Institut d'études politiques de Paris
Ecole Doctorale de Sciences Po
Programme doctoral en économie
Département d'économie
Doctorat en sciences économiques

Essays on the determinants of wage inequality

Sophie Cetre

Thesis supervised by
Yann Algan, Professeur des universités en économie, Sciences Po
Claudia Senik, Professeur des universités en économie, Paris School of Economics

defended on October 16^{th}, 2020

Jury:

Yann Algan	Professeur des universités en économie, Sciences Po *(supervisor)*
Denis Fougère	Directeur de recherche, CNRS-OSC *(referee)*
Elise Huillery	Professeur des universités en économie, Université Paris-Dauphine – PSL
Andrea Ichino	Professor of economics, European University Institute
Nicolas Jacquemet	Professeur des universités en économie, Paris School of Economics *(referee)*
Claudia Senik	Professeur des universités économie, Paris School of Economics *(supervisor)*

Contents

Note to the reader **vii**

Acknowledgments **ix**

Introduction **1**

References . . . 22

1 Updating of academic tastes and ability signals **33**

1.1 Introduction . . . 34

1.2 Context and data . . . 36

1.2.1 Eliciting academic tastes . . . 36

1.2.2 Academic data . . . 39

1.3 Academic tastes predict observed academic choices . . . 41

1.4 Determinants of taste updating . . . 43

1.4.1 Empirical strategy . . . 43

1.4.2 Both student-specific ability signals and dominant high-school tastes predict future tastes . . . 46

1.4.3 Heterogeneity in student characteristics . . . 49

1.4.4 No significant effect of relative ability signals . . . 50

1.5 Conclusion . . . 50

References . . . 54

Appendix to Chapter 1 **57**

1.A Letters of motivation . . . 57

1.A.1 Guidelines for letters of motivation . . . 57

1.A.2 Hand-labeling . . . 58

1.A.3 Keywords . . . 61

1.B Peer groups . . . 69

1.C Typology of student heterogeneity . . . 74

1.D Appendix Tables and Figures . . . 77

2 Preferences over Income Distribution: Evidence from a Choice Experiment **83**

2.1 Introduction . . . 84

2.2 Design of the experiment . . . 86

2.2.1 Overview . . . 86

2.2.2 The Choices . . . 87

2.2.3 Merit versus Luck . . . 89

2.2.4 Control variables . . . 89

2.2.5 Implementation . . . 90

2.3 Results . . . 91

2.3.1 Choices behind the veil of ignorance . . . 91
2.3.2 Choices with known position . . . 92
2.3.3 Observable heterogeneity . . . 95
2.4 Within-subject variation . . . 96
2.5 Robustness Checks . . . 96
2.5.1 Are subjects consistent within a category of choices? . . . 96
2.5.2 Is choosing behind the veil of ignorance the same as choosing between two lotteries? 97
2.5.3 Does the significant treatment effect pick up on over-confidence? . . . 97
2.6 Conclusion . . . 98
References . . . 98

Appendix to Chapter 2 **103**
2.A Tables and Figures . . . 104
2.A.1 Tables . . . 104
2.A.2 Figures . . . 110
2.B Hypothetical Decisions . . . 116
2.C Screenshots of the Experiment . . . 117
2.D Post-experimental Questionnaire . . . 122
2.E Experimental Instructions . . . 125
2.E.1 Introduction of the experiment . . . 125
2.E.2 Part 1 . . . 125
2.E.3 Part 2 . . . 126
2.E.4 Real effort task . . . 129
2.E.5 Part 3 . . . 130
2.E.6 Questionnaire *(Luck and Merit groups)* . . . 131
References . . . 131

3 Principals' Distributive Preferences and the Incentivization of Agents **133**
3.1 Introduction . . . 134
3.2 Managers' preferences: survey evidence . . . 137
3.3 Experimental design . . . 139
3.3.1 Lab setting . . . 139
3.3.2 Workers . . . 139
3.3.3 Principals . . . 141
3.3.4 Summary statistics . . . 144
3.4 Results . . . 145
3.4.1 Effort choices and effort beliefs . . . 145
3.4.2 Belief-based contract trade-offs . . . 145
3.4.3 Principals' choices . . . 148
3.5 Structural Characterization of Distributive Preferences . . . 152
3.5.1 Making distributive decisions ex-ante . . . 153
3.5.2 Utility function specification . . . 153
3.5.3 Pooled results . . . 156
3.5.4 Characterizing heterogeneity in preferences . . . 157

3.5.5 Results from the finite mixture model . . . 159
3.5.6 Counterfactual analyses . . . 160
3.6 Conclusion . . . 167
References . . . 167

Appendix to Chapter 3 **171**
3.A Tables and Figures . . . 171
3.B Robustness Checks . . . 180
3.C Deriving the density of the likelihood on the individual level . . . 181
3.D Instructions . . . 181
3.D.1 Principals . . . 182
3.D.2 Workers . . . 184
3.E Comprehension test . . . 186
3.E.1 Questions principals . . . 186
3.E.2 Workers' questions . . . 189
3.E.3 Comprehension test performance . . . 191
3.F Aptitude test . . . 193
3.F.1 French Questions . . . 193
3.F.2 Logic questions . . . 194
3.F.3 General knowledge . . . 194

4 Ethnic bias, economic success, and trust: findings from large sample experiments in Germany and the U.S. **197**
4.1 Introduction . . . 198
4.2 Experimental design . . . 202
4.2.1 Trustlab methodology . . . 202
4.2.2 Duration and payoffs payment . . . 207
4.3 Results . . . 207
4.3.1 Overview of the results . . . 207
4.3.2 Ethnic block . . . 209
4.3.3 Ethnic plus Income block . . . 216
4.4 Discussion . . . 227
4.5 Conclusions . . . 231
References . . . 232

Appendix to Chapter 4 **239**
4.A Figures . . . 239
4.B Tables . . . 241
4.C Behavior consistency . . . 245
4.D Experimental instructions . . . 249

Résumé en français **261**
References . . . 272

List of figures **274**

List of tables **276**

Note to the reader

The four chapters of this dissertation are self-contained research articles and can be read separately. They are preceded by an introduction that explains my research agenda and the main research questions I address. The second chapter has been published:

Cetre, S., Lobeck, M., Senik, C., & Verdier, T. (2019). Preferences over income distribution: Evidence from a choice experiment. *Journal of Economic Psychology*, 74, 102202.

Acknowledgments

I have spent four wonderful years working on this thesis and there is a long list of persons to which I want to express my profound gratitude. They are largely responsible for smoothing the whole process and making it a pleasant, joyful and intellectually fruitful journey.

I want to thank my supervisors, Claudia Senik and Yann Algan, for their guidance. I first met Claudia in 2015, at the end of the first year of my master's degree at the Paris School of Economics. I sent her a long email "*Economie du bonheur questions*", to ask her about her fascinating research topics. We quickly started a collaboration, alongside Andrew Clark, to work on the effect of parenthood on happiness, which led to my first published article in 2016. We kept working together on other projects, such as the second chapter of this thesis, co-authored with Max Lobeck and Thierry Verdier. I have always loved our discussions and how I would walk out of her office with tons of new ideas to explore. She taught me to never forget the big picture, to work on the topics that interest me the most, and she gave me key insights on all my projects. I also warmly thank Yann Algan, who agreed to supervise my thesis at Sciences Po. I benefited a lot from his honest advice and he never hesitated to put me forward in our meetings with other researchers. I am particularly grateful to him for introducing me to the *Trustlab* network, where I had the chance to meet and work with talented researchers.

This thesis also owes immensely to Max Lobeck, with whom I have co-authored Chapters 2 and 3. Research is easy when we can brainstorm with brilliant minds like his. I was amazed by his numerous ideas, his great sense of detail (which is crucial when designing experiments), his perseverance and the incredibly large amount of work he was able to do in short periods of time. These two papers would be nowhere as good as their as now without his input.

I also want to warmly thank the *Trustlab team*, especially Gianluca Grimalda, Fabrice Murtin, Louis Putterman, Vincent Siegerink, Lara Fleischer and Arnstein Aassve. Gianluca and Louis basically taught me how to write a good research article and everyone doing research knows how hard writing can be. I thank Fabrice for his genuine care for junior researchers and his suggestion that I should first-author our article.

I am also grateful to Ghazala Azmat, Anne Boring and Roberto Galbiati, with whom I have co-authored the first chapter of this thesis, and who have also helped and guided me throughout my doctoral years. I also thank our four research assistants, Marguerite Clanché, Jeanne Dorlencourt, Léa Dousset and Guillaume Lagacherie, who have helped us with the data cleaning process.

I have had the chance to visit Stanford University for a semester, thanks to the invitation of Matthew Gentzkow. I am grateful to him and also to Alvin Roth and Paul Oyer for providing feedback on my thesis.

The articles constituting this thesis have also benefited from many discussions that me or my co-authors have had with other researchers, seminar speakers and economists met during conferences. I have special thanks to address to Béatrice Boulu-Reshef, for her detailed comments on the second chapter of this thesis; Nicolas Jacquemet for his regular feedback on Chapters 2 and 3; Radu Vranceanu, for inviting me to the ESSEC to present my research; Julia Cagé, Maxim Frolov, Jeanne Hagenbach and Benjamin Marx. I also thank Stefano DellaVigna, Jana Friedrichsen, Shachar Kariv,

Dorothea Kubler, Fabrice Le Lec and Thierry Mayer.

I am also immensely grateful to the members of the jury, Denis Fougère, Elise Huillery, Andrea Ichino and Nicolas Jacquemet, for accepting to review my thesis. I look forward to receiving their comments that will undoubtedly help improving my research.

Pour finir avec la liste des professeurs, permettez-moi de passer au français afin de remercier ma professeure de Sciences Economiques et Sociales au lycée, Mathilde Cotin Leandri. Elle m'a prise sous son aile dès la 1ère, et m'a fortement encouragée à passer le concours de Sciences Po. Son exigence et sa confiance ont été de profonds moteurs pour moi et son apport a été très certainement décisif dans la réussite de mes études supérieures.

On a different note, research is not only the result of some brain juice. All the research articles constituting this thesis have received financial support from various institutions. Chapter 1 has received funding from the Laboratory for Interdisciplinary Evaluation of Public Policies (LIEPP). Chapter 2, has benefited from the financial support of the European Research Council (ERC TECTACOM), CEPREMAP, and the French National Research Agency, through the program Investissements d'Avenir, ANR-10-LABX-93-01. Chapter 3 has received funding from CEPREMAP. The research leading to the fourth chapter of this thesis has received funding from the European Research Council under the European Union's Horizon 2020 research and innovation programme (Grant Agreement no.647870); from the Kiel Institute for the World Economy for the survey carried out in Germany; and from Brown University, including its Department of Economics and Population Studies and Training Center, for the U.S. survey.

If I have enjoyed so much these four years and if I feel a bit nostalgic right now, it is certainly thanks to all the PhD students at Sciences Po with whom I have worked in *les combles*. My dear friends, so many good memories come to my mind when I think about you! I remember all these nice and restful tea/coffee breaks we have had, which have grown longer and more frequent over time (and we all know who is to blame for that!). We have come through so many things together! We have survived to the tasteless (but cheap) food at the CROUS; we have chased and killed mice together (the perfect team building activity); fought drowsiness in the afternoon seminars armed with liters of tea and coffee; we have organized a secret santa (thanks Camille for the cutest cactus) and welcomed two baby economists (congratulations Etienne and Arthur!). You are all truly amazing human beings, incredibly smart and talented, with unique characters and personalities. I have nourished myself with our conversations and the quality of our debates. You have endured with patience my increasingly annoying and invasive ecological concerns, especially my strong aversion to air-conditioning (I am so sorry Ludovic and Florin for constantly arguing with you about that). I have also found emotional support when needed and benevolent ears to listen to my frustrations (and my gossips). On this aspect in particular, I am tremendously grateful to Elisa, for truly enlightening these years and supporting me on all matters, personal and professional alike. *Mille mercis* also to Julia, who is such a strong and independent woman I look up to; Gustave, whose good mood and optimism are so communicative; Jérôme, because life would me much less fun without him and his peculiar musical tastes; Laurène, for helping me and supporting me during my job search – it was truly comforting to go through the same challenges together. And for all the fun and interesting conversations we have had together I warmly thank Ludovic, Camille, Pauline, Etienne, Edgard, Arthur, Charles, Florin, Nicolò, Vladimir, Jan, Pierre C., Pierre D., Pierre V., Ségal (I am so sorry for the counteless kicks under our desk), Paul, Axelle, Jean-Benoît, Zydney, Max, Victor S., Victor A., Alais, Jean-Louis, Clémence, Aseem,

Edoardo, Lucas, Mario, Stefan, Marco, Daniel, Oliver, and of course, Florence and Anja, my favorite political scientists.

Many thanks to all my students at Sciences Po, they probably don't know how much *they* have taught me.

I am also grateful to my new employer, the *Institut de Radioprotection et de Sûreté Nucléaire* (IRSN), and more particularly the team of the economics laboratory, Ludivine Pascucci-Cahen, Maxime Cavalgante and Felipe Ramirez Goni, for giving me the opportunity to finish and defend this thesis in the best conditions. I am so excited by the new research avenues and challenges that lie ahead!

Je passe au français pour remercier mes amis, pas forcément économistes, heureusement ! Mon amie d'enfance, Carine, qui a suivi une voie tellement différente de la mienne, et qui me permet d'avoir des conversations d'autant plus rafraichissantes. Carmelo, pour son soutien indéfectible dans à peu près tous les aspects de ma vie ; Hannah, pour son ouverture d'esprit et ses propositions de sorties culturelles toujours originales ; Christine, pour avoir partagé mes hésitations au sujet de la recherche en économie. Je ne t'oublie pas, même si tu habites si loin ! Je remercie aussi Agathe, Bérengère et Simon, et également Laureline, Samy, Antoine, Jean-Baptiste, Amandine, Brune, Vincent, Hadrien, Camille, Gilles et Guillaume. Il faut aussi savoir se détendre pendant la thèse, et toutes ces personnes y ont contribué très activement !

J'embrasse aussi mes parents, Michèle et Jean-Claude, qui ont été d'une bienveillance et d'un calme à toute épreuve, aussi loin que mes souvenirs portent. Ils m'ont appris à savoir être heureuse et résiliente, et à trouver mon bonheur en toutes circonstances. Je remercie évidemment ma soeur jumelle, Marion – qui va soutenir sa thèse en médecine en même temps que moi à quelques semaines près – notre complicité fait partie des choses qui sont les plus importantes dans ma vie ; ma soeur aînée, Céline, qui m'a appris la tolérance face à l'altérité ; et mon frère Nicolas, pour m'avoir appris à ne pas être susceptible face aux taquineries ! Je remercie également pour leur soutien mes deux belles-soeurs, Laurence et Audrey, et mon beau-frère Nicolas. J'ai aussi une pensée très affectueuse pour Tatie Claudine et mes deux nièces, Romane et Louise.

Je remercie également ma belle-famille, qui m'a accueillie à bras ouverts et a suivi avec attention toutes les étapes de cette thèse : Frédérique, Pierre, Laurence, Mamie Lili, Papi Jean-Marie et Mamie Odette.

Et pour finir, un immense merci à Jean-François, mon compagnon, confident et partenaire de fous rires, pour tout ce qu'il m'apporte au quotidien, tant sur le plan intellectuel qu'affectif. Si j'ai pu avoir des doutes quant à cette thèse, il n'en a manifestement jamais eus. Son soutien indéfectible, sa confiance aveugle en mes capacités et sa qualité d'écoute exceptionnelle ont été des renforts précieux tout au long de ces années. Je le remercie enfin pour m'avoir lue à haute voix toutes les fables de La Fontaine pendant le confinement, ce qui a constitué une belle source d'inspiration pour la rédaction de l'introduction de cette thèse.

Introduction

La Génisse, la Chèvre et la Brebis, en société avec le Lion.

La Génisse, la Chèvre et leur soeur la Brebis,
Avec un fier Lion, seigneur du voisinage,
Firent société, dit-on, au temps jadis,
Et mirent en commun le gain et le dommage.
Dans les lacs de la Chèvre un cerf se trouva pris.
Vers ses associés aussitôt elle envoie.
Eux venus, le Lion par ses ongles compta,
Et dit: "Nous sommes quatre à partager la proie".
Puis, en autant de parts le cerf il dépeça ;
Prit pour lui la première en qualité de Sire :
"Elle doit être à moi, dit-il, et la raison,
C'est que je m'appelle Lion :
A cela l'on n'a rien à dire.
La seconde, par droit, me doit échoir encor :
Ce droit, vous le savez, c'est le droit du plus fort.
Comme le plus vaillant, je prétends la troisième.
Si quelqu'une de vous touche à la quatrième,
Je l'étranglerai tout d'abord.

Jean de la Fontaine, *Fables*, Livre premier, Fable VI

Inequality at the heart of the economic profession

My main argument throughout this thesis is that income inequality is first and foremost a micro-founded object. Individuals make choices, have preferences, views and perceptions of fairness, agree with, reject or renew societal norms. These forces shape the lives and economic outcomes of each individual through their own choices – constrained or otherwise – and also through the actions that other individuals take. The aggregation and interactions of these microeconomic forces have macroeconomic implications. Collectively, individuals shape ideologies, change economic institutions and political regimes. This is particularly relevant to the topic of income inequality. Before starting to measure aggregated indices such as the Gini coefficient[1], or the share of income that accrues to the top 1% of the income distribution, it is beneficial to reflect upon the underlying microeconomic processes. People choose whether or not to go to college, express their preferences by voting, make choices on the labor market, etc.

