness. *Experimental Economics*, 9(3) :193–208, 2006.
- Atkinson, Anthony B, Thomas Piketty, and Emmanuel Saez. Top Incomes in the Long Run of History. *Journal of Economic Literature*, 49(1) :3–71, 2011.
- Babcock, Linda, George Loewenstein, Samuel Issacharoff, and Colin Camerer. Biased Judgments of Fairness in Bargaining. *The American Economic Review*, 85(5) :1337–1343, 1995.
- Barr, Abigail, Luis Miller, and Paloma Ubeda. Moral consequences of becoming unemployed. *Proceedings of the National Academy of Sciences*, 113(17) :4676–4681, 2016.
- Barron, Kai, Robert Stüber, and Roel van Veldhuizen. Motivated motive selection in the lying-dictator game. *Working Paper*, 2020.
- Bartling, Björn and Ferdinand A Von Siemens. The intensity of incentives in firms and markets : Moral hazard with envious agents. *Labour Economics*, 17(3) :598–607, 2010.
- Bartling, Björn and Ferdinand A. von Siemens. Wage inequality and team production : An experimental analysis. *Journal of Economic Psychology*, 32(1) :1–16, 2011.
- Bauer, Michal, Alessandra Cassar, Julie Chytilová, and Joseph Henrich. War’s Enduring Effects on the Development of Egalitarian Motivations and In-Group Biases. *Psychological Science*, 25(1) :47–57, 2014.
- Bellemare, Charles, Sabine Kröger, and Arthur Van Soest. Measuring inequity aversion in a heterogeneous population using experimental decisions and subjective probabilities. *Econometrica*, 76(4) :815–839, 2008.
- Bénabou, Roland and Efe A Ok. Social mobility and the demand for redistribution : The POUM hypothesis. *The Quarterly Journal of Economics*, 116(2) :447–487, 2001.
- Bénabou, Roland and Jean Tirole. Self-Confidence and Personal Motivation. *The Quarterly Journal of Economics*, 117(3) :871–915, 2002.

- Bénabou, Roland and Jean Tirole. Belief in a Just World and Redistributive Politics. *The Quarterly Journal of Economics*, 121(2) :699–746, 2006.
- Bentham, Jeremy. *An Introduction to the Principles of Morals and Legislation*. T. Payne, 1789.
- Bjerk, David. In front of and behind the veil of ignorance : An analysis of motivations for redistribution. *Social Choice and Welfare*, 47(4) :791–824, 2016.
- Bolton, Gary E and Axel Ockenfels. ERC : A theory of equity, reciprocity, and competition. *American Economic Review*, 90(1) :166–193, 2000.
- Bracha, Anat, Uri Gneezy, and George Loewenstein. Relative Pay and Labor Supply. *Journal of Labor Economics*, 33(2) :297–315, 2015.
- Breza, Emily, Supreet Kaur, and Yogita Shamdasani. The morale effects of pay inequality. *The Quarterly Journal of Economics*, 133(2) :611–663, 2018.
- Bruhin, Adrian, Ernst Fehr, and Daniel Schunk. The many faces of human sociality : Uncovering the distribution and stability of social preferences. *Journal of the European Economic Association*, 17(4) :1025–1069, 2018.
- Buser, Thomas, Gianluca Grimalda, Louis Putterman, and Joël van der Weele. Overconfidence and gender gaps in redistributive preferences : Cross-Country experimental evidence. *Journal of Economic Behavior & Organization*, 178 :267–286, 2020.
- Cappelen, Alexander, John List, Anya Samek, and Bertil Tungodden. The Effect of Early-Childhood Education on Social Preferences. *Journal of Political Economy*, 128(7) :2739–2758, 2020a.
- Cappelen, Alexander W. and Bertil Tungodden. Rewarding effort. *Economic Theory*, 39(3) :425–441, 2009.
- Cappelen, Alexander W, Astri Drange Hole, Erik Ø Sørensen, and Bertil Tungodden. The pluralism of fairness ideals : An experimental approach. *The American Economic Review*, 97(3) :818–827, 2007.
- Cappelen, Alexander W, James Konow, Erik Ø Sørensen, and Bertil Tungodden. Just luck : An experimental study of risk-taking and fairness. *The American Economic Review*, 103(4) :1398–1413, 2013.
- Cappelen, Alexander W., Ranveig Falch, and Bertil Tungodden. Fair and Unfair Income Inequality. In Zimmermann, Klaus F., editor, *Handbook of Labor, Human Resources and Population Economics*, pages 1–25. Springer International Publishing, Cham, 2020b.
- Cappelen, Alexander W., Ranveig Falch, Erik Ø. Sørensen, and Bertil Tungodden. Solidarity and fairness in times of crisis. *Journal of Economic Behavior & Organization*, 186 :1–11, 2021.

- Card, David, Alexandre Mas, Enrico Moretti, and Emmanuel Saez. Inequality at work : The effect of peer salaries on job satisfaction. *The American Economic Review*, 102(6) :2981–3003, 2012.
- Carlsson, Fredrik, Dinky Daruvala, and Olof Johansson-Stenman. Are People Inequality-Averse, or Just Risk-Averse? *Economica*, 72(287) :375–396, 2005.
- Cassar, Lea and Arnd H. Klein. A Matter of Perspective : How Failure Shapes Distributive Preferences. *Management Science*, 65(11) :5050–5064, 2019.
- Charness, G. and M. Rabin. Understanding Social Preferences with Simple Tests. *The Quarterly Journal of Economics*, 117(3) :817–869, 2002.
- Charness, Gary and Peter Kuhn. Does Pay Inequality Affect Worker Effort? Experimental Evidence. *Journal of Labor Economics*, 25(4) :693–723, 2007.
- Charness, Gary and Peter Kuhn. Lab Labor : What Can Labor Economists Learn from the Lab? *Handbook of Labor Economics*, 4 :229–330, 2011.
- Checchi, Daniele and Antonio Filippin. An experimental study of the POUM hypothesis. In *Inequality, Welfare and Income Distribution : Experimental Approaches*, pages 115–136. Emerald Group Publishing Limited, 2004.
- Cherry, Todd L, Peter Frykblom, and Jason F Shogren. Hardnose the Dictator. *The American Economic Review*, 92(4) :1218–1221, 2002.
- Clark, Andrew E., Nicolai Kristensen, and Niels Westergård-Nielsen. Job satisfaction and co-worker wages : Status or signal? *The Economic Journal*, 119(536) :430–447, 2009.
- Clark, Andrew E, David Masclet, and Marie Claire Villeval. Effort and comparison income : Experimental and survey evidence. *Industrial and Labor Relations Review*, 63(3) :407–426, 2010.
- Cohn, Alain, Ernst Fehr, Benedikt Herrmann, and Frédéric Schneider. Social comparison and effort provision : Evidence from a field experiment. *Journal of the European Economic Association*, 12(4) :877–898, 2014.
- Cooper, David J. and John H. Kagel. Other-Regarding Preferences : A selective Survey of Experimental Results. In Kagel, John H and Alvin E Roth, editors, *The Handbook of Experimental Economics*, Vol. 2, pages 2017–2282. Princeton University Press, Princeton, 2016.
- Cruces, Guillermo, Ricardo Perez-Truglia, and Martin Tetaz. Biased perceptions of income distribution and preferences for redistribution : Evidence from a survey experiment. *Journal of Public Economics*, 98 :100–112, 2013.
- Cullen, Zoe and Ricardo Perez-Truglia. How Much Does Your Boss Make? The Effects of Salary Comparisons. *Working Paper*, 2018.

- Dalton, Hugh. The Measurement of the Inequality of Incomes. *The Economic Journal*, 30(119) : 348, 1920.
- Dana, Jason, Roberto A. Weber, and Jason Xi Kuang. Exploiting moral wiggle room : Experiments demonstrating an illusory preference for fairness. *Economic Theory*, 33(1) :67–80, 2007.