The choice of this approach has been, like any research endeavor, deeply influenced by the economic, social, political and academic context surrounding the beginning of my doctoral work. The past few years have witnessed the rise of populist views embodied by radical parties[2], strong civilian discontent such as the Yellow Vest movement in France, and the declining trust in political institutions across several countries (OECD, 2017; Citrin and Stoker, 2018). This troubled context has attracted the interest of researchers active in a particularly rich academic field, further confirming its relevance. My thesis stems largely from the intellectual debate following the pioneering works of Anthony B. Atkinson, Branko Milanovic, Thomas Piketty, Emmanuel Saez, Gabriel Zucman and others. Piketty's *Capital in the Twenty-First Century* was first published in French three years before I started my thesis and in English two years before. New data on wealth and income inequality was released and presented to the whole world – putting figures on what were previously mere feelings or intuitions – which triggered national debates on redistributive policies (Atkinson, 2014; Ostry et al., 2019; Saez and Zucman, 2020). This groundbreaking research had just enough time to permeate the economic profession that, during the last year of my master's degree and the first year of my PhD, research seminars were full of economists who had appropriated the topic in their own way, enriching it with their own sets of methods and data, producing new perspectives and putting forward new questions. Certain authors who were already tackling the issue of income distribution decades ago – like Kuznets (1955) – gained renewed attention.

[1] The Gini coefficient is a synthetic measure of the distribution of income across the population of a country. It ranges from 0 (full equality, everyone has the same income) to 1 (extreme inequality, one person earns everything).

[2] Many scholars have attempted to pinpoint the origins of such dynamics, emphasizing the importance of economic factors (Algan et al., 2017; Guriev, 2018). The rise of income inequality (Piketty, 2013), the polarizing effects of globalization on wages creating winners (high-skilled workers) and losers (low-skilled workers) (Autor et al., 2014, 2016) and the 2008 financial crisis with its dramatic consequences for the least well-off, are triggers of resentment, discontent and mistrust toward the elite and political institutions (Algan et al., 2017, 2019a). This dichotomous vision of the world pitting the people against a spoiled elite is often seen as a consensual minimal definition of populism (Bonikowski and Gidron, 2016; Mudde, 2017). Another hypothesized cause of populism is a cultural shift with the revival of a racial animus that may have been dormant for several decades, masked by the apparent vitality of progressive ideas (Stephens-Davidowitz, 2014; Inglehart and Norris, 2016). Such a cultural backlash, to use the terminology of Norris and Inglehart (2019), may have been amplified by increasing ethnic diversity and the migrant crisis, resulting in an "us vs them" identity redefinition.

Research questions

This dissertation analyzes the course of several micro-events leading individuals to receive different wages. The first question I consider is: *why* are different people paid unequally? I then go on to examine individuals' views about income inequality: *should* different people be paid unequally? The distinction between "different" and "unequal" is important. Wages are unequal in the sense that they vary according to a unique monetary metric, which implies that we can order wage outcomes in a hierarchical way. While wages are unidimensional, individuals vary along multidimensional lines, which cannot be ordered in a hierarchical manner. Individuals can vary in terms of their innate talent, ability, effort, merit, personality etc. As an economist, my goal is to analyze how markets, institutions and individuals map out these multidimensional inputs into a unidimensional outcome: wages. Answering the first question is a lifelong undertaking, and I do not claim to provide an exhaustive answer. I have thus focused my research on three determinants of wage inequality: educational choices (Chapter 1), managers' normative preferences and incentivization concerns (Chapter 3), and ethnic discrimination (Chapter 4).

The second question is also vast, and I want to emphasize that even if it resembles a normative question, it is not. In Chapters 2 and 3, I show how *individuals* answer this question, with their own views and perceptions of fairness and equity – concepts that I will describe at greater length in this introduction. When answering this question, my objective is not only to clarify the social and political preferences of our societies, I also want to describe how these preferences vary from one individual to the next. Accounting for heterogeneity enables us to step back from the dominant narratives and ideologies, and better understand the behavior of the least well-off (Chapter 2).

A short tale of inequality

The chapters of my thesis do not set out to answer these two questions sequentially. Instead, they follow a chronological narrative. Since my thesis is all about the choices and preferences of individuals, it could be entirely summarized in the form of a tale. Chapter 1 begins with a student hesitating between different fields of study. Her choices will strongly influence her professional career and wage path. In high school, she was highly interested in economics and regularly read newspaper articles on this topic. Once at university, she realizes that, even if her grades in economics are very respectable, she is better at political science and decides to follow this track instead. Her friend, on the contrary, sticks to his initial vocation, despite having fairly low grades in his preferred field of study. She wonders why he disregards his grades when she is so performance-driven.

In Chapter 2, she is still pursuing her studies and can vote for the first time in national elections. She has to think about which candidate she would like to see in office. Reading through the different manifestos, she identifies the various underlying ideologies, one of which considers that people should be paid according to their relative merits and talents, which makes the resulting economic inequality fair. Another view considers that even in a context of merit-based inequality, differences in income may still be indecently large, which warrants State intervention to redistribute income. She carries out a short thinking exercise in order to make up her mind, inspired by John Rawls' *Theory of Justice*, which she has been reading in her political science classes. She wonders about the type of society she would like to grow up in if she had no idea about her future life circumstances, talents, and family environment. The realization dawns and she knows for whom she will vote.

In Chapter 3, she gets her first job and meets her first boss. Everyone in her team is paid almost the same wage, while in her friend's firm, wages vary widely and are indexed on performance. While chatting with him, she realizes that their respective bosses have significantly different views about what constitutes a fair wage distribution. Her own boss considers that people doing the same job should be paid the same, while her friend's boss thinks that employees need monetary incentives. She realizes that these different views of fairness have major implications in terms of the firms' entire wage structures and management styles.

Finally, in Chapter 4, she has more and more trouble working with certain colleagues, who seem to distrust her for some reason. She has been the target of racist jokes, and fears that her skin color might explain her white colleagues' unwillingness to cooperate with her, which affects her performance at work. She talks to her boss about this and they devise a strategy together: her boss will publicly recognize her hard work and merit in order to put an end to the discrimination she is confronted with. It works – her white colleagues respect her more – but the strategy backfires when coworkers of her own ethnicity become less friendly with her and jealous of her success.

What conclusions can we draw from this tale? Income inequality stems from the aggregation of millions of individual economic, social and political choices, made at different points in time. Moreover, these choices are shaped by people's preferences, perceptions of fairness and beliefs. Finally, people do not only make choices that influence their own wage path, they also take actions that can be beneficial or detrimental to the people with whom they interact.

Before outlining each chapter of my thesis, I will return to the broader framework in which my research is embedded.

Research framework

My objective in the next pages is to clarify the connections between the chapters of my thesis and present the research that has been most influential to my work. This effort to establish links aims to provide a comprehensive framework of analysis for my dissertation. I will also revisit the main concepts underlying my thesis.

The normative view on the level of economic inequality

One of Thomas Piketty's claims is that the *level* of inequality is too high in many countries, both in terms of the share of income accruing at the top – which is the primary focus of his *Capital in the Twenty-First Century* – and in terms of the income differences between those at the top and the bottom of the distribution (Piketty et al., 2018).[3] but it would be an oversimplification to infer from his work that the *level* of inequality is the only factor that matters.

First, several economists have argued that people are largely ignorant of the true levels of inequality and income distributions in their countries (Gimpelson and Treisman, 2018). Hauser and Norton (2017) show that individuals typically underestimate the extent of inequality in the US and UK, while overestimation is more common in France and Germany. Estimating the levels of national inequality is a hard task, and many individuals actually use heuristic judgments and gauge the level of inequality surrounding them (Cruces et al., 2013). In the end, it is not so much the level of inequality that

[3]Regarding the latter, he refers to the abundant literature on poverty, its determinants and the means to alleviate it (Sen, 1992; Duflo and Banerjee, 2011),

correlates with preference for redistribution and social conflicts, but rather the *perceived* inequality (Gimpelson and Treisman, 2018). This stream of research requires a behavioral approach in order to question individuals' beliefs, perceptions, and preferences.

Second, people may care as much (or even more) about the *origin* of inequality as its level (Starmans et al., 2017). Atkinson (2014) hints at this in his paper *After Piketty?*. Even if he clearly states that the level of inequality should be addressed: *"I take the widespread popular reaction to Capital as an indication that many people agree that the present level of inequality is intolerable, and not justified by the common good."*, he explains that equality of opportunity, and more precisely *competitive* equality of opportunity[4] is often defined as the objective. According to him, *"even if there were competitive equality of opportunity, the reward structure is too unequal and [...] ex post inequality needs to be reduced."*. As many economists have pointed out, equality of opportunity is rarely achieved, even in rich countries (Causa and Johansson, 2010; Chetty et al., 2014, 2017).[5] The fact that family background, neighborhood and ethnicity correlate so well with adults' income is the sign that equality of opportunity is still elusive. If people do not start out with the same chances, the resulting income distribution will often be deemed *unfair*. Piketty (1995)'s early work already theorized that the extent of equality of opportunity, measured through individuals' social mobility experiences, is a major determinant of political orientation.

Is inequality fair? The contribution of behavioral economics

People often have a gut feeling that a student who carries out an internship in her father's company, or a young adult who rents several apartments after receiving an early inheritance, do not fully *deserve* their privileged situations. The fact that other people have not shared their good fortune seems *unfair*. Experimental and behavioral economists, as well as philosophers, have pondered this question, formulating explicit theories and frameworks underlying the vague terms of "fairness" and "social justice". How can we define them, after all? Konow (2001) writes: *"What is remarkable, and perhaps unique, about justice is that, despite the strong sentiments and vigorous actions it incites, no consensus has emerged about what it is"*.[6] We will see that the difficulty with the concept of fairness is that there are as many definitions of fairness as there are individuals. Considerable effort has been made in experimental economics to characterize this heterogeneity in the form of equity principles.

Early work and methods

The early work on fairness by psychologists and experimental economists was not directly related to inequality. It was part of a sometimes provocative movement whose intended goal was to shake up standard economic theory as a whole. As Kahneman et al. (1986) put it: *"The absence of considerations of fairness and loyalty from standard economic theory is one of the most striking contrasts between this body of theory and other social sciences – and also between economic theory and lay intuitions about human behavior."*.[7] Kahneman et al. (1986) ask in this paper whether it would be in-

[4] Competitive equality of opportunity is defined as a situation in which everyone has "an equal chance to take place in a race", in contrast to non-competitive opportunity in which everyone has "an equal chance to fulfill their – independent – life projects".

[5] High income inequality could also cause low social mobility and thus impede equality of opportunity (Corak, 2013).

[6] In this respect, the definition provided by the Oxford dictionary of fairness is also very vague: "*the quality of treating people equally or in a way that is reasonable*".

[7] The idea that people are other-regarding and not completely selfish is not entirely new in economics. Several prominent economists developed theories incorporating these aspects, such as Smith (1759) and Becker (1974). See Fehr

sightful to augment economic theory with considerations of fairness, generosity and rebellion, since *"it is often viewed as an embarrassment to the basic theory that people vote, do not always free ride, and commonly allocate resources equitably to others and to themselves when they are free to do otherwise"*. A substantial body of anecdotal evidence shows that fairness considerations matter in everyday life[8], Kahneman et al. (1986) suggest using experimental games, which are simple role plays, in order to quantify the relevance of fairness in economic transactions. They use an ultimatum game (Güth et al., 1982): player A is asked to divide a sum of money between herself and another (anonymous) player B. Player B may either accept or refuse the offer. If B refuses, both A and B leave the experiment without receiving anything. If B accepts, they receive the amounts corresponding to A's share. A basic game theoretic approach predicts that A should send the smallest amount possible, e.g. one cent, and B should accept any positive offer since this is still better than nothing. In a meta-analysis spanning three decades of research, Tisserand (2014) shows that the average offer is about 42% of the amount at stake. Some people resist unfairness, and individuals largely anticipate this behavior.

This kind of experimental work flourished and broadened in scope to create the so-called *social preferences* literature. Researchers have increased the number of games in the experimental economist's toolbox to quantify a variety of "other-regarding" attitudes and preferences. The ultimatum game was simplified to create the famous dictator game designed to measure *altruism* (Kahneman et al., 1986; Forsythe et al., 1994; Engel, 2011). The game is the same as the ultimatum game, except that now player B cannot reject the offer made by A (player A thus becomes a *dictator*). Engel (2011) conducted a meta-study of 131 research articles using dictator game experiments. Figure 1 shows the aggregated results of the studies providing individual level data. It plots the distribution of give rates selected by Player A. On average, the give rate is about 28% of the initial endowment, far above the 0% predicted by economic theory. Moreover, the distribution shows interesting patterns, with a bunching around a 50-50 split (16.7% of the sample). The trust game was also designed by Berg et al. (1995) to measure *interpersonal trust* and *trustworthiness*. The premise is also simple: player A (the trustor), is endowed with a sum of money, say 10 euros. Player A can send any fraction of this endowment to player B (the trustee), who receives the fraction multiplied by 3. If A sends 5 euros, B will therefore receive 15 euros. Finally, B can return any fraction of the 15 euros back to A. The amount that A sends initially is a measure of her trust in B [9], while the amount that B sends back to A is a measure of her trustworthiness. The public good game completes the set of basic games used by psychologists and experimental economists, by measuring *cooperation* (Fehr and Gächter, 2000a).[10]

These experimental games ultimately have much to do with inequality, especially the dictator game. However, they have limited scope since income is distributed between two players only, which

and Schmidt (2006) for a review.

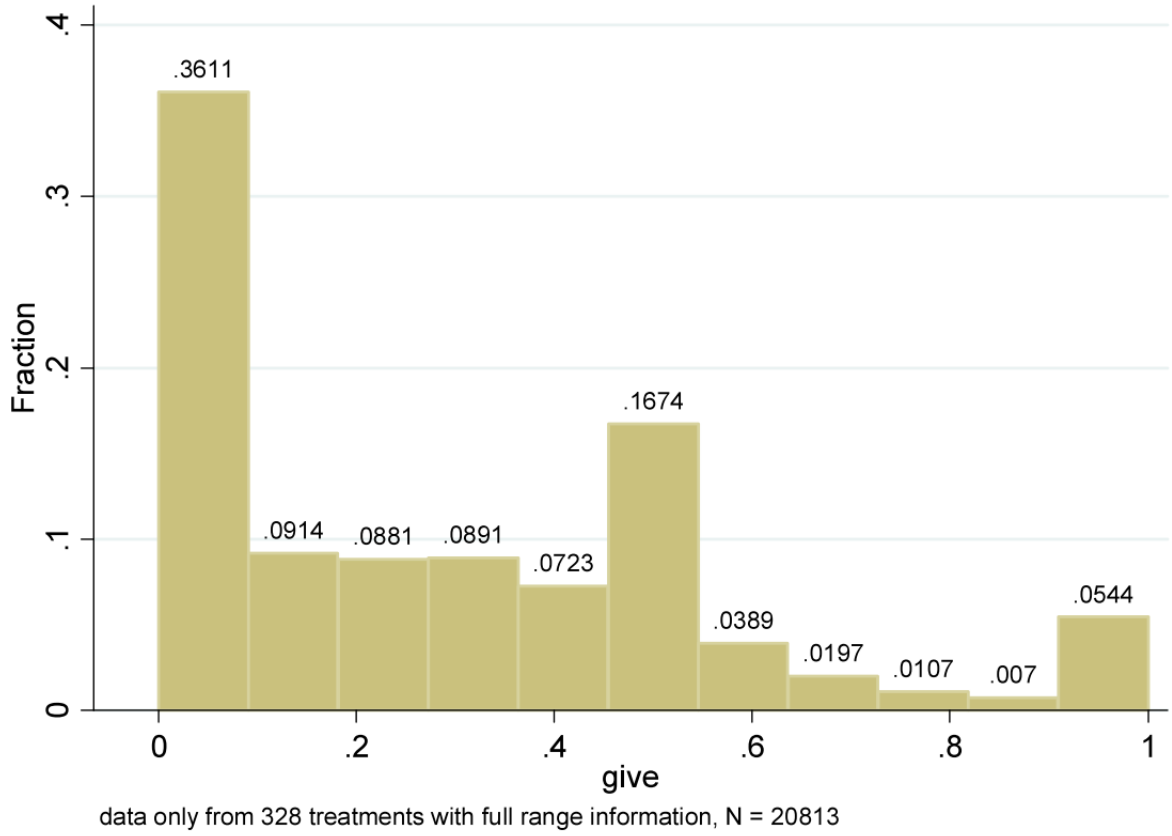
[8] Kahneman et al. (1986) give the example of a situation in which a football team has the choice of selling tickets either through an auction (tickets are sold to the highest bidders), by lottery, or on a first-come first-served basis. Economic theory predicts that auctions are the optimal allocation method. When asking 191 individuals about what they considered to be the fairest and least fair allocation methods, auctions were deemed to be the least fair in 75% of the cases.