- De Bruyn, Arnaud and Gary E. Bolton. Estimating the Influence of Fairness on Bargaining Behavior. *Management Science*, 54(10) :1774–1791, 2008.
- Deffains, Bruno, Romain Espinosa, and Christian Thöni. Political self-serving bias and redistribution. *Journal of Public Economics*, 134 :67–74, 2016.
- Di Tella, Rafael, Sebastian Galiani, and Ernesto Schargrotsky. The Formation of Beliefs : Evidence from the Allocation of Land Titles to Squatters. *The Quarterly Journal of Economics*, 122(1) :209–241, 2007.
- Duesenberry, James S. *Income, Saving, and the Theory of Consumer Behavior*. Harvard University Press, first edition edition, 1949.
- Durante, Ruben, Louis Putterman, and Joël van der Weele. Preferences for Redistribution and Perception of Fairness : An Experimental Study. *Journal of the European Economic Association*, 12(4) :1059–1086, 2014.
- Dworkin, Ronald. *Sovereign Virtue : The Theory and Practice of Equality*. Harvard University Press, Cambridge, 2002.
- Engel, Christoph. Dictator games : A meta study. *Experimental Economics*, 14(4) :583–610, 2011.
- Engelmann, Dirk and Martin Strobel. Inequality aversion, efficiency, and maximin preferences in simple distribution experiments. *The American Economic Review*, 94(4) :857–869, 2004.
- Engelmann, Dirk and Martin Strobel. Preferences over income distributions experimental evidence. *Public Finance Review*, 35(2) :285–310, 2007.
- Englmaier, Florian and Achim Wambach. Optimal incentive contracts under inequity aversion. *Games and Economic Behavior*, 69(2) :312–328, 2010.
- Epper, Thomas, Ernst Fehr, and Julien Senn. Other-Regarding Preferences and Redistributive Politics. *Working Paper*, 2020.
- Fehr, Dietmar and Martin Vollmann. Misperceiving Economic Success : Experimental Evidence on Meritocratic Beliefs and Inequality Acceptance. *Working Paper*, 2020.
- Fehr, E., G. Kirchsteiger, and A. Riedl. Does Fairness Prevent Market Clearing? An Experimental Investigation. *The Quarterly Journal of Economics*, 108(2) :437–459, 1993.

- Fehr, Ernst and Klaus M Schmidt. A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, 114(3) :817–868, 1999.
- Fehr, Ernst, Michael Naef, and Klaus M Schmidt. Inequality aversion, efficiency, and maximin preferences in simple distribution experiments : Comment. *The American Economic Review*, 96(5) :1912–1917, 2006.
- Festinger, Leon. *A Theory of Cognitive Dissonance*. Stanford University Press, 1957.
- Fisman, Raymond, Shachar Kariv, and Daniel Markovits. Individual preferences for giving. *American Economic Review*, 97(5) :1858–1876, 2007.
- Fisman, Raymond, Pamela Jakiela, and Shachar Kariv. How did distributional preferences change during the Great Recession? *Journal of Public Economics*, 128 :84–95, 2015a.
- Fisman, Raymond, Pamela Jakiela, Shachar Kariv, and Daniel Markovits. The distributional preferences of an elite. *Science*, 349(6254), 2015b.
- Fisman, Raymond, Pamela Jakiela, and Shachar Kariv. Distributional preferences and political behavior. *Journal of Public Economics*, 155 :1–10, 2017.
- Fisman, Raymond, Pamela Jakiela, Shachar Kariv, and Silvia Vannutelli. The Distributional Preferences of Americans. *Working Paper*, 2020.
- Fleurbaey, Marc. *Fairness, Responsibility, and Welfare*. Oxford University Press, 2008.
- Fong, Christina. Social preferences, self-interest, and the demand for redistribution. *Journal of Public Economics*, 82(2) :225–246, 2001.
- Forsythe, Robert, Joel L. Horowitz, N.E. Savin, and Martin Sefton. Fairness in Simple Bargaining Experiments. *Games and Economic Behavior*, 6(3) :347–369, 1994.
- Frank, Robert H. Are Workers Paid their Marginal Products? *The American Economic Review*, 74(4) :549–571, 1984.
- Frank, Robert H. *Success and Luck : Good Fortune and the Myth of Meritocracy*. Princeton University Press, 2017.
- Frohlich, Norman, Joe A. Oppenheimer, and Cheryl L. Eavey. Choices of Principles of Distributive Justice in Experimental Groups. *American Journal of Political Science*, 31(3) :606, 1987.
- Frohlich, Norman, Joe Oppenheimer, and Anja Kurki. Modeling other-regarding preferences and an experimental test. *Public Choice*, 119(1-2) :91–117, 2004.
- Gächter, Simon and Christian Thöni. Social comparison and performance : Experimental evidence on the fair wage–effort hypothesis. *Journal of Economic Behavior & Organization*, 76 (3) :531–543, 2010.

- Gächter, Simon, Daniele Nosenzo, and Martin Sefton. The Impact of Social Comparisons on Reciprocity : Impact of social comparisons on reciprocity. *The Scandinavian Journal of Economics*, 114(4) :1346–1367, 2012.
- Gärtner, Manja, Johanna Mollerstrom, and David Seim. Individual risk preferences and the demand for redistribution. *Journal of Public Economics*, 153 :49–55, 2017.
- Gärtner, Manja, Johanna Möllerström, and David Seim. Income Mobility, Luck/Effort Beliefs, and the Demand for Redistribution : Perceptions and Reality. *Working Paper*, 2019.
- Giuliano, Paola and Antonio Spilimbergo. Growing up in a Recession. *The Review of Economic Studies*, 81(2) :787–817, 2014.
- Godechot, Olivier and Claudia Senik. Wage comparisons in and out of the firm. Evidence from a matched employer-employee French database. *Journal of Economic Behavior & Organization*, 117 :395–410, 2015.
- Greiner, Ben, Axel Ockenfels, and Peter Werner. Wage transparency and performance : A real-effort experiment. *Economics Letters*, 111(3) :236–238, 2011.
- Grosfeld, Irena and Claudia Senik. The emerging aversion to inequality. *Economics of Transition*, 18(1) :1–26, 2010.
- Gross, Till, Christopher Guo, and Gary Charness. Merit pay and wage compression with productivity differences and uncertainty. *Journal of Economic Behavior and Organization*, 117 : 233–247, 2015.
- Güth, Werner, Rolf Schmittberger, and Bernd Schwarze. An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior & Organization*, 3(4) :367–388, 1982.
- Harsanyi, John C. Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *Journal of Political Economy*, 63(4) :309–321, 1955.
- Hedegaard, Morten, Rudolf Kerschbamer, Daniel Müller, and Jean-Robert Tyran. Distributional preferences explain individual behavior across games and time. *Games and Economic Behavior*, 128 :231–255, 2021.
- Hennig-Schmidt, Heike, Abdolkarim Sadrieh, and Bettina Rockenbach. In Search of Workers’ Real Effort Reciprocity—a Field and a Laboratory Experiment. *Journal of the European Economic Association*, 8(4) :817–837, 2010.
- Hirschman, Albert O and Michael Rothschild. The Changing Tolerance for Income Inequality in the Course of Economic Development\* With A Mathematical Appendix. *The Quarterly Journal of Economics*, 87(4) :544–566, 1973.
- Hoffman, Elizabeth, Kevin McCabe, Keith Shachat, and Vernon Smith. Preferences, property rights, and anonymity in bargaining games. *Games and Economic behavior*, 7(3) :346–380, 1994.

- Jacquemet, Nicolas and Olivier L'Haridon. *Experimental Economics : Method and Applications*. Cambridge University Press, Cambridge, 2018.
- Jakiela, Pamela. How fair shares compare : Experimental evidence from two cultures. *Journal of Economic Behavior & Organization*, 118 :40–54, 2015.