[9] Later works have emphasized that the amount transferred by A to B is influenced by other motives such as risk aversion (Eckel and Wilson, 2004), betrayal aversion (Bohnet and Zeckhauser, 2004) and altruism (Cox, 2004)

[10] Levitt and List (2007) also list the gift-exchange game as another way to measure social preferences. This game is more specific because it is more contextualized than the others as it is framed as an employer-employee relationship. Player A is the employer and B the employee. The employer offers the employee a wage (lump-sum transfer) and asks for a desired level of effort from the employee in return. The employee chooses a (costly) effort level, with a greater effort increasing the payoff for the employer. This game also measures trust and trustworthiness in employer-employee relationships (also called *positive reciprocity*).

makes this game relevant for bilateral relationships but much less so for group or even society-wide interactions. A promising branch of the literature has added multiple players to the dictator game in order to focus on *distributive preferences*, which is the type of experimental game we use in Chapter 2 (Engelmann and Strobel, 2007; Krawczyk, 2010; Durante et al., 2014; Bjerk, 2016; Lefgren et al., 2016).

Figure 1 – Distribution of individual give rates



Source: Engel (2011)

Implications for economic theory

Study after study, these games have shown in a striking way that social preferences and fairness are of paramount importance and relevant even to bare-bones interactions between two anonymous individuals.[11] A large proportion of individuals go against their self-interest and are willing to forgo money in order to be nice to people that have treated them nicely (*positive reciprocity*), and punish people that behave badly (*negative reciprocity*) (Fehr and Gächter, 2000b). Kahneman et al. (1986)'s call for the inclusion of such behavioral motives in economic theory has been heard. The most famous contributions in this respect are the models of social preferences by Rabin (1993), Fehr and Schmidt (1999), Bolton and Ockenfels (2000) and Charness and Rabin (2002). Broadly speaking, these models aimed to provide a valid theory to explain the empirical evidence produced by the experimental games.

[11] In practice, these experiments are conducted in laboratory settings with subjects in front of computer screens displaying the instructions. They are sometimes alone in a room, and do not see the other player in the game at all. We can argue that such experiments constitute lower bounds for the importance of fairness attitudes and preferences, given the richness of real human interactions.

They became well known because they managed to be very simple and yet insightful. To give a flavor of these models, I will briefly return to the two-player solution provided by Fehr and Schmidt (1999). They account for the co-existence of both fair-minded agents, i.e. agents choosing non-zero transfers to player B in the dictator game (63% of the sample according to the meta-study shown in Figure 1), and selfish agents (36%). Their solution is to model fairness as *"self-centered inequity aversion"*: people resist inequitable outcomes, where inequity is understood in relative terms.[12]

Fehr and Schmidt (1999) considers two individuals i and j with their respective monetary payoffs, x_i and x_j. The utility function of i is

$$U_i = x_i - \alpha_i \max\{x_j - x_i, 0\} - \beta_i \max\{x_i - x_j, 0\} \tag{1}$$

where $\beta_i \leq \alpha_i$ and $0 \leq \beta_i < 1$. The parameters α_i and β_i describe the extent of disadvantageous inequality aversion ($x_j > x_i$) and advantageous inequality aversion ($x_i > x_j$), respectively. Individuals are typically more upset by inequality when they are behind than when they are ahead, hence the assumption that $\beta_i \leq \alpha_i$. When both β_i and x_i are 0, then the model boils down to a simple payoff-maximizing utility function. From this utility function, they can derive rational behavior that is consistent with the stylized facts obtained from experimental games on social preferences.

This type of model is fairly narrow in scope as the economic interactions considered are minimalist (money transfer between two or more individuals). Nevertheless, these theoretical efforts have inspired, and are inspired by, broader movements in economics to incorporate non-selfish or non-monetary considerations into theories of individual behavior.[13]

The pluralism of fairness ideals

The description of the Fehr and Schmidt (1999) model shows that the concept of fairness remained vague at that time. Modeling it as a relative comparison of *outcomes* is already a debatable choice. In a way, Fehr and Schmidt (1999) confused fairness and equality. Inequitable outcomes are simply *unequal* outcomes in their model: one player receives more money than the other player in the game. Many other subsequent research works have operationalized the fairness concept and introduced a clearer distinction between inequality and inequity. This has been mostly done by complexifying the context of choices in order to incorporate production, productivity, and efficiency, and consequently the concepts of merit, effort, and luck. They measured the extent to which the *origin* of income inequality mattered to people, both in terms of preferences and behavior.[14]. We will see that *merit* is a powerful underlying theme in all this literature. It is useful to take a step back and understand why economists paid so much attention to this particular dimension of fairness.

First, a simple explanation for the dominance of this concept could lie in its elusiveness. It is almost impossible to provide a concrete definition of fairness, or rather, there are simply too many possible definitions. Its simplest definition is probably a negative one, being the opposite of luck (Rosanvallon,

[12] Fehr and Schmidt (1999) builds upon a large body of empirical literature showing that people compare themselves with others and are unhappy when they obtain less favorable outcomes than others Clark and Oswald (1994).

[13] To name but a few, Akerlof and Yellen (1990) show how fairness concerns influence wage-setting on the labor market. Akerlof and Kranton (2000) models identity, *"a person's sense of self"* and its influence on behavior and economic payoffs in various situations such as ethnic and gender discrimination, poverty and educational choices (Akerlof and Kranton, 2002). Bisin and Verdier (2001) show the importance of social norms and cultural transmission between generations in explaining economic behavior.

[14] Justice principles become entirely *context-dependent* Konow (2001)

2011). We do not know where merit starts and where it ends, so experimental economists have tried to investigate all its aspects.

Another explanation could be found in a broader cultural context. According to Piketty (2019), societies develop a coherent ideological framework to rationalize and legitimate inequality. For instance, he analyzes the ideologies of the *Ancien Régime*, of colonial and slave societies and the caste system in India. To some extent, in each of these cases, Piketty describes the fairness ideal of the *dominant* class, the equity principles of the economic, social, and political winners. In non-democratic societies, these fairness principles are not necessarily those that foster the adhesion of the majority. However, with the extension of voting rights to an increasingly large proportion of the population, the dominant class could no longer justify inequality by some natural rule or purely inherited rights. Thomas Piketty argues that it is necessary to bear this particular context in mind in order to understand the success of *merit* as a fairness principle.[15] This is an extremely powerful device as it makes the victims responsible for their lower social and economic statuses and can foster adhesion beyond the dominant class, thus having the power to stabilize inequality even in a democracy (Duru-Bellat, 2006; Dubet, 2014; Duru-Bellat, 2019). According to François Dubet, meritocracy has become a *"necessary fiction"*. As a matter of fact, Emile Boutmy, the founder of Sciences Po (formerly named the *Ecole libre des sciences politiques*), openly justifies the creation of his merit-based school as a strategy for the social reproduction of the dominant class. Favre (1981) quotes Boutmy: *this new education is intended for the classes that have an established position and the time to grow intellectually. These classes have been politically dominant so far; but they are threatened... Forced to succumb to the power of majority, the so-called upper classes cannot maintain their political hegemony any longer unless they invoke the right of the more capable. Behind the crumbling façade of their prerogatives and tradition, the onslaught of democracy should run into a second rampart consisting of bright and useful merits, indubitable superiority, competencies that we would be mad to ignore.*[16] Boutmy's vision can help explain why merit and its related ideological regime – *meritocracy* – have been so successful, especially in Western democracies.

It can be considered that the broad focus on merit as fairness in experimental economics stems from this historical and political heritage. With its investigation of the heterogeneity of fairness ideals, this literature has also tempered the importance of such a definition of fairness by showing that this meritocratic ideal is not shared by everyone. This is also one of the conclusions we reached in Chapters 2 and 3 of this dissertation.

This literature on fairness can be summarized by two heuristic additions to the basic social preference games. First, the introduction of a *production phase* so that the subsequent distribution of monetary payoffs is based on tangible outputs, and second, the distinction between *impartial spectators* and stakeholders.

[15] See Piketty (2019) pp. 825–829

[16] My translation. Original text in *Quelques idées sur la creation d'une Faculté libre d'enseignement superieur*, Lettres de E. Boutmy et E. Vinet, Programme, Paris, Impr. de A. Laîné, 1871,28 p. cited from Favre (1981): *L'enseignement nouveau s'adresse aux classes qui ont une position faite et le loisir de cultiver leur esprit. Ces classes ont eu jusqu'ici la prépondérance politique; mais elles sont menacées... Contraintes de subir le droit du plus nombreux, les classes qui se nomment elles-mêmes les classes élevées ne peuvent conserver leur hégémonie politique qu'en invoquant le droit du plus capable. Il faut que, derriere l'enceinte croulante de leurs prérogatives et de la tradition, le flot de la democratie se heurte à un second rempart fait de merites éclatants et utiles, de supériorités dont le prestige s'impose, de capacités dont on ne puisse pas se priver sans folie.*

Property rights and production phase The idea of a property right or a production phase preceding the distribution phase can clarify the origin of the endowment that Player A receives in a dictator game, for instance. In the case of property right, the subject's ranking within the income distribution – the right to be the person sharing out the endowment in the dictator game, or the right to occupy the top position within an income distribution in a multi-player dictator game (as in Chapter 2 in this dissertation) – is determined by a given rule. Differences in sharing behavior in the distribution phase through variations in property-right-allocation rules enable the measurement of justice principles. Property rights can be allocated on the basis of luck, which could be seen as unfair, or of merit or effort. Hoffman et al. (1994) shows that when the first movers in the ultimatum game earn their endowment by performing better at a general knowledge quiz, they tend to be much less altruistic in their sharing behavior with the second mover.Cherry et al. (2002) describe a similar result for dictator games: dictators are less generous when they earn larger endowments due to better performance in a quiz. Similarly, Oxoby and Spraggon (2008) find that if dictators earn the total payoff, they will behave selfishly, but if receivers earn it, on average, they will allocate a significant amount to the receiver.

Moving beyond property rights and building up a complete production phase leads to an even richer typology of justice principles. In this case, the endowment usually provided by the experimenter without any counterpart is now "produced", unit by unit, by carrying out a real-effort task. This leads to productivity differences across subjects that can be measured and compared to the differences in the final payoffs that subjects obtain.[17] This is the method followed in Chapter 3 of this dissertation. In this vein, Cappelen et al. (2007) test the relative importance of three fairness ideals: (*i*) *strict egalitarianism*: preference for equality of outcomes in all situations; (*ii*) *libertarianism*: it is fair to reward individuals in the proportion of their production; (*iii*) *liberal egalitarianism*: it is fair to reward individuals in proportion to their production only if differences in production originate from factors under the individual's control. They find that that 43.5% of their sample of Norwegian subjects are strict egalitarians, 38.1% are liberal egalitarians (which is arguably the fairness ideal closest to "merit") and 18.4% are libertarians. This study thus shows that merit-based narratives in order to justify inequality are not shared by everyone.

Almås et al. (2020) compare the fairness ideals of Americans and Norwegians, both within and between countries. They study how distributive preferences vary according to the origin of earnings. Inequality could be either based on pure luck or merit (individual productivity), and they also vary the cost of redistribution (efficiency). They find that overall, Americans accept more inequality than Norwegians, and that there is a socioeconomic gradient in the US, whereby highly educated subjects favor inequality more when productivity is the basis.

Some experiments have investigated other fairness views that are based on *agency* arguments. For instance, Cappelen et al. (2013) asks whether inequality acceptance varies according to whether it is based on pure luck or risk-taking. They find that most participants favor the equalization of earnings when inequality is the result of good or bad luck. But when inequality stems from poor choices (people making a risky bet and losing instead of choosing a safe outcome) then inequality is much more frequently accepted. Akbas et al. (2019) introduces a distinction between procedural justice and agency. Procedural justice is violated when the choice set of an individual is restricted so that she cannot fully exercise her agency. Agency is achieved if someone can be held fully responsible for the

[17] This is related to the *Accountability principle* in Konow (2001).

choices leading to earning outcomes. They find that subjects do not wish to reduce inequality only if both equality of opportunity and agency are combined.

Impartial spectators and choices behind the veil of ignorance The second innovation relies on the status of the subjects of the experiment. In most of the experimental games presented above, the subjects are *stakeholders* in the sense that they are directly impacted by the distributional choices they make. For example, in a regular dictator game, when the first mover decides on a 50-50 share with a second mover rather than keeping everything for herself, she is facing a trade-off: her sharing decision involves her and someone else. It is often argued that this design is necessary to ensure the experiment's incentive compatibility (Cappelen et al., 2007) [18], but this also pollutes the elicitation of distributive and fairness preferences, since selfish and other-regarding motives are conflicted.

The *impartial spectator* method addresses this concern by asking experimental subjects to make distributional decisions for *other* people (other subjects in the lab). For instance, in a dictator game with a production phase, two subjects produce an endowment and the dictator is a third person who must decide how to distribute this endowment between both "workers". Since the money is, in reality, distributed according to the spectator's will at the end of the experiment, the experiment remains incentive-compatible to some extent. This method is inspired by Smith (1759) and has been largely developed by the philosopher Rawls (1971, 2001), who suggested that decisions concerning the just allocation of resources among members of society should be made behind *the veil of ignorance*. People should choose distributive principles without knowing which position they will occupy within the society. They should not even be able to guess the probability of achieving any position. This is the only way to attain *unbiased justice* (Konow, 2001). Such mental games have been extensively reproduced in the lab (Konow, 2001; Engelmann and Strobel, 2004; Konow, 2009; Cappelen et al., 2013). In Chapter 2, we use the method of choices behind the veil of ignorance and in Chapter 3, we compare impartial spectators with stakeholders in order to isolate distributive preferences from selfish concerns.

What happens when fairness ideals are violated? There is strong evidence that justice principles matter tremendously and that people reject outcomes that violate their fairness principles. Zizzo and Oswald (2001) shows in a lab experiment that many people are willing to pay in order to burn the money of other subjects, especially those that received a high payoff purely through luck. Outside of the lab, extensive literature shows how income comparisons affect people's well-being and wealth (Card et al., 2012). In Chapter 2, we show that many people believe that inequality is acceptable when it is merit-based, when choosing behind the veil of ignorance. However, the behavior of the losers in the system can be polarized. Some will accept inequality, but other will resist it and engage in money burning. This kind of attitude reveals that the meritocratic narrative is not all-powerful, and even if *in principle* it can foster broad adhesion, once outcomes are realized, significant bitterness can result. The Yellow Vests movement in France exemplifies such an argument. It has often been depicted as being fueled by anger about injustice, especially concerning economic inequality and the lack of redistribution.[19]

[18] If subjects do not have any monetary stakes in the experiment, they may be more subject to social desirability bias and experimenter effects. They may not report truthfully on their preferences and may pay less attention to the experiment.

[19] See (Algan et al., 2019b) and *Gilets jaunes : une enquête pionnière sur la "révolte des revenus modestes", Le Monde, 11 décembre 2018*

Inequality as the result of individuals' choices

Ex post inequality is the result of both an *ex ante* distribution of labor and capital income *and* an organized redistribution phase. Income distribution *before* redistribution is shaped by myriads of individual and collective choices. In my dissertation, Chapter 1 is about educational choices, which lead to different career tracks and wage paths; Chapter 3 is about wage contract choices, which imply different levels of wage inequality within firms; and Chapter 4 is about ethnic discrimination, which can reduce cooperation across groups and be detrimental to the labor market opportunities of ethnic minorities.

Higher education aspirations and choices

Education is a key institution when it comes to inequality, as it can be either a tool for social reproduction or social mobility, depending on how it operates (Haveman and Smeeding, 2006; Piketty, 2019). Economists analyze education as an investment that brings a return, which can be monetary (wage) or of a different nature (psychological gain, prestige, health, etc.). It is also associated with a cost, since years spent studying are costly, both in terms of effort and forgone income (Becker, 1964). Differences in years of education (Card, 1999; Dickson and Harmon, 2011) and also differences in terms of the educational track (field of study) pursued (Zafar, 2013; Arcidiacono, 2004) cause income inequality. In the US, Carnevale et al. (2013) shows that the highest-earning college majors (Petroleum Engineering, Pharmaceutical Science, Compute Science, etc.) earn three to four times more than the lowest-earning majors (Counseling Psychology, Early Childhood Education, etc.), at the median. In Chapter 1, we use data on the cohort of students admitted to Sciences Po to better understand their study choices. Even within Sciences Po, these choices already lead to large earning differences one year after graduation. Students graduating from the Finance and Strategy master's degree earn much more on average than those graduating from the Urban school, for instance.

Given these large starting-wage differences – which stack up over the individuals' careers – the literature has endeavored to understand what motivated those different choices, and asks whether inefficiencies and frictions constrain them. It has unearthed a large variety of motives and frictions, sometimes institution-based or individual-centered. For instance, the under-investment of students from disadvantaged backgrounds has been shown to stem from financial constraints (Solis, 2017), imperfect information about educational paths (Guyon and Huillery, 2019), self-censorship and aspiration failures (Dalton et al., 2016; Genicot and Ray, 2017; Carlana et al., 2017). Similarly, girls tend to make less ambitious choices than boys, despite having similar or even better academic records (Boring and Brown, 2020). By taking stock of all these potential determinants of higher education choices, a strand of the literature has tried to disentangle all the underlying mechanisms. Educational choices result from a mixture of beliefs, expectations, aspirations and learning, social norms, and stereotypes (Filippin and Ichino, 2005; Zafar, 2011; Beffy et al., 2012; Zafar, 2013; Stinebrickner and Stinebrickner, 2014; Wiswall and Zafar, 2015). The first chapter of this dissertation contributes to this literature by focusing more specifically on updating academic tastes and by analyzing the step-by-step learning process leading to master's degree choices.