- Johansson-Stenman, Olof, Fredrik Carlsson, and Dinky Daruvala. Measuring Future Grandparents' Preferences for Equality and Relative Standing. *The Economic Journal*, 112(479) : 362–383, 2002.
- Kahneman, Daniel, Jack L. Knetsch, and Richard Thaler. Fairness as a Constraint on Profit Seeking : Entitlements in the Market. *The American Economic Review*, 76(4) :728–741, 1986a.
- Kahneman, Daniel, Jack L. Knetsch, and Richard H. Thaler. Fairness and the Assumptions of Economics. *The Journal of Business*, 59(4) :S285–S300, 1986b.
- Konow, James. A positive theory of economic fairness. *Journal of Economic Behavior & Organization*, 31(1) :13–35, 1996.
- Konow, James. Fair Shares : Accountability and Cognitive Dissonance in Allocation Decisions. *American Economic Review*, 90(4) :1072–1091, 2000.
- Konow, James. Which Is the Fairest One of All? A Positive Analysis of Justice Theories. *Journal of Economic Literature*, 41(4) :1188–1239, 2003.
- Kosse, Fabian, Thomas Deckers, Pia Pinger, Hannah Schildberg-Hörisch, and Armin Falk. The Formation of Prosociality : Causal Evidence on the Role of Social Environment. *Journal of Political Economy*, 128(2) :434–467, 2020.
- Kranton, Rachel, Matthew Pease, Seth Sanders, and Scott Huettel. Deconstructing bias in social preferences reveals groupy and not-groupy behavior. *Proceedings of the National Academy of Sciences*, 117(35) :21185–21193, 2020.
- Kritikos, Alexander and Friedel Bolle. Distributional concerns : Equity-or efficiency-oriented? *Economics Letters*, 73(3) :333–338, 2001.
- Krosnick, Jon A. and Duane F. Alwin. Aging and susceptibility to attitude change. *Journal of Personality and Social Psychology*, 57(3) :416–425, 1989.
- Ku, Hyejin and Timothy C. Salmon. The Incentive Effects of Inequality : An Experimental Investigation. *Southern Economic Journal*, 79(1) :46–70, 2012.
- Kuziemko, Ilyana, Michael I Norton, Emmanuel Saez, and Stefanie Stantcheva. How elastic are preferences for redistribution? Evidence from randomized survey experiments. *The American Economic Review*, 105(4) :1478–1508, 2015.
- Ledyard, John O. Public Goods : A survey of Experimental Research. In Kagel, John H and Alvin E Roth, editors, *The Handbook of Experimental Economics*, pages 111–149. Princeton University Press, Princeton, 1995.

- Lefgren, Lars J., David P. Sims, and Olga B. Stoddard. Effort, luck, and voting for redistribution. *Journal of Public Economics*, 143 :89–97, 2016.
- Lerner, Melvin J. The Belief in a Just World. In Lerner, Melvin J., editor, *The Belief in a Just World : A Fundamental Delusion*, Perspectives in Social Psychology, pages 9–30. Springer US, Boston, MA, 1980.
- Luttmer, Erzo F P. Group loyalty and the taste for redistribution. *Journal of political Economy*, 109(3) :500–528, 2001.
- Luttmer, Erzo F. P and Monica Singhal. Culture, Context, and the Taste for Redistribution. *American Economic Journal : Economic Policy*, 3(1) :157–179, 2011.
- Meltzer, Allan H. and Scott F. Richard. A Rational Theory of the Size of Government. *Journal of Political Economy*, 89(5) :914–927, 1981.
- Michelbach, Philip A, John T Scott, Richard E Matland, and Brian H Bornstein. Doing Rawls justice : An experimental study of income distribution norms. *American Journal of Political Science*, 47(3) :523–539, 2003.
- Mollerstrom, Johanna, Bjørn-Atle Reme, and Erik Ø Sørensen. Luck, choice and responsibility – An experimental study of fairness views. *Journal of Public Economics*, 131 :33–40, 2015.
- Müller, Daniel and Sander Renes. Fairness views and political preferences : Evidence from a large and heterogeneous sample. *Social Choice and Welfare*, 56(4) :679–711, 2021.
- Nielsen, Kai. *Equality and Liberty : A Defense of Radical Egalitarianism*. Rowman & Littlefield, 1985.
- Obama, Barack. Remarks by the President on Economic Mobility. *The White House Office of the Press Secretary*, 2013.
- Okun, Arthur M. *Equality and Efficiency : The Big Tradeoff*. Brookings Institution Press, 1975.
- Owens, Lindsay A. and David S. Pedulla. Material Welfare and Changing Political Preferences : The Case of Support for Redistributive Social Policies. *Social Forces*, 92(3) :1087–1113, 2014.
- Piketty, Thomas. Social mobility and redistributive politics. *The Quarterly Journal of Economics*, 110(3) :551–584, 1995.
- Piketty, Thomas. *Capital in the Twenty-First Century*. Harvard University Press, Cambridge, 2014.
- Piketty, Thomas. *Capital and Ideology*. Harvard University Press, Cambridge, 2020.
- Rawls, John. *A Theory of Justice*. Harvard university press, Cambridge, 1971.
- Roemer, John E. *Equality of Opportunity*. Harvard University Press, Cambridge, 1998.



- Romer, Thomas. Individual welfare, majority voting, and the properties of a linear income tax. *Journal of Public Economics*, 4(2) :163–185, 1975.
- Roth, Alvin E. Laboratory Experimentation in Economics : A Methodological Overview. *The Economic Journal*, 98(393) :974, 1988.
- Roth, Alvin E. Is Experimental Economics Living Up to Its Promise ? In Fréchette, Guillaume R. and Andrew Schotter, editors, *Handbook of Experimental Economic Methodology*, pages 13–40. Oxford University Press, 2015.
- Roth, Christopher and Johannes Wohlfart. Experienced inequality and preferences for redistribution. *Journal of Public Economics*, 167 :251–262, 2018.
- Ruffle, Bradley J. More is better, but fair is fair : Tipping in dictator and ultimatum games. *Games and Economic Behavior*, 23(2) :247–265, 1998.
- Saez, Emmanuel and Gabriel Zucman. The Rise of Income and Wealth Inequality in America : Evidence from Distributional Macroeconomic Accounts. *Journal of Economic Perspectives*, 34(4) :3–26, 2020.
- Sandel, Michael J. *The Tyranny of Merit : What's Become of the Common Good?* Farrar, Straus and Giroux, 2020.
- Schildberg-Hörisch, Hannah. Is the veil of ignorance only a concept about risk? An experiment. *Journal of Public Economics*, 94(11) :1062–1066, 2010.
- Sen, Amartya. Utilitarianism and Welfarism. *The Journal of Philosophy*, 76(9) :463–489, 1979.
- Senik, Claudia. When information dominates comparison : Learning from Russian subjective panel data. *Journal of Public Economics*, 88(9) :2099–2123, 2004.
- Shayo, Moses. A Model of Social Identity with an Application to Political Economy : Nation, Class, and Redistribution. *American Political Science Review*, 103(2) :147–174, 2009.
- Snowberg, Erik and Leeat Yariv. Testing the Waters : Behavior across Participant Pools. *American Economic Review*, 111(2) :687–719, 2021.
- Stantcheva, Stefanie. Understanding Tax Policy : How Do People Reason? *Working Paper*, 2020.
- Tyran, Jean-Robert and Rupert Sausgruber. A little fairness may induce a lot of redistribution in democracy. *European Economic Review*, 50(2) :469–485, 2006.
- Valero, Vanessa. Redistribution and Beliefs about the Source of Income Inequality. *Working Paper*, 2020.
- Veblen, Thorstein. *The Theory of the Leisure Class*. Courier Corporation, 1899.
- Yaari, M. E. and M. Bar-Hillel. On Dividing Justly. *Social Choice and Welfare*, 1(1) :1–24, 1984.