Ethnic discrimination

In Chapter 4 of this dissertation, we study the patterns of ethnic discrimination in the US and Germany. With regard to discrimination, we are now interested in how the choices and preferences of *others* affect one's own income. The seminal works by Becker (1957) and Arrow (1972) profoundly influenced the study of discrimination by respectively arguing that it either stems from taste-based (pure dislike), or statistics-based considerations (ethnicity is a signal for characteristics relevant to the employer such as skills). These theoretical works have been accompanied by a vast empirical literature aiming to identify the existence of discrimination and quantifying the relative importance of taste-based vs statistics-based motives (Guryan and Charles, 2013). Proving the existence of discrimination – be it based on ethnicity, gender, or age – is a difficult task: it is hard to separate discrimination from simple differences in characteristics. For instance, if a firm needs to choose between two people for a post (a black or a white candidate), choosing the white candidate is not a proof of discrimination per se since it could mean that the white candidate simply has better credentials. The main innovation of the labor economics literature in this respect has been to use ingenious methods, including the so-called *fictitious resumes* experiments, in order to overcome these confounding factors (Bertrand and Mullainathan, 2004; Lahey, 2008; Oreopoulos, 2011; Valfort, 2018). The basic idea is to draft two resumes, which are as equivalent as possible in terms of education, professional experience, hobbies etc., but differ only with regard to the applicant's name. For instance, Bertrand and Mullainathan (2004) uses white-sounding and black-sounding names and Valfort (2018) uses the most frequently given Muslim and Christian names. The next step is to use these resumes to apply to (real) job ads, randomizing the name appearing on top of the resume. The researcher then only needs to wait for the firm to get in touch in order to find out whether or not a candidate has been asked to attend an interview. Bertrand and Mullainathan (2004) finds a 50% difference in callback rates between black and white candidates. Valfort (2018) finds more modest differences between Christian and Muslim candidates, but the gap widens when the resumes of Muslims candidates indicate a higher degree of religiosity[20]

Experiments on ethnic discrimination can also be conducted in the laboratory, as we see in Chapter 4. This has the advantage of analyzing the mechanisms in greater detail and also of studying them in more general situations, not necessarily on the labor market. This literature stems from experimental social psychology that has been studying the foundations of group identity and discrimination for many years(Allport et al., 1954; Tajfel et al., 1971; Brewer, 1999). One of the main findings is that the tendency to discriminate in favor of those identified as belonging to the same group (the so-called "ingroup") at the expense of others identified as belonging to another group (the so-called "outgroup") is endemic (Balliet et al., 2014; Lane, 2016). Such findings are obtained by means of the social preference games we described above. Consider a trust game, for instance, as described in Chapter 4: if we disclose the ethnic group of the second mover in this game, we can track how this information affects the transfers the first mover chooses to "invest". If transfers are lower when the second mover is not from the same ethnic group as that of the first mover, we can conclude that the first mover presents an ethnic ingroup bias. Such a bias has been found in experiments conducted with groups differing on the basis of for their nationality (Romano et al., 2017; Dorrough and Glöckner, 2016), ethnicity (Fershtman and Gneezy, 2001; Fershtman et al., 2005), religious affiliation (Chuah

[20]Religiosity is also manipulated in an exogenous manner by randomizing statements that the applicant enrolled in *Scouting* activities: "Scouts and Guides of France" for Christians and "Muslims Scouts of France" for Muslims.

et al., 2013, 2014), political groups (Rand et al., 2009; Weisel and Böhm, 2015), and also when groups differed according to purely arbitrary characteristics induced in the laboratory (Tajfel et al., 1971; Charness et al., 2007; Güth et al., 2009; Hargreaves and Zizzo, 2009).

Inequality as the result of a trade-off between efficiency and distributive preferences

To add the final touches to this portrait of the literature I rely on, I will use the insights from the Personnel Economics literature. We have seen that people hold views about inequality, whether or not it is fair, and under which conditions it is perceived as just. Moreover, people make choices on a daily basis that are based on their preferences, beliefs, expectations and also on their biases, which in turn affect their own income and the income of others, and therefore concern inequality. In this section, I would like to insist on the fact that inequality may be chosen, even if people oppose it, because it is associated (or they *believe* it is associated) with other more positive outcomes. In other words, inequality can be the result of a trade-off between two desirable things. People do not only have distributive preferences, they also have efficiency concerns (Kritikos and Bolle, 2001; Engelmann and Strobel, 2004), and both inequality and efficiency are related to each other. The philosopher Rawls (2001) argues that this is actually one justice principle under which inequality becomes acceptable. Inequality must be productive (efficient) in the sense that it improves the lot of the worse-off (*maximin* condition).[21]

How could efficiency and inequality be related? The rules governing the payoff distribution will actually have an impact on the production process. This idea is intuitive: consider a team of blue-collar workers producing certain goods, for example. The employer may choose to implement performance pay, so that workers are paid per unit produced instead of at a fixed wage. By providing these extrinsic (monetary) incentives, workers are expected to exert more effort and thus produce more in order to increase their wages. The total production (and therefore efficiency) then becomes higher under performance pay, but this also implies more unequal wages for workers with different ability levels, whereas equality would have been achieved with fixed equal wage. Shearer (2004) shows that this mechanism works in a tree-planting field experiment. Tree planters were randomly allocated to a fixed-wage treatment or a piece-rate wage treatment. Productivity was 20% higher under the piece-rate condition.[22]

This situation raises two questions: first, does this simple mechanism hold true in other settings and, more broadly-speaking, at a society-wide level? Second, do people *believe* that inequality can lead to a higher social surplus and in this way *choose* to favor efficiency over equality? The economic literature does not provide definitive answers to either of these questions.

First, in more complex interactions, monetary incentives (and therefore inequality) can backfire, especially within teams. Literature on *social incentives* (Ashraf and Bandiera, 2018) has shown that workers are sensitive to their co-workers' wages and dislike pay differences, which can undermine effort (Clark et al., 2010; Bartling and von Siemens, 2010; Gross et al., 2015)[23], group cooperation (Anderson et al., 2008; Breza et al., 2017) and also have adverse effects on the well-being of workers

[21] We must emphasize that this is not a sufficient condition, nor the most important one. Rawls (2001) explains that the foremost condition is the *Liberty principle*, whereby people have democratic rights and the freedom of speech and conscience. The second justice principle is that economic and social inequality must arise from equal opportunities, and then finally satisfy this maximin condition.

[22] This has been shown in other similar contexts (Lazear, 2000; Bloom and Reenen, 2010; DellaVigna and Pope, 2018).

[23] The literature is not unanimous: Charness and Kuhn (2007) and Bartling and von Siemens (2011) find no evidence that workers' efforts are affected by their co-workers' wages.

(Card et al., 2012). Aggregating all these situations leads to ambiguous predictions regarding a society-wide relationship between inequality and efficiency. Theoretical works have modeled how inequality could affect economic growth, sometimes concluding in a positive or negative relationship, depending on the assumptions and modeling choices.[24] Proponents of the positive relationship argue that inequality stimulates innovation, which fuels economic growth (Foellmi and Zweimüller, 2006; Jones and Kim, 2018).[25] Opponents argue that inequality can lead to unproductive investments by the rich, lower levels of human capital and political unrest (Halter et al., 2014; Aghion et al., 1999, for a review of these arguments). The empirical evidence tends to lean toward a negative correlation between inequality and growth, but the result is dependent on the type of data, the econometric methods, inequality measurement choices, the countries studied and the time frame (Aghion et al., 1999; Ehrhart, 2009).

The second question – whether people *believe* that inequality increases efficiency – has rarely been studied, and this is the gap in the literature that we aim to fill in the third chapter of this dissertation. We show that even if managers want to grant the same wages to their workers, their normative distributive preferences for low inequality are partially crowded out because they anticipate that workers would be less productive with lower monetary incentives.

This belief is key because it influences both the pre-tax income distribution and redistributive policies. In the absence of capped wages, employers are responsible for the wage spread within their firms, and largely influence the pre-tax income distribution with their own views about what is an efficient *and* a fair wage distribution. Second, the belief about the existence of an equality-efficiency trade-off influences voting and hence the political choices of a society regarding redistribution. We can only speculate about the evolution of such a belief in the population in the future. Inequality economists' careful and precise counterarguments arguably shook people's views about the existence of such a trade-off. In France, President Emmanuel Macron decided to abolish the wealth tax in 2018 and he has extensively used rhetoric revolving around the benefits brought by the *"premiers de cordée"*. Through this metaphorical image relating economic leaders to climbers that lead the way, he embraces the belief that inequality creates a larger economic surplus that benefits everyone. This discourse has generated significant debate concerning the underlying "trickle down" theories, which have very often been denounced and even mocked[26]. The appropriation and critical appraisal of complex economic theories by the masses, and the political instability personified by the Yellow Vests movement suggest that beliefs are indeed changing.

Outline of this doctoral dissertation

Over the coming pages, we will now outline the four research works constituting this thesis.

[24] Seminal work on the question has tended to focus on the opposite relationship: the impact of economic development on inequality. Kuznets (1955) argues that inequality is a by-product of early development stages, since economic growth tends to be initially captured by the wealthy, thanks to their higher savings (and therefore investment) capacity. Over the course of economic development, the concentration of savings among a limited number of people is counteracted by political decisions and the rise of the welfare state, leading to the famous inverted U-shaped Kuznet curve. Later on, the theoretical literature starts drifting toward the opposite causal relationship, from inequality to growth.

[25] See Aghion et al. (2019) for a paper showing the empirical relevance of this relationship in the US. A counterargument often put forward is the fact that economic growth does not benefit everyone. Alvaredo et al. (2018) shows that over the 1980-2016 period, the top 1% captured 27% of total growth.

[26] Many media have published articles explaining this theory and about its critics, and argue about whether or not Emmanuel Macron's policies share similar ideological roots. For example, see: Le Monde, France Culture, Europe 1, L'Express, Libération, Le Figaro, La Croix.

Chapter 1: Updating of academic tastes and ability signals

In this Chapter co-authored with Ghazala Azmat, Anne Boring and Roberto Galbiati, we focus on the *why* question outlined earlier, in the context of higher education choices. We document the learning process that leads students to choose a master's degree. When students enter college, they receive ability signals and discover new fields of study, which cause them to update their beliefs about their academic preferences. This taste-updating process is important because it eventually leads to choices that have significant consequences on the students' labor market outcomes. For instance, starting salaries upon graduation differ substantially across different master's degrees and majors (Black et al., 2003; Carnevale et al., 2013).

Moreover, knowing how and why students change their academic tastes can inform the debate on the flexibility of higher educational systems. There are large institutional differences in the way students choose their majors. In countries like the US, students experiment with different fields before specializing, while in other countries, like the UK and France, majors are chosen even before entering university. A low-flexibility system has the drawback of hindering changes in academic tracks, but has the advantage of accelerating skills accumulation (Malamud, 2011).

In this chapter, we explore this process by documenting how and why students' beliefs about their academic preferences are updated throughout their undergraduate studies. To conduct our analysis, we use a unique dataset from a French university specializing in social sciences – Sciences Po. This dataset has several distinct advantages. First, it is longitudinal, as we can follow students over the course of their studies, from their last year of high school through to their graduate studies. We measure academic tastes at two points in time using the content of letters of motivation that all students have to write. The first one is submitted as part of their application package to Sciences Po, while they are still in high school, and the second one is written two years later when they justify their choices of universities for the mandatory third-year study abroad program. Second, this university system is flexible in the sense that students specialize late in their studies, although they have no latitude in their choices of courses in first year. The curriculum is identical for all first-year students; they must try all the major social science fields (Economics, History, Law, Political Science and Sociology) before making any important academic choices. This feature is rare in other universities where students often self-select into their courses and do not try out their affinities and abilities in all majors. Self-selection is therefore not a concern in our context. Finally, there is no supply-side constraint on master's degree choices, which alleviates strategic concerns that would arise if students had to anticipate other students' choices would choose if they wanted to shy away from competition, or on the contrary, challenge themselves.

We find that academic tastes elicited both in high school and in second year correlate with revealed choices (master's degree choices), this relationship being stronger with the latter than the former. This indicates that students are changing their mind after exposure to the curriculum of the university. We then investigate the reasons why this updating process takes place. We focus on three types of information available to students. First, we use data on students' priors. They do not start higher education with the same degree of information about their academic preferences. Some students may hold very strong beliefs and already be passionate about a subject, while others are more uncertain. We proxy the degree of certainty of initial beliefs using a keyword-based measure of the strength of the dominant academic taste of the student in high school. Second, we study the effect of several ability signals students receive: (i) they learn about their performance in the field in which they

initially wanted to major; (ii) they learn about their ability in other fields; (iii) they can compare their performance with respect to their peers. We also study whether the precision of the information about peers' performance matters for taste updating.

We find that about 75% of students update their academic tastes over the first two years of their studies. Ability signals play an important role in this process, they are drivers of perseverance when they indicate a good fit with initial tastes, and drivers of change when they are a signal that the student has higher ability in a different field. Nonetheless, for students who have particularly strong beliefs that they will eventually choose a certain field of study, we find a stronger persistence in tastes, even when students receive negative information about their academic ability. Further analyses suggest that relative performance through comparisons with peers does not play a significant role. A higher degree of precision regarding peers' performance does not change this result. Nor do we find differences in the taste-updating process according to the students' gender or socioeconomic background.

Chapter 2: Preferences over income distribution, evidence from a choice experiment

In this Chapter co-authored with Max Lobeck, Claudia Senik and Thierry Verdier, we focus on the second main research question of my thesis by analyzing how individuals answer the following question: *should different people be paid unequally?*. We assess how people's distributive preferences are revealed according to the context of the choice, focusing on three aspects: (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 a 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 (based either on luck or a real effort task). Within subjects, we vary whether or not one distribution is Pareto-dominant in relation to the other. We also ask subjects to choose successively behind a veil of ignorance, hence without knowing their future rank and payoff, and then with information about their position within their group. The series of binary choices that subjects must 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 to Merit and Luck treatments. In the Merit treatment, people's position within their group of five people 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 randomly determined.

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 disappears once positions within the income distributions are fixed, i.e. when subjects know their own ranking before they choose. In this 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.

Chapter 3: Principals' distributive preferences and the incentivization of agents

In this Chapter co-authored with Max Lobeck, we tackle the issue of people's beliefs regarding the equality-efficiency trade-off and their fairness ideals and how they translate into wage compensation choices. We focus on employers and managers, which are the relevant entities in terms of pre-tax wage distribution.

We provide evidence that there is 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. 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 believe that performance pay motivates workers or whether it is likely to create tensions, etc. This correlation is also robust for a wide array of manager- and firm-specific controls.

Establishing causality in such a context is complicated and to work around this issue, we run a principal-agent lab experiment, randomizing subjects into manager (principal) or worker (agent) positions. Each principal is matched with two workers that differ according to their 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). Spectators can implement their preferred income distribution at no cost, which gives us a measure of the distribution of the income principals believe to be 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.

Moreover, our framework allows us to precisely pin down the relative importance of various fairness ideals (egalitarian, efficiency-minded, and equal-procedure) among principals. Piece-rate wage contracts are an innovation in relation to the existing literature because the comparison of the piece rates chosen for each worker, depending on their ability level, will lead to a direct classification into three distributive preference types. Choosing to reward the high-ability worker with a higher piece rate is

evidence of efficiency concerns 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 paying them proportionally to the output they have produced. This leads to procedural fairness since both workers are treated equally with the same piece rate. Finally, giving a higher piece rate to the low-ability worker shows an egalitarian concern, since differences in productivity will be compensated for. We calibrate these egalitarian contracts in such a way that if both workers exert the same level of effort, they will be paid the same final total wage. This boils down to a common situation in real firms where both workers are paid the same final wage, despite different production levels.

The analysis crucially depends on (i) whether or not agents best 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 the piece rate. This provides control over the efficiency-equality trade-off that principals believe they face before workers start working.

We find that, despite the firm-like setting and the moral hazard situation, principals do have egalitarian concerns. They are, on average, willing to trade off higher output for a decrease in within-firm inequality. This willingness is significantly lower if principals are Stakeholders (extensive margin incentives) and it is also the case within treatment when there is a large trade-off between efficiency 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 equality in procedure as such is not seen as a particularly attractive contract characteristic and principals are more interested in distributive outcomes.

We run a heterogeneity analysis to measure principals' profiles using a finite mixture model. We quantify the importance principals attach to the payoff of the high- and the 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. We assign 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 weight to the low-ability agent's income, and (3) an intermediate group that attaches significant importance to the low ability agent's income if the difference in piece rate becomes too great.

The structural estimates allow us to make counterfactual statements on how the implications of these preferences on firms' performance change once we assume that agents hold horizontal social preferences. Counterfactual simulations that alter *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 do better for moderate levels of agents' inequality aversion.

Chapter 4: Ethnic bias, economic success, and trust

The last chapter is co-authored with a team of *Trustlab* researchers: Yann Algan, Gianluca Grimalda, Fabrice Murtin, Louis Putterman, Ulrich Schmidt, and Vincent Siegerink. We return to the question of the origins of inequality by focusing on ethnic discrimination in an experimental setting. We focus on minimal economic interactions between individuals belonging to different ethnic groups in

order to better understand the extent and the patterns of ethnic-based discrimination and to test whether discrimination can be reduced. We focus on two large Western countries – the United States and Germany. While previous research typically focused on cross-national discrimination, or within-country discrimination between two ethnic groups, our use of large samples makes it possible to study discriminatory patterns between the ethnic majority, two specified minority groups, and a residual group, in both countries. In this way we can study whether discrimination is selective or treats other outgroups similarly. We quantify ingroup bias for both the ethnic majority and the ethnic minorities within both countries. We determine whether ingroup favoritism stems from accurate expectations of low trustworthiness or whether it is purely taste-based. We also test potential treatments in a controlled manner to reduce such bias. Discrimination often hinges upon stereotypical beliefs that ethnic minorities do not share the same work ethic as the ethnic majority. People from ethnic minorities are often depicted as being lazy and taking advantage of welfare benefits (Gilens, 2009; Alesina et al., 2018). We test whether releasing information that people from ethnic minorities are economically successful alleviates discriminatory attitudes from the ethnic majority.

To this end, we conduct a module on ethnic discrimination on the *Trustlab* platform, a large-scale multi-country incentivized online experiment designed to study social preferences, generalized trust and trust in institutions using experimental games (Murtin et al., 2018; Aassve et al., 2018a,b). The module was implemented in the US and in Germany on about 1,000 subjects, representative of the national population of each country. The module consists of several trust games (TGs) involving pairs of players. Both receive an endowment of 10 dollars/euros. The first mover can transfer any fraction of this endowment to a second mover. The transferred amount is multiplied by 3 and the second mover can then return any amount out of this multiplied transfer and her own endowment to the first mover. Our key experimental manipulation is to disclose the second mover's ethnic group to the first mover. In the US, first movers from any ethnic groups are matched in random order with a non-Hispanic White (henceforth "White" for the sake of brevity), an African-American, and a Hispanic second mover. In Germany, first movers are matched in a similar fashion with rooted Germans, a subject of Turkish descent and a subject with Eastern European origins. We measure the prevalence of ingroup favoritism – also referred to as parochial attitudes (Romano et al., 2017) – by comparing the first mover transfers across the different ethnic groups. In our experiment, discrimination coincides with ingroup favouritism. In other words, this is the propensity to transfer larger sums to people from one's ingroup than to people from one's outgroup. Since we record the first mover's ethnicity, we are able to study how favoritism varies according to the ethnic types, thereby making a distinction between the discriminatory behavior of the ethnic majority compared to ethnic minorities. We also study bias selectivity, i.e. whether first mover transfers depend on the type of outgroup second movers. Are subjects more biased against one outgroup compared to another? Is there an ethnic group that is discriminated against or favored by all groups, or are bias patterns completely ethnicity-specific?

The second part of the experiment tests whether information on second movers' incomes can alleviate ethnic ingroup bias. We run another round of TGs where first movers are now matched with rich second movers, whose incomes belong to the top 20% of the country's income distribution. We still vary the second mover's ethnicity. Ethnic-majority participants are thus confronted with rich ethnic minorities, which contradicts the usual populist narratives picturing immigrants or minorities as idle welfare recipients. We also analyze how ethnic minorities react to being matched with rich people from their own or other ethnic minorities.

Overall, we find that members of all ethnic groups have a significant ingroup bias, except for participants of Eastern European descent in Germany. This bias is particularly substantial for African Americans in the US, and rooted Germans and participants of Turkish descent in Germany. We further show that ethnic discrimination is selective in Germany. Rooted Germans discriminate twice as much against Turkish descent participants as against those of Eastern European descent. On the contrary, first movers of Eastern European and Turkish descent discriminate against each other, but are more neutral in their trust relative to rooted Germans. In the US, the ethnic groups have a more homogeneous non-selective ingroup bias.

We can break down ethnic discrimination into a taste-based and a statistical-based component using first movers' beliefs regarding second movers' transfer. Controlling for expected trustworthiness (expected transfer from second mover to first mover), we infer that 80% of the ingroup bias is driven by taste-based discrimination and 20% by statistical discrimination in both countries. We further show that low-trustworthiness stereotypes are mostly inaccurate, except for those concerning Turkish descent second-movers, who transfer significantly less money to first-movers than other groups.

Although participants of all ethnicities reduce transfers when the receiver is known to be rich, matching participants with rich second movers attenuates ethnic discrimination. Ethnic ingroup favoritism almost completely disappears except for African Americans and rooted German first-movers who still favor their own ingroup, even if the second mover is rich, but to a much lower extent than when income information is not released. Moreover, we uncover the existence of a rich ethnic minority premium in Germany. Rooted Germans engage in less discrimination against rich Turkish second movers than against poor ones, suggesting that narratives of successful ethnic minorities could help to change stereotypes. However, we also show that this treatment can backfire and generate distrust within minority groups. In the US, we also observe a rich but smaller ethnic minority premium. White Americans had a smaller ingroup bias in the first place, so there was probably less room for intervention. In the US, the treatment does not generate any backlash within minorities. We also find that first movers belonging in the top 20% of income distribution display ingroup loyalty across income lines, as they transfer more to fellow top 20% income earners than to first-movers from the bottom 80% of the income distribution. However, this income ingroup bias is significant only for rooted Germans.

Bibliography

Aassve, Arnstein, Pierluigi Conzo, and Francesco Mattioli. Was Banfield right ? New insights from a nationwide laboratory experiment. 2018a.

Aassve, Arnstein, Gianmarco Daniele, and Marco Le Moglie. Never Forget the First Time : The Persistent Effects of Corruption and the Rise of Populism in Italy. 2018b.

Aghion, Philippe, Eve Caroli, and Cecilia García-peñalosa. Inequality and Economic Growth : The Perspective of the New Growth Theories. *Journal of Economic Literature*, XXXVII(December): 1615–1660, 1999.

Aghion, Philippe, Ufuk Akcigit, Antonin Bergeaud, Richard W. Blundell, and David Hémous. Innovation and Top Income Inequality. *The Review of Economic Studies*, 86(1):1–45, 2019.

Akbas, Merve, Dan Ariely, and Sevgi Yuksel. When is inequality fair? An experiment on the effect of procedural justice and agency. *Journal of Economic Behavior and Organization*, 161:114–127, 2019.

Akerlof, George A and Rachel E Kranton. Economics and Identity. *Quarterly Journal of Economics*, CXV(3):715–753, 2000.

Akerlof, George A and Rachel E Kranton. Identity and Schooling: Some Lessons for the Economics of Education. *Journal of Economic Literature*, 40(4):1167–1201, 2002.

Akerlof, George A and Janet L Yellen. The Fair Wage-Effort Hypothesis and Unemployment. *The Quarterly Journal of Economics*, 105(2):255–283, 1990.

Alesina, Alberto, Armando Miano, and Stefanie Stantcheva. Immigration and Redistribution. *NBER Working Paper 24733*, 2018.

Algan, Yann, Sergei Guriev, Elias Papaioannoun, and Evgenia Passari. The European Trust Crisis and the Rise of Populism. *Brookings Papers on Economic Activity*, (Fall):309–382, 2017.

Algan, Yann, Elizabeth Beasley, Daniel Cohen, and Martial Foucault. *Les Origines du populisme*. Le Seuil, 2019a.

Algan, Yann, Elizabeth Beasley, Daniel Cohen, Martial Foucault, and Madeleine Péron. Qui sont les Gilets jaunes et leurs soutiens ? *Observatoire du Bien-être du CEPREMAP et CEVIPOF*, 3, 2019b.

Allport, Gordon Willard, Kenneth Clark, and Thomas Pettigrew. *The nature of prejudice*. Addison-wesley Reading, MA, 1954.

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.

Alvaredo, Facundo, Lucas Chancel, Thomas Piketty, Emmanuel Saez, and Gabriel Zucman. World Inequality Report 2018. Technical report, 2018.

Anderson, Lisa R., Jennifer M. Mellor, and Jeffrey Milyo. Inequality and public good provision: An experimental analysis. *Journal of Socio-Economics*, 37(3):1010–1028, 2008.

Arcidiacono, Peter. Ability sorting and the returns to college major. *Journal of Econometrics*, 121: 343–375, 2004.

Arrow, Kenneth J. Some Mathematical Models of Race Discrimination in the Labor Market. In *Racial Discrimination in Economic Life*, pages 187—-294. 1972.

Ashraf, Nava and Oriana Bandiera. Social Incentives in Organizations. *Annual Review of Economics*, 10:439–463, 2018.

Atkinson, A. B. After Piketty? *British Journal of Sociology*, 65(4):619–638, 2014.

Autor, David, David Dorn, Gordon Hanson, and Kaveh Majlesi. Importing Political Polarization? The Electoral Consequences of Rising Trade Exposure. *NBER Working Paper*, 2016.

Autor, David H, David Dorn, Gordon H Hanson, and Jae Song. Trade Adjustment: Worker-Level Evidence. pages 1799–1860, 2014.

Balliet, Daniel, Junhui Wu, and Carsten K.W. De Dreu. Ingroup favoritism in cooperation: a meta-analysis. *Psychological bulletin*, 140(6):1556–1581, 2014.

Bartling, Björn and Ferdinand A. Siemensvon . 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. Siemensvon . Wage inequality and team production: An experimental analysis. *Journal of Economic Psychology*, 32(1):1–16, 2011.

Becker, Gary. *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education (Second Edition)*. University of Chicago Press, 1964.

Becker, Gary S. *The economics of discrimination*. University of Chicago press, 1957.

Becker, Gary S. A Theory of Social Interactions. *Journal of Political Economy*, 82(6):1063–1093, 1974.

Beffy, Magali, Denis Fougere, and Arnaud Miguel. Choosing the Field of Study in Postsecondary Education: Do Expected Earnings Matter? *The Review Of Economics And Statistics*, 94(1):334–347, 2012.

Berg, Joyce, John Dickhaut, and Kevin McCabe. Trust, Reciprocity, and Social History, 1995. ISSN 08998256.

Bertrand, Marianne and Sendhil Mullainathan. Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination. *American Economic Review*, 94 (4):991–1013, 2004.

Bisin, Alberto and Thierry Verdier. The Economics of Cultural Transmission and the Dynamics of Preferences. *Journal of Economic Theory*, 97:298–319, 2001.

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.

Black, Dan A., Seth Sanders, and Taylor Lowell. The economic reward for studying economics. *Economic Inquiry*, 41(3):365–377, 2003.

Bloom, Nicholas and John Van Reenen. Human Resource Management and Productivity. In *Handbook of labor economics*, number 4, pages 1697–1767. 2010.

Bohnet, Iris and Richard Zeckhauser. Trust, risk and betrayal. *Journal of Economic Behavior and Organization*, 55:467–484, 2004.

Bolton, Gary E and Axel Ockenfels. A Theory of Equity, Reciprocity, and Competition. *American Economic Review*, 90(1):166–193, 2000.

Bonikowski, Bart and Noam Gidron. Multiple Traditions in Populism Research : Toward a Theoretical Synthesis. *APSA Comparative Politics Newsletter*, 26(12):7–14, 2016.

Boring, Anne and Jen Brown. Gender, Competition and Choices in Higher Education. 2020.

Brewer, Marilynn B. The Psychology of Prejudice : Ingroup Love or Outgroup Hate ? *Journal of Social*, 55(3):429–444, 1999.

Breza, Emily, Supreet Kaur, and Yogita Shamdasani. The Morale Effects of Pay Inequality. *The Quarterly Journal of Economics*, (April):611–663, 2017.

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

Cappelen, By Alexander W, James Konow, Erik Ø Sørensen, and Bertil Tungodden. Just Luck : An Experimental Study of. *American Economic Review*, 103(4):1398–1413, 2013.

Card, By David, Alexandre Mas, Enrico Moretti, and Emmanuel Saez. Inequality at Work : The Effect of Peer Salaries on Job Satisfaction. *American Economic Review*, 102(6):2981–3003, 2012.

Card, David. *The causal effect of education on earnings*, volume 3. 1999.

Carlana, Michela, Eliana La Ferrara, and Paolo Pinotti. Goals and Gaps: Educational Careers of Immigrant Children *. 2017.

Carnevale, Anthony P, Jeff Strohl, and Michell Melton. What's it worth? The economic value of college majors. Technical report, 2013.

Causa, Orsetta and Asa Johansson. Intergenerational social mobility in OECD countries. *OECD Journal: Economic Studies*, 2010:33–76, 2010.

Charness, By Gary, Luca Rigotti, and Aldo Rustichini. Individual Behavior and Group Membership. *American Economic Review*, 97(4):1340–1352, 2007.

Charness, G and M Rabin. Understanding Social Preferences with Simple Tests. *Quarterly Journal of Economics*, 117(August):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.

Cherry, By Todd L, Peter Frykblom, and Jason F Shogren. Hardnose the Dictator. *The American Economic Review*, 92(4):1218–1221, 2002.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez. Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States. *The Quarterly Journal of Economics*, 129(4):1553–1623, 2014.

Chetty, Raj, David Grusky, Maximilian Hell, Nathaniel Hendren, Robert Manduca, and Jimmy Narang. The fading American dream: Trends in absolute income mobility since 1940. *Science*, 406(April):398–406, 2017.

Chuah, Swee Hoon, Reema Fahoum, and Robert Hoffmann. Fractionalization and trust in India: A field-experiment. *Economics Letters*, 119(2):191–194, 2013.

Chuah, Swee-hoon, Robert Hoffmann, Bala Ramasamy, and Jonathan H W Tan. Religion , ethnicity and cooperation : An experimental study. *Journal of Economic Psychology*, 45:33–43, 2014.

Citrin, Jack and Laura Stoker. Political Trust in a Cynical Age. *Annual Review of Political Science*, 21(1):49–70, 2018.

Clark, Andrew E and Andrew J Oswald. Unhappiness and Unemployment. *The Economic Journal*, 104(424):648–659, 1994.

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.

Corak, Miles. Income Inequality, Equality of Opportunity, and Intergenerational Mobility. *Journal of Economic Perspectives*, 27(3):79–102, 2013.

Cox, James C. How to identify trust and reciprocity. *Games and Economic Behavior*, 46(2):260–281, 2004.

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.

Dalton, Patricio S., Sayantan Ghosal, and Anandi Mani. Poverty and Aspirations Failure. *Economic Journal*, 126(590):165–188, 2016.

DellaVigna, Stefano and Devin Pope. What Motivates Effort? Evidence and Expert Forecasts. *Review of Economic Studies*, 85(2):1029–1069, 2018.

Dickson, Matt and Colm Harmon. Economics of Education Review Economic returns to education : What We Know, What We Don't Know , and Where We Are Going – Some brief pointers. *Economics of Education Review*, 30(6):1118–1122, 2011.

Dorrough, Angela Rachael and Andreas Glöckner. Multinational investigation of cross-societal cooperation. *Proceedings of the National Academy of Sciences*, 113(39):10836–10841, 2016.

Dubet, François. *La préférence pour l'inégalité*. Le Seuil, la républi edition, 2014.

Duflo, Esther and Abhijit Banerjee. *Poor Economics: A Radical Rethinking of the Way to Fight Global*. PublicAffairs, 2011.

Durante, Ruben, Louis Putterman, and J. van der Weele Joël. Preferences for redistribution and perception of fairness: An experimental study. *Journal of the European Economic Association*, 12 (4):1059–1086, 2014.

Duru-Bellat, Marie. *L'inflation scolaire: les désillusions de la méritocratie*. Le Seuil, 2006.

Duru-Bellat, Marie. *Le mérite contre la justice*. Presses de Sciences Po, 2019.

Eckel, Catherine C. and Rick K. Wilson. Is trust a risky decision? *Journal of Economic Behavior and Organization*, 55:447–465, 2004.

Ehrhart, Christophe. The effects of inequality on growth : a survey of the theoretical and empirical literature. 2009.

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.

Favre, Pierre. Les sciences d'Etat entre déterminisme et libéralisme: Emile Boutmy (1835-1906) et la création de l'Ecole libre des sciences politiques. *Revue française de sociologie*, 22(3):429–465, 1981.

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

Fehr, Ernst and Simon Gächter. Fairness and retaliation: The economics of reciprocity. *Journal of Economic Perspectives*, 14(3):159–181, 2000b.

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 and Klaus M. Schmidt. Chapter 8 The Economics of Fairness, Reciprocity and Altruism - Experimental Evidence and New Theories. *Handbook of the Economics of Giving, Altruism and Reciprocity*, 1(06):615–691, 2006.

Fershtman, Chaim and Uri Gneezy. Discrimination in a segmented society: An experimental approach. *Quarterly Journal of Economics*, 116(1):351–377, 2001.

Fershtman, Chaim, Uri Gneezy, and Frank Verboven. Discrimination and Nepotism: The Efficiency of the Anonymity Rule. *The Journal of Legal Studies*, 34(2):371–396, 2005.

Filippin, Antonio and Andrea Ichino. Gender wage gap in expectations and realizations. *Labour Economics*, 12:125–145, 2005.

Foellmi, Reto and Josef Zweimüller. Income distribution and demand-induced innovations. *The Review of Economic Studies*, 73(4):941–960, 2006.

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.

Genicot, Garance and Debraj Ray. Aspirations and Inequality. *Econometrica*, 85(2):489–519, 2017.

Gilens, Martin. *Why Americans hate welfare: Race, media, and the politics of antipoverty policy*. University of Chicago Press, 2009.

Gimpelson, Vladimir and Daniel Treisman. Misperceiving inequality. *Economics and Politics*, 30(1): 27–54, 2018.

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.

Guriev, Sergei. Economic Drivers of Populism. *AEA Papers and Proceedings*, pages 200–203, 2018.

Guryan, Jonathan and Kerwin Kofi Charles. Taste-based or Statistical Discrimination: the Economics of Discrimination returns to its roots. *The Economic Journal*, 123(2006):417–432, 2013.

Güth, Werner, Rolf Schmittberger, and Bernd Schwarze. An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior and Organization*, 3(4):367–388, 1982.

Güth, Werner, Matteo Ploner, and Tobias Regner. Determinants of in-group bias : Is group affiliation mediated by guilt-aversion ? *Journal of Economic Psychology*, 30(5):814–827, 2009.

Guyon, Nina and Elise Huillery. Biased aspirations and social inequality at school: evidence from French teenagers. 2019.

Halter, Daniel, Manuel Oechslin, and Josef Zweimüller. Inequality and growth : the neglected time dimension. *Journal of Economic Growth*, (December 2013):81–104, 2014.

Hargreaves, Shaun P and Daniel John Zizzo. The Value of Groups. *The American Economic Review*, 99(1):295–323, 2009.

Hauser, Oliver P and Michael I Norton. (Mis)perceptions of inequality. *Current Opinion in Psychology*, 18:21–25, 2017.

Haveman, Robert and Timothy Smeeding. The Role of Higher Education in Social Mobility. *The Future of Children*, 16(2):125–150, 2006.

Hoffman, Elizabeth, Kevin McCabe, Keith Shachat, and Vernon Smith. Preferences, Property Rights, and Anonymity in Bargaining Games. *Games and Economic Behavior*, 7:346–380, 1994.

Inglehart, Ronald F and Pippa Norris. Trump, Brexit, and the Rise of Populism: Economic Have-Nots and Cultural Backlash. 2016.

Jones, Charles I and Jihee Kim. A Schumpeterian Model of Top Income Inequality. *Journal of Political Economy*, 126(5):1785–1825, 2018.

Kahneman, Daniel, Jack L. Knetsch, and Richard H. Thaler. Fairness and the Assumptions of Economics. *The Journal of Business*, 59(4), 1986.

Konow, James. Fair and square: The four sides of distributive justice. *Journal of Economic Behavior and Organization*, 46(2):137–164, 2001.

Konow, James. Is fairness in the eye of the beholder? An impartial spectator analysis of justice. *Social Choice and Welfare*, 33(1):101–127, 2009.

Krawczyk, Michał. A glimpse through the veil of ignorance: Equality of opportunity and support for redistribution. *Journal of Public Economics*, 94(1-2):131–141, 2010.

Kritikos, Alexander and Friedel Bolle. Distributional concerns: Equity- or efficiency-oriented? *Economics Letters*, 73(3):333–338, 2001.

Kuznets, Simon. Economic Growth and Income Inequality. *The American Economic Review*, 45(1): 1–28, 1955.

Lahey, Joanna N. Age, Women and Hiring An Experimental Study. *The Journal of Human Resources*, 43(1):30–56, 2008.

Lane, Tom. Discrimination in the laboratory: A meta-analysis of economics experiments. *European Economic Review*, 90:375–402, 2016.

Lazear, Edward P. Performance pay and productivity. *American Economic Review*, 90(5):1346–1361, 2000.

Lefgren, Lars J., David P. Sims, and Olga B. Stoddard. Effort, luck, and voting for redistribution. *Journal of Public Economics*, 143:89–97, 2016.

Levitt, Steven D and John A List. What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World? *Journal of Economic Perspectives*, 21(2):153–174, 2007.

Malamud, Ofer. Discovering one's talent: Learning from academic specialization. *Industrial and Labor Relations Review*, 64(2):375–405, 2011.

Mudde, Cas. Populism: An Ideational Approach. In *The Oxford Handbook of Populism*, number October. 2017.

Murtin, Fabrice, Lara Fleischer, Vincent Siegerink, Arnstein Aassve, Yann Algan, Santiago Gonzalez, Gianluca Grimalda, Rafael Hortala, Vallve Soonhee, Kim David, Lee Zsuzsanna, Lonti Louis, Putterman Ulrich, and Schmidt Conal. Trust and its Determinants : Evidence from the Trustlab Experiment. 2018.

Norris, Pippa and Ronald Inglehart. *Cultural backlash: Trump, Brexit, and authoritarian populism.* Cambridge University Press, 2019.

OECD, . *OECD Guidelines on Measuring Trust.* 2017.

Oreopoulos, Philip. Why Do Skilled Immigrants Struggle in the Labor Market? A Field Experiment with Thirteen Thousand Resumes. *American Economic Journal: Economic Policy*, 3(November): 148–171, 2011.

Ostry, Jonathan D, Prakash Loungani, and Andrew Berg. *Confronting inequality: how societies can choose inclusive growth*. Columbia University Press, 2019.

Oxoby, Robert J. and John Spraggon. Mine and yours: Property rights in dictator games. *Journal of Economic Behavior and Organization*, 65(3-4):703–713, 2008.

Piketty, Thomas. Piketty social mobility and redistributive politics. *The Quarterly Journal of Economics*, 110(3):1057–1106, 1995.

Piketty, Thomas. *Le capital au XXIe siècle*. Le Seuil, 2013.

Piketty, Thomas. *Capital et Idéologie*. Le Seuil, 2019.

Piketty, Thomas, Emmanuel Saez, and Gabriel Zucman. Distributional National Accounts: Methods and Estimates for the United States. *Quarterly Journal of Economics*, 133(2):553–609, 2018.

Rabin, Matthew. Incorporating Fairness into Game Theory and Economics. *American Economic Review*, 83(5):1281–1302, 1993.

Rand, David G, Thomas Pfeiffer, Anna Dreber, Rachel W Sheketoff, Nils C Wernerfelt, and Yochai Benkler. Dynamic remodeling of in-group bias during the 2008 presidential election. *Proceedings of the National Academy of Sciences*, 106(15):6187–6191, 2009.

Rawls, John. *A theory of justice*. 1971.

Rawls, John. *Justice as Fairness, A Restatement*. Harvard University Press, 2001.

Romano, Angelo, Daniel Balliet, Toshio Yamagishi, and James H. Liu. Parochial trust and cooperation across 17 societies. *Proceedings of the National Academy of Sciences*, 114(48):12702–12707, 2017.

Rosanvallon, Pierre. *La société des égaux*. 2011.

Saez, Emmanuel and Gabriel Zucman. *Le Triomphe de l'injustice. Richesse, évasion fiscale et démocratie*. Le Seuil, 2020.

Sen, Amartya. *Inequality reexamined*. Oxford University Press, 1992.

Shearer, Bruce. Piece rates, fixed wages and incentives: Evidence from a field experiment. *Review of Economic Studies*, 71(2):513–534, 2004.

Smith, Adam. *The Theory of Moral Sentiments*. 1759.

Solis, Alex. Credit Access and College Enrollment Alex Solis. *Journal of Political Economy*, 125(2): 562–622, 2017.

Starmans, Christina, Mark Sheskin, and Paul Bloom. Why people prefer unequal societies. *Nature Human Behaviour*, 1(4):1–7, 2017.

Stephens-Davidowitz, Seth. The cost of racial animus on a black candidate : Evidence using Google search data. *Journal of Public Economics*, 118:26–40, 2014.

Stinebrickner, Ralph and Todd R. Stinebrickner. A major in science? Initial beliefs and final outcomes for college major and dropout. *Review of Economic Studies*, 81(1):426–472, 2014.

Tajfel, Henri, M.G. Billig, R.P. Bundy, and Claude Flament. Social categorization and intergroup behaviour. *European Journal of Social Psychology*, 1(2):149–178, 1971.

Tisserand, Jean-Christian. Ultimatum Game : A Meta-Analysis of the past three Decades of Experimental Research. *Proceedings of International Academic Conferences*, 2014.

Valfort, Marie-anne. Anti-Muslim Discrimination in France : Evidence from a Field Experiment. 2018.

Weisel, Ori and Robert Böhm. "Ingroup love" and "outgroup hate" in intergroup conflict between natural groups. *Journal of Experimental Social Psychology*, 60:110–120, 2015.

Wiswall, Matthew and Basit Zafar. Determinants of college major choice: Identification using an information experiment. *Review of Economic Studies*, 82(2):791–824, 2015.

Zafar, Basit. How do College Students Form Expectations? 2011.

Zafar, Basit. College Major Choice and the Gender Gap. *Journal of Human Resources*, 48(3):545–595, 2013.

Zizzo, and Oswald. Are People Willing to Pay to Reduce Others' Incomes? *Annales d'Économie et de Statistique*, (63/64):39, 2001.

Chapter 1

Updating of academic tastes and ability signals

This chapter is a joint work with Ghazala Azmat, Anne Boring and Roberto Galbiati.

1.1 Introduction

There are large institutional differences in the way students choose their major in higher educational systems. In some countries like the US, students experiment with different fields before specializing, while in other countries like the UK, majors are chosen even before entering university. A low-flexibility system has the drawback of hindering changes in academic tracks, but has the advantage of accelerating skills accumulation (Malamud, 2011). Being locked into a specific academic track at an early stage has important consequences, since the field of study in which students specialize impacts the returns to higher education and labor market outcomes (Altonji et al., 2016).

To assess the relative merits of each system, it is necessary to understand how beliefs about academic preferences are formed, to what extent they are updated over time, and for what reasons this occurs. This is notoriously difficult because major choices are often made under constraints and with incomplete information. Before entering college, students have been exposed to only fragmentary information about different fields of study. They hold a certain set of – potentially biased – beliefs regarding their academic ability and preferences for different fields. Moreover, researchers often only observe the revealed choices and not the entire decision-making process. Through the first college courses they take, the grades they receive and the peers and teachers to whom they are exposed, students receive new information that causes them to update their beliefs. A student's choice of a major or a master's degree is thus the result of a complex learning process (Arcidiacono, 2004; Stinebrickner and Stinebrickner, 2014b,a).

In this paper, we explore this process by documenting how and why students' beliefs about their academic preferences are updated throughout their undergraduate studies. To conduct our analysis, we use a unique dataset from a French university specializing in social sciences – Sciences Po. This dataset has several distinct advantages. First, it is longitudinal, as we can follow students over the course of their studies, from their last year of high school to their graduate studies. We measure academic tastes at two points in time using the content of letters of motivation that all students have to write. The first one is submitted as part of their application package to Sciences Po, while they are still in high school, and the second one is written two years later when they justify their choices of universities for the mandatory third-year study abroad program. Eliciting tastes and beliefs is traditionally carried out by conducting surveys (Stinebrickner and Stinebrickner, 2014b; Zafar, 2013), which often suffer from attrition and a low-attention bias. In contrast, letters of motivation are high-stake components of the application procedures and this type of data is exhaustive. We also collect administrative data on students' academic choices: their decisions regarding specialist courses in the second year, and master's degrees. Second, this university education has an intermediary status concerning flexibility, in the sense that students specialize late in their studies but they have no latitude in their choices of courses in the first year. The curriculum is identical for all first-year students; they must try all major branches of social sciences (Economics, History, Law, Political Science and Sociology) before making any important academic choices. This feature is rare in other universities where students often self-select into their courses and do not try out their affinities and abilities in all majors. Self-selection is therefore not a concern in our context. Finally, there is no supply-side constraint on master's degree choices, which alleviates strategic concerns that might arise if students had to anticipate what other students would choose if they wanted to shy away from competition, or on the contrary, challenge themselves.

Throughout the paper, we use different concepts pertaining to the economics of education literature, which we consider useful to clarify here. Students have underlying *academic preferences*, which we assume to be fixed over the time frame of our study. Students discover them as they undertake a learning process, and in this way develop *beliefs* about their true preferences. With letters of motivation, we measure these beliefs in the form of self-declared academic aspirations, which we call *academic tastes* for the sake of simplicity. We thus make a conceptual distinction between tastes and preferences: tastes may be updated and change over time – following new informational signals – while preferences are fixed and unknown ex ante, although students have more or less strong *priors*. Finally, preferences are revealed themselves in the form of master's degree choices.

We find that both academic tastes elicited in high school and in the second year of university correlate with revealed choices, with this relationship being stronger with the latter than the former. This indicates that students are changing their mind after exposure to the program. We then investigate the reasons why this updating process takes place. We focus on three types of information available to students. First, we use data on students' priors. They do not start higher education with the same degree of information about their academic preferences. Some students may hold very strong beliefs and already be passionate about a subject, while others are more uncertain. We proxy the degree of certainty of initial beliefs using a keyword-based measurement of the strength of the dominant academic taste of the student in high school. Second, we study the effect of several ability signals students receive: (i) they learn about their performance in the field in which they initially wanted to major; (ii) they learn about their ability in other fields; (iii) they can compare their performance with respect to their peers. We also study whether the precision of the information about peers' performance matters for taste updating.

We find that about 75% of students update their academic tastes over the first two years of their studies. Ability signals play an important role in this process, they are drivers of stickiness when they indicate a good fit with initial tastes, and drivers of change when they are a signal that the student has a higher ability in a different field. Nonetheless, for students who have particularly strong beliefs that they will eventually choose a certain field of study, we find a stronger persistence in tastes, even when students receive negative information about their academic ability. Further analyses suggest that relative performance does not play a significant role in updating students' tastes. A higher degree of precision regarding peers' performance does not change this result. We do not find differences in the process of taste updating by gender, nor by the socioeconomic background of students.

We contribute to the large body of economic literature on the factors that determine students' choices for majors (Altonji et al., 2012, 2016, for reviews), and more specifically, papers incorporating subjective beliefs and expectations about educational or labor market outcomes (Montmarquette et al., 2002; Arcidiacono, 2004; Zafar, 2011; Beffy et al., 2012; Zafar, 2013; Stinebrickner and Stinebrickner, 2014b; Wiswall and Zafar, 2015; Delavande and Zafar, 2019). Stinebrickner and Stinebrickner (2014b) surveyed students' beliefs regarding their probability of achieving certain outcomes (dropping out or choosing a certain major) twelve times each year and connected these beliefs to realized outcomes. They found that students tend to be overoptimistic about their chances to graduate in science, and that they revise their beliefs downward following ability signals. We depart from earlier works by introducing a novel type of data – letters of motivation – and by simultaneously looking at a wider array of ability signals, including relative ones.

Our paper is also consistent with papers that examine the impact of ability signals on students'

choices, both in terms of absolute and relative performance (Chevalier et al., 2009; Ost, 2010; Rask, 2010; Stinebrickner and Stinebrickner, 2014a). Elsner et al. (2018) study the ability signal that students obtain from their ordinal rank. They use a quasi-natural experiment at a business school university, where students are randomly assigned to teaching sections. They causally identify the impact of ordinal rank on student outcomes, controlling for students' own individual ability. They find that higher-ranked students tend to be more persistent in their field of study. They also find large gender differences: a higher rank causes male students to increase their effort as measured by the number of study hours, but does not impact effort by female students. Other papers that study the impact of class rank are studies of primary and secondary schools (Murphy and Weinhardt, 2018; Elsner and Isphording, 2017, 2018). Elsner and Isphording (2017) in particular find that a higher high school rank is related to students having higher expectations regarding their future labor market outcomes, and beliefs in their ability. We depart from these studies by looking at multiple types of performance signals, absolute and relative, and with different degrees of precision.

Finally, this paper contributes to the literature on the differences in study choices, aspirations, expectations and beliefs between male and female students (Filippin and Ichino, 2005; Beaman et al., 2012; Zafar, 2013; Speer, 2017; Astorne-Figari and Speer, 2018), and between students from different socioeconomic backgrounds (Hoxby and Turner, 2015; Carlana et al., 2017; Goux et al., 2017; Landaud et al.; Guyon and Huillery, 2019).

This paper is structured as follows: we describe the data we use and how we measure academic tastes in Section 1.2. We then show that these elicited tastes predict observed academic choices (Section 1.3). We detail our main results regarding the determinants of taste updating in Section 1.4, before concluding (Section 4.5).

1.2 Context and data

We collected data on the cohort of students admitted to Sciences Po in 2014. We detail below how we elicited students' dominant academic tastes using information provided by students in their letters of motivation (Section 1.2.1). We also describe the administrative data we used for the analysis of choices once students are in college (Section 1.2.2).

1.2.1 Eliciting academic tastes

To measure students' academic tastes and how these tastes change over time, we elicited information provided by students in two letters of letters of motivation they are required to write. First, all students applying to Sciences Po have to write a letter of motivation when they are in their final year of high school. In this letter, they explain why they are interested in obtaining a bachelor's degree from the university, and how the university's course offering corresponds to their academic and professional ambitions. We use the information that admitted students write in these letters to elicit their beliefs concerning their true academic preferences, i.e. academic tastes. Students can only apply when they are in their final year of high school, which enables the comparison of tastes before any students are treated with any information about their aptitude for higher education. Sciences Po specializes in social sciences, and no field of study is looked upon more favorably than any other by the Admission Committee. Students therefore have no incentive to be strategic when describing their

academic tastes in these letters.[1] We describe the guidelines that students receive on writing their letter of motivation in Appendix 1.A.

Second, all students admitted to Sciences Po have to spend the third and final year of their bachelor's degree abroad. The majority of students choose to study abroad. These students must provide a ranked list of six universities out of the 470 institutions on all continents that have a partnership with Sciences Po. They enclose a two-page letter to justify their rankings of universities. The administration, which then selects the university to which the student will be assigned, encourages students to explain in these letters why they want to study at each of the six universities. Students therefore provide details about the courses they would like to take and in which field of study they are most interested. They must define a project that is consistent with their academic and professional aspirations.[2] Students write these letters of motivation in the second year of their bachelor's degree, i.e. half-way through their undergraduate studies. We elicit information from these letters to measure whether students' beliefs about their academic preferences have changed since their high school letters. We can then impute changes to the informational treatments that students received during their first year of studies.[3]

We use both letters to elicit students' beliefs about their academic preferences in the form of dominant tastes in either one of the five core fields of study at the university: Economics, History, Law, Political Science and Sociology. To construct measurements of academic tastes, we combine a qualitative and a quantitative approach, relying on hand-labeling to identify dominant taste and keyword counts linked to the five fields of interest to obtain an objective measurement of the intensity of these tastes.[4] High-school taste intensity is a proxy for the degree of certainty of students' beliefs about their academic preferences. We can also interpret hand-labeled dominant tastes as an extensive margin construct and the keyword-based method as an intensive margin measure.

Dominant tastes were hand-coded by four research assistants (RAs) who read all the letters to determine students' dominant academic tastes. A student's "dominant taste" is our main variable of interest. Eliciting students' beliefs about their preferences through hand-labeling was a relatively

[1] Sciences Po is a highly selective university. To be admitted, applicants are likely to be keen to show themselves in their best light, and some may embellish their own achievements. However, students do not have incentives to hide their true beliefs about their *tastes*. They are encouraged to describe their "personal and professional goals" (see Appendix 1.A). If they do not have any clear aspirations yet or are still hesitant, writing the letter encourages them to reflect on their own preferences. The letter may be the result of this personal reflection process, acting as a revealer. Of course, there is always the possibility of a minority of applicants having no particular tastes and inventing them. Since all of the five tastes are legitimate (as they constitute the core of the first-year curriculum), there is no reason for applicants to express more interest in one field than another.

[2] Boring and Brown (2020) provide a detailed description of this exchange program and the mechanism leading to students' allocations to universities.

[3] Concerning the second-year letter of motivation for the third year abroad, being truthful is a good strategy, given that students will eventually have to visit the departments of the universities to which they apply. If a student dislikes economics, applying to a university's economics department would be irrational.

[4] A third type of approach would infer academic tastes using machine learning and would require hand-labeled data in order to train the algorithms. There is a flourishing literature on textual analysis, machine-learning methods and their application to economics (Gentzkow and Shapiro, 2010; Gentzkow et al., 2017; Hansen et al., 2018). We do not use this approach, as there are insufficient data points to train an algorithm effectively. Indeed, the 2014 cohort is composed of about 1,000 students. Moreover, this is not a simple binary classification task as in many machine learning applications. We would need an algorithm capable of detecting the dominant taste among the five that interest us. This is a complex task because we need to predict a non-binary outcome composed of five categories, some of which are under-represented, such as having a dominant taste for sociology. Only 2.8% of students have a dominant taste in sociology (31 observations). To effectively train a classification algorithm, the machine must learn from thousands of instances in each category. The data we have is simply too limited for machine learning. Moreover, we have letters written in both French (90%) and English (10%). A machine learning algorithm would need to be trained in both languages separately, which further reduces our statistical power. However, we can easily scan the letters for both English and French keywords separately.

easy task to perform, since the letters are fairly substantial, and students write the letters with the purpose of describing their academic tastes. The RAs coded and ranked each student's four preferred academic fields.[5] These fields could be any of the five main social sciences fields that we focus on in our analysis or any other field in which students can express an interest (foreign languages, literature, philosophy, mathematics, journalism, etc.). We then aggregated outcomes into seven categories, as described in Table 1.A1: "Economics/Finance/Business", "History", "Law", "Political Science and International Relations", "Sociology", "Other taste" and "No expressed taste. The "Other" category includes tastes that do not correspond to any of the five main fields (e.g. an interest in urban studies or journalism). To construct our measurement of dominant taste, we proceed in the following way: if the first-ranked taste is not one of the five fields of interest, we move down and use the second-ranked taste. If the second-ranked taste is not yet one of the five fields, we move down to the third-ranked taste, etc. as far as the fourth-ranked taste. If none of the five fields is ever identified by our hand-labeled procedure, but the student nonetheless has another identified taste (e.g. for urban studies), then we code the dominant taste as "Other taste". If the student does not mention any academic field, we code the dominant taste as "No expressed taste". In the admission letters, 8.86% of students have another dominant taste and 6% have no dominant taste. This situation mainly occurs when students only talk about extra-curricular activities (volunteering, traveling etc.). We further detail this hand-labeling procedure in Appendix 1.A. Section 1.4.1 describes how we measure the intensity of the dominant taste using keyword counts and Appendix 1.A.3 details the procedure we followed to select these keywords and also shows that keywords correlate well with – and are good predictors of – hand-labeled dominant tastes.

Table 1.1 shows the flows of tastes between high school and the second year of college, in the form of a transition matrix, which includes frequencies (first number of a cell), row and column percentages (with high school tastes in rows and second year tastes in columns). Figure 1.1 summarizes the data visually using a Sankey diagram. The column percentages in the Total column (right of the table) display the distribution of dominant high school tastes.

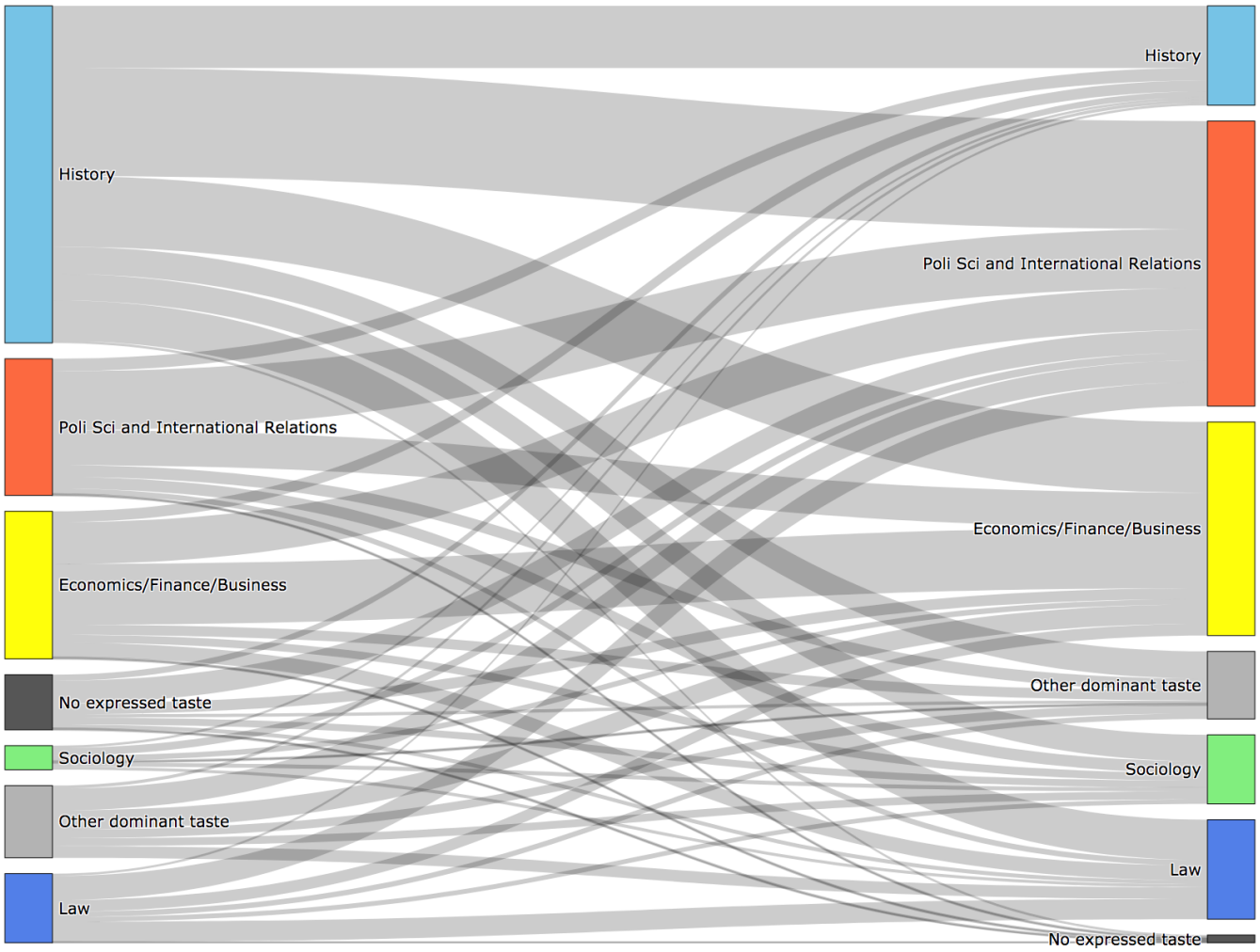
We find that students tend to express higher levels of interest for fields to which they are more likely to have been exposed in high school. The most popular dominant taste is History (for 40.2% of students), which is the field that all students studied in high school, irrespective of whether they specialized in sciences, social sciences, or humanities before starting college. All students had therefore received information about this field before starting college. The second- and third-most dominant tastes are Economics/Finance/Business (17.7%) and Political Science/International Relations (16.4%). High school students are generally less exposed to Law, and so only 8.3% express a strong interest in this field. Sociology comes last with 2.9% students having a dominant taste for this field at the end of high school.

Halfway through their undergraduate studies, we find that the majority of students (about 75%) have changed their beliefs about their preferences. The distribution of dominant tastes in the second year is described in the Total row (bottom of the table). History declines sharply, with only 11.9% of students still expressing a strong interest in this field in the second year. The History row in Table 1.1, which gives the distribution of the second-year tastes of students who favored History in high school, suggests that many students switch to Political Science/International Relations (32.1%) or develop a dominant taste for Economics/Finance/Business (21%). The fields that attract the

[5]To account for heterogeneity in RAs, all our regressions included RA fixed effects.

largest number of students in second year are Political Science/International Relations (33.8%) and Economics/Finance/Business (25.6%).

Figure 1.1 – Evolution of academic tastes between high school and second year



Notes: The Sankey diagram describes the flows of dominant academic tastes from high school (categories on far left) to the second year (categories on far right). The size of the colored rectangles represents the distribution of dominant tastes. The size of the gray arrows is proportional to the size of the flows from one dominant taste to another between both time periods.

1.2.2 Academic data

We collected data on students from their first year to their master's degree choice. The database contains the socio-demographic characteristics of students' gender, family background, nationality, and type of high school diploma[6], the campus on which the student is studying[7], and whether the student is enrolled in a dual degree program.

The database also includes information about the first-year and second-year college courses in which students are enrolled, the grades they obtained and their peer groups.[8] A key feature of the

[6]More specifically, the database includes information about whether students earned their diploma from a foreign high school or from a French high school (in which case we know about their specialization in the Scientific, Literature, Economics and Social Sciences or Technical tracks).

[7]The university has several campuses, in Paris, Reims, Poitiers, Nancy, Le Havre, Dijon and Menton. Except for Paris, campuses are specialized in a geographical region and specific languages.

[8]However, the database does not include any measurement of a student's academic ability prior to enrolling in higher

Table 1.1 – Transition matrix of dominant tastes

	Dominant taste in 2nd year							
Dominant taste in high school	**Econ/Finance/Biz**	**History**	**Law**	**Poli Sci/Inter. Rel.**	**Sociology**	**Other taste**	**No expressed taste**	**Total**
Econ/Finance/Biz	78	14	19	54	10	13	2	190
(row %)	41.1%	7.4%	10.0%	28.4%	5.3%	6.8%	1.1%	100.0%
(col %)	28.4%	11.2%	14.8%	14.8%	11.2%	14.9%	25.0%	17.7%
History	91	80	52	139	34	35	2	433
(row %)	21.0%	18.5%	12.0%	32.1%	7.9%	8.1%	0.5%	100.0%
(col %)	33.1%	64.0%	40.6%	38.2%	38.2%	40.2%	25.0%	40.2%
Law	15	3	26	31	6	7	1	89
(row %)	16.9%	3.4%	29.2%	34.8%	6.7%	7.9%	1.1%	100.0%
(col %)	5.5%	2.4%	20.3%	8.5%	6.7%	8.0%	12.5%	8.3%
Poli Sci/Inter. Rel.	45	16	7	76	15	15	2	176
(row %)	25.6%	9.1%	4.0%	43.2%	8.5%	8.5%	1.1%	100.0%
(col %)	16.4%	12.8%	5.5%	20.9%	16.9%	17.2%	25.0%	16.4%
Sociology	7	3	4	10	5	2	0	31
(row %)	22.6%	9.7%	12.9%	32.3%	16.1%	6.5%	0.0%	100.0%
(col %)	2.5%	2.4%	3.1%	2.7%	5.6%	2.3%	0.0%	2.9%
Other taste	25	4	15	28	10	11	0	93
(row %)	26.9%	4.3%	16.1%	30.1%	10.8%	11.8%	0.0%	100.0%
(col %)	9.1%	3.2%	11.7%	7.7%	11.2%	12.6%	0.0%	8.6%
No expressed taste	14	5	5	26	9	4	1	64
(row %)	21.9%	7.8%	7.8%	40.6%	14.1%	6.2%	1.6%	100.0%
(col %)	5.1%	4.0%	3.9%	7.1%	10.1%	4.6%	12.5%	5.9%
Total	275	125	128	364	89	87	8	1,076
(row %)	25.6%	11.6%	11.9%	33.8%	8.3%	8.1%	0.7%	100.0%
(col %)	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Notes: The two-way table shows the joint distribution of dominant tastes in high school (in rows) and in dominant second-year tastes (in columns). It shows that 78 students had a dominant taste for Economics/Finance both in high school and in the second year, and they represent 41.1% of the students who had a dominant taste for Economics/Finance in high school, and 28.4% of the students who had a dominant taste for Economics/Finance in the second year.

educational setting is that all students, across all campuses, take mandatory courses in all five main fields of study. We can therefore measure how receiving information in each field may impact students' beliefs regarding their tastes for each field. In their first year, all students follow two courses in Economics (Microeconomics and Macroeconomics), one course in History, one in Law, one in Political Science, and one in Sociology. These mandatory courses all include plenary lectures and tutorials (24 hours each). Students obtain a tutorial grade, which is based on a continuous assessment carried out by the teaching assistant, and a final exam, which is graded anonymously. We use the final course grade, which is equal to 1/3 final exam + 2/3 tutorial grade.[9] In France, the grading scheme ranges from 0 to 20 where 10 is the passing grade. In social sciences, it is hard to attain a grade above 18 and 14 is considered a very decent mark. Appendix Table 3.E1 shows the average final course grades students obtained in each first-year course. All tutorials are attended with the same classmates. We describe how these tutorial groups are formed in Appendix 1.B.

1.3 Academic tastes predict observed academic choices

In this section, we reveal two findings: (i) the elicited academic tastes are meaningful in the sense that they predict observed academic choices; (ii) dominant second-year tastes correlate more strongly with choices than high-school tastes, which suggests that taste updating takes places over time.

We run the following OLS regressions at the student level i, for each field j separately.

$$\begin{aligned} Y_i = \alpha + \beta_1 \mathbb{1}(\textit{High school taste in field } j) + \beta_2 \mathbb{1}(\textit{2nd year taste in field } j) \\ + X_i\delta + \pi_i + \gamma_i + u_i \end{aligned} \tag{1.1}$$

Where Y_i corresponds to the choice of master's degree in Table 1.2. In Appendix 3.A, we also show results based on two other outcome variables: the choice of second-year specialist courses [10] (Table 1.D3) and the choice of second-year elective courses[11] (Table 1.D4). Students also have more freedom in choosing the courses they want to take in their second year than in their first year. The main explanatory variables are two dummies for whether the student has a dominant taste in high school and in a given field j in the second year. We control for a vector X_i of socio-demographic characteristics: gender, high family socio-economic status, three baccalaureate dummies, two admission track dummies, six region of nationality dummies and the log of the number of words written in both letters of motivation. All these variables are described in Appendix Table 3.E1. We further control for research assistant fixed effects (π_i) in order to address fixed differences in hand-labeling across the four research assistants who read the letters. We also control for campus fixed effects to account for the specific sorting of students across study locations (γ_i).

In Table 1.2, we regress binary variables for whether the student chooses a master's degree in Economics, Business or Finance (columns 1 and 2), in Law (columns 3 and 4), and Political Science or International Relations (columns 5 and 6), on dummies for whether the corresponding taste was

education.

[9] The final exam is generally a written exam common to the whole cohort. The tutorial grade is a continuous assessment carried out during the tutorials. It includes a midterm (written exam) – also common to the whole cohort – and additional grades that are tutorial-specific (quizzes, oral presentations, readings summaries, etc.).

[10] Lectures specializing in one of the five fields.

[11] Small-group courses that students have to choose, on top of the main curriculum.

Table 1.2 – Master's degree choice and dominant tastes

	Econ/Finance/Business master		Law master		Poli Sci/Inter. Rel. master	
	(1)	(2)	(3)	(4)	(5)	(6)
Econ/Finance dominant taste in HS	0.136***	0.0740**				
	(0.0339)	(0.0296)				
Econ/Finance dominant taste in 2nd year		0.345***				
		(0.0307)				
Law dominant taste in HS			0.203***	0.113***		
			(0.0500)	(0.0388)		
Law dominant taste in 2nd year				0.473***		
				(0.0448)		
Poli Sci/ Inter. Rel. dominant taste in HS					0.0890**	0.0777**
					(0.0385)	(0.0377)
Poli Sci/ Inter. Rel. dominant taste in 2nd year						0.121***
						(0.0269)
Constant	0.401	0.370	-0.237	-0.226	0.0175	0.0555
	(0.383)	(0.343)	(0.322)	(0.282)	(0.406)	(0.403)
Ind. Controls	Yes	Yes	Yes	Yes	Yes	Yes
RA FE	Yes	Yes	Yes	Yes	Yes	Yes
Campus FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1074	1074	1074	1074	1074	1074
R^2	0.050	0.243	0.057	0.265	0.099	0.118

Notes: OLS regressions where the dependent variable is binary and equal to 1 if the student chooses a master's degree in Economics or Finance (columns 1 and 2), in Law (columns 3 and 4) or in Political Science/International Relations (columns 5 and 6). The explanatory variables are binary variables for whether the students have a dominant taste in the corresponding field in high school (HS) or in the second year. Individual controls include binary variables for gender, high socio-economic background, 3 baccalaureate dummies, 2 admission track dummies, 6 region of nationality dummies and the log of the number of words written in both letters of motivation. See Appendix table 3.E1 for a more detailed description of these variables. RA fixed effects are 3 dummy variables controlling for the differences in hand-labeling practices across our 4 research assistants. Campus FEs are 6 dummy variables controlling for the students' different campuses (Dijon, Le Havre, Menton, Nancy, Paris, Poitiers and Reims). Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

dominant in high school or in the second year. We do not show the results of tastes for History and Sociology because there are too few students enrolled in the master's degrees related to these fields.[12] Table 1.2 shows that both dominant high-school taste and dominant second-year taste in field j significantly predict choosing the associated master's degree, for all three fields.

After controlling for second-year tastes, the coefficient of dominant high-school taste decreases in size, implying that high-school and second-year tastes are positively correlated. Moreover, second-year tastes are more strongly associated with choices of graduate studies. This result indicates that students discover their true preferences as they progress through their bachelor studies. Having a dominant taste for economics-related topics in the second year increases the likelihood of pursuing graduate studies in that field by 34.5 percentage points, while having such a dominant taste in high school only increases the likelihood of choosing the corresponding master's degrees by 7.4 percentage points (significant at the 5% level). Second-year tastes also have greater predictive power. The R-squared is almost multiplied by five when second-year tastes are included for both Economics/Finance/Business degrees and Law degrees, and by two for Political Science/International Relations degrees.

Tables 1.D3 and 1.D4 perform the same exercise for the choice of specialist courses and elective courses respectively. We can focus on the five dominant tastes since courses on all topics are available to students. The only predictor that we can use is dominant high-school taste, because second-year letters of motivation are written *after* the students have enrolled in second year courses. We still find a positive and significant correlation between tastes and choices for Economics/Finance/Business, History and Law, but the evidence is weaker for Political Science/International Relations and Sociology.[13]

1.4 Determinants of taste updating

We now turn to the learning process that leads students to update their beliefs about their true preferences. Section 1.4.1 describes our general empirical strategy. Section 1.4.2 shows that both ability signals and the strength of initial tastes matter in the taste updating process.

1.4.1 Empirical strategy

To characterize the evolution of dominant taste, we create a binary dependent variable for whether or not the student's dominant taste changed between high school and the second year. Dominant tastes are measured using hand-labeled data (see Section 1.2.1 and Appendix 1.A). Our empirical model is described by Equation (1.2) below:

$$\begin{aligned} \textit{Change taste}_i = \alpha + \beta_1\ \textit{Grade}_i + \beta_2\ \mathbb{1}(\textit{Best grade not in HS taste}) + \beta_3\ \textit{Relative ability}_i \\ +\beta_4\ \textit{Strength HS taste}_i \\ +X_i\delta + \pi_i + \gamma_i + \eta_i + u_i \end{aligned} \tag{1.2}$$

The explanatory variables account for several dimensions of the trade-off that taste updating involves. Section 1.4.1 details the variables relative to ability signals, while 1.4.1 describes the variables related

[12] Only research-oriented graduate studies exist for History and Sociology at Sciences Po. Only 9 students are enrolled in the History graduate school and 12 in the Sociology graduate school. The other master's degrees that Sciences Po offers, such as urbanism and public affairs (see Table 1.D2), are interdisciplinary and cannot be tied to a single field.

[13] This could be explained by the fact that Sciences Po does not offer students an add-and-drop period in the second year, which reduces the possibility of enrolling in the most appropriate courses.

to the importance of having strong priors.

Three types of ability signals

We focus on three ability signals: the first one, $Grade_i$, is the grade received in the first year *in the field of the student's dominant high school taste*. It measures the match between the student's ability and her high-school taste. We use the final (weighted) course grade. We standardize this grade at the field level, so that it has a mean equal to 0 and a standard deviation equal to 1, within each field.[14] This accounts for the fact that having a 14 in History and a 14 in Sociology does not deliver the same signal. As the Appendix Table 3.E1 shows, History grades are lower on average (12.95 out of 20) and Sociology marks tend to be the highest (14.26 on average).[15] We expect $\beta_1 < 0$, i.e. having a good grade in one's dominant high-school taste should be a driver of taste persistence ($Change\ taste_i = 0$).

The next performance variable we use, $\mathbb{1}(Best\ Grade\ not\ in\ HS\ taste)$, is a signal of the student's ability in *other* fields and is thus a potential driver of change in dominant taste ($\beta_2 > 0$). It is a binary variable equal to 1 if the student's highest performance is *not* in the field of her dominant high-school taste. This time, to compare the student's grades in different subjects, we use non-standardized figures. It is likely that students compare absolute grades, given that they are more salient. Moreover, students have only limited information about the grade distribution of their cohort, so it may be difficult to make comparisons across standardized grades.

$Relative\ ability_i$ constitutes a third type of ability signal, but this time it has a comparative dimension since it concerns the performance of student i relative to her peers, in her high school taste. We expect that students are more likely to maintain their initial taste if they are comparatively better at it (hence $\beta_3 < 0$). We measure $Relative\ ability_i$ in two ways, in order to account for students' degree of precision vis-à-vis their peers' grades. The first variable is the student's ranking in terms of the final weighted grade, among her local peers, in the field of her dominant high-school taste. A high value for the ranking variable means that the student has a top rank and is thus of higher ability than many of her local peers in the field of her initial taste. We normalize this ranking by dividing by the peer-group size to allow for comparisons across groups of different sizes.[16] Local peers are defined as the smallest observable group of peers with which the student interacts regularly throughout the first year. When available[17], we use the peers from the tutorial groups, which are groups of about 15 students formed on a quasi-random basis in August, before the start of the academic year and before students start interacting with each other. These groups meet for six hours per week during three tutorial classes each semester. We provide a detailed description of the formation process of these groups in Appendix 1.B. We show the results of several random allocation tests suggesting that group formation is as good as random, once students enrolled in dual bachelor's degrees are dropped.[18] When data on tutorial groups is not available[19], we use the other students in the same cohort on the

[14] Note that the variable is standardized *ex ante* . All students receive five grades in their first year at Sciences Po. We first standardize each one of these five variables so that within the cohort, they all have mean 0 and standard deviation 1. Then, we create the $Grade_i$ variable that is the standardized grade in the field of the student's dominant high-school taste. The results are similar if we use non-standardized grades (not shown).

[15] This grading difference can be explained by the different types of examinations conducted in each field. History relies on traditional "French-style" dissertations with strong analytical and abstraction requirements, and for which grading is notoriously tougher. Sociology relies on shorter lecture questions, for which is it easier to obtain good grades.

[16] If we do not normalize, the best student of a group could be ranked 40th and the best student of another group could be ranked 20th, simply because the study group is larger in the first case.

[17] On the Paris and Nancy campuses (62% of the sample)

[18] Students enrolled in dual degrees are highly selected in terms of both tastes and ability.

[19] On the other campuses, tutorial groups are not fixed and student change peers at each tutorial

same campus as the relevant local peers.[20] In Appendix 1.B, we show that there is a higher risk of self-selection regarding the allocation of students into *campuses* compared to the allocation of students into *tutorial groups*. Therefore, we provide robustness checks on results focusing on tutorial groups only, which additionally drop students enrolled in dual degrees (see Table 1.D5).

The second measurement of $Relative\ ability_i$ is based on information known with certainty by students since it is on their transcripts. Students have access to an approximate measurement of their position within the cohort in the form of letter grades entered on their transcripts at the end of each semester. They are awarded a grade "A" if they are among the top 10%, a "B" if they are among the next 25%, "C" for the next 30%, "D" for the next 25%, and "E" for the bottom 10%. We construct a binary variable equal to 1 if the student is within the top 10% of her cohort in her dominant high-school taste. We do not use the rest of the letters because the relative performance signal contained in the other letters is much weaker; the categories are very broad from "B" onwards.

Strength of dominant high-school taste

Finally, to identify the strength of students' priors regarding their true academic preferences, we add the variable $Strength\ HS\ taste_i$, which measures the strength of the dominant high-school taste of the student i and is thus a driver of taste persistence ($\beta_4 < 0$). Intuitively, if $Strength\ HS\ taste_i$ is large, it means that students' belief distribution is more tightly centered around their true academic preferences and that uncertainty is lower. This variable is calculated as follows: for all five fields, we first compute the *High-school taste intensity* of the student i in field j as the share of keywords used in the letter about the field relative to the total number of keywords used in the five fields, as described by Equation (1.3). The procedure used to select keywords is described in Appendix 1.A.3.

$$High\ school\ taste\ intensity_{ij} = \frac{keyword\ counts_{ij}}{\sum_{j=1}^{5} keyword\ counts_{ij}} \tag{1.3}$$

$Strength\ HS\ taste_i$ is therefore the $High\ school\ taste\ intensity_{ij}$ of the student's *dominant* taste, according to hand-labeled data. This variable is, by definition, missing for students who do not have a dominant taste or have one in another field than the five mentioned (about 14% of the students).

Table 1.3 shows how high-school taste intensity varies across hand-labeled dominant tastes. Numbers on the diagonal of the top panel correspond to the mean $Strength\ HS\ taste_i$ across the five fields, i.e. high-school taste intensity in the dominant taste. The rest of the numbers in the top panel show the keyword intensity for the other (non-dominant) tastes. The bottom panel shows how high-school taste intensity varies with dominant second-year tastes.

We see that taste intensity is greater (closer to 1) on the diagonal, which shows that taste intensity and dominant tastes (based on hand-labeled data) are closely correlated. For instance, on average, students whose dominant taste in high school is History used 42% of History keywords in their high school letters of motivation, and only 5% of sociology keywords. This correlation between keyword-measured high-school taste intensity and hand-labeled dominant taste still exists in the second year (bottom panel of the table). Overall, these results are reassuring, given that both statistics are measuring the same thing – academic tastes – except that keywords measure the intensive margin and hand-labeled data the extensive margin. Appendix 1.A.3 describes in greater detail how keyword

[20] These campuses are fairly small in size (about 60 students for Dijon, Le Havre, Menton, Poitiers and 140 for Reims).

measurements of tastes are good predictors of hand-labeled dominant tastes.

Table 1.3 – High-school taste intensity by (hand-labeled) dominant taste

	High school taste intensity				
	Econ/Fin/Biz	History	Law	PoliSci/InterRel	Sociology
Dominant taste in high school					
Economics/Finance/Business	0.52	0.19	0.08	0.16	0.05
History	0.27	0.42	0.09	0.17	0.05
Law	0.23	0.13	0.45	0.14	0.04
Poli Sci/Inter. Rel.	0.32	0.23	0.05	0.37	0.03
Sociology	0.32	0.22	0.07	0.15	0.23
Other taste	0.38	0.29	0.07	0.23	0.04
No expressed taste	0.35	0.21	0.10	0.34	0.00
Total	0.33	0.30	0.11	0.21	0.05
Dominant taste in 2nd year					
Economics/Finance/Business	0.43	0.25	0.09	0.19	0.04
History	0.29	0.41	0.06	0.19	0.05
Law	0.29	0.27	0.21	0.19	0.04
Poli Sci/Inter. Rel.	0.30	0.30	0.12	0.24	0.04
Sociology	0.33	0.30	0.10	0.19	0.09
Other taste	0.30	0.31	0.14	0.21	0.05
No expressed taste	0.45	0.20	0.11	0.21	0.04
Total	0.33	0.30	0.11	0.21	0.05

Notes: High-school taste intensity is measured using high school letters and according to Equation (1.3). Dominant taste in high school and in the second year are measured using hand-labeled data. We read the table as follows: students who have a dominant taste for Law in high school use 45% of Law keywords in their high school letters of motivation (among all the keywords related to the five academic tastes).

The last row of Equation (1.2) shows the control variables we use. These variables are the same as those used in Equation (1.1) (Section 1.3), but we additionally control for fixed effects, η_i of dominant high-school tastes. We therefore compare students who had the same dominant high-school taste, but who received different grades, and whose initial taste was of a different strength. These fixed effects account for cohort trends in the evolution of taste distribution. For instance, students who liked History in high school are more likely to lose this taste than other students, simply because History was probably "over-rated" initially. If History lovers had specific characteristics, such as lower grades on average, then we would mistakenly attribute the decline of History as a dominant taste to low grades, whereas the drop would actually be driven by cohort-level trends. Focusing on within-taste variations enables this potential issue to be taken into account.

1.4.2 Both student-specific ability signals and dominant high-school tastes predict future tastes

In this subsection, we describe the effect of student-specific variables. Before turning to regressions, we can now provide some descriptive evidence for the relevance of ability signals and the strength of dominant high-school tastes in taste updating. The left-hand panel of Figure 1.2 shows the distribution of performance in the student's favorite field in high school (blue bars) and in the second year (white bars). Performance in the dominant taste is measured as an "internal" ranking. Each student obtains five grades at the end of the first year, which can be ranked. If a student obtains her best grade in her dominant high-school taste, then she will be added to the blue bar on the far left. On the contrary, if

she obtains her worst grade in her dominant high-school taste, she will be added to the blue bar on the far right. The same distribution is plotted with dominant second-year tastes (white bars). We observe two patterns: first, the blue bars have an increasing trend, which means that students' performance in their dominant high-school taste tends to be fairly poor. This result can be explained by the fact that 40.2% of the students have a dominant taste for History in high school, which happens to be the field in which it is harder to obtain a good grade (see Appendix Table 3.E1). Second, white bars representing the ability distribution relative to dominant second-year tastes clearly decrease, meaning that students tend to update their dominant tastes toward fields in which they perform well. For instance, 30% of the students have a dominant second-year taste in the field in which they obtained their best grade. The difference between both distributions (Kolmogorov-Smirnov test) is significant at the 1% level.

The far right-hand panel in Figure 1.2 qualifies the picture by showing that an opposite force – the strength of priors – can contradict ability signals. It plots the distribution of the strength of dominant high-school taste, as measured in Section 1.4.1, starting from Equation (1.3). It thus shows the share of keywords related to the student's dominant taste (in high school), among all the keywords that could be used to express the five academic tastes. The green bars show this distribution for students who maintain their initial taste, while the white bars shows the distribution for students who change their dominant taste. On average, students who maintain their taste used 50% of academic keywords related to their dominant taste, compared to 40% for students who changed taste. Both the difference in means (t-test) and the difference in distribution (Kolmogorov-Smirnov test) are significant at any conventional level.

Figure 1.2 shows that two opposite forces are at play. When the student does not obtain a good grade in her dominant high-school taste, she is conflicted and must decide whether she wants to rely on this bad-fit signal, or stick to her initial vocation. If students already had very strong priors, there is little change in their beliefs, even when they receive bad-fit signals. These students may be so passionate about a field that not having good grades is irrelevant, they already know what their true (and fixed) preferences are. This could be rationalized by the behavior of students who prefer to learn extensively about the field they love – reading overly specialized books, attending conferences – but overlook the importance of gaining the right set of methodological skills to perform well at exams.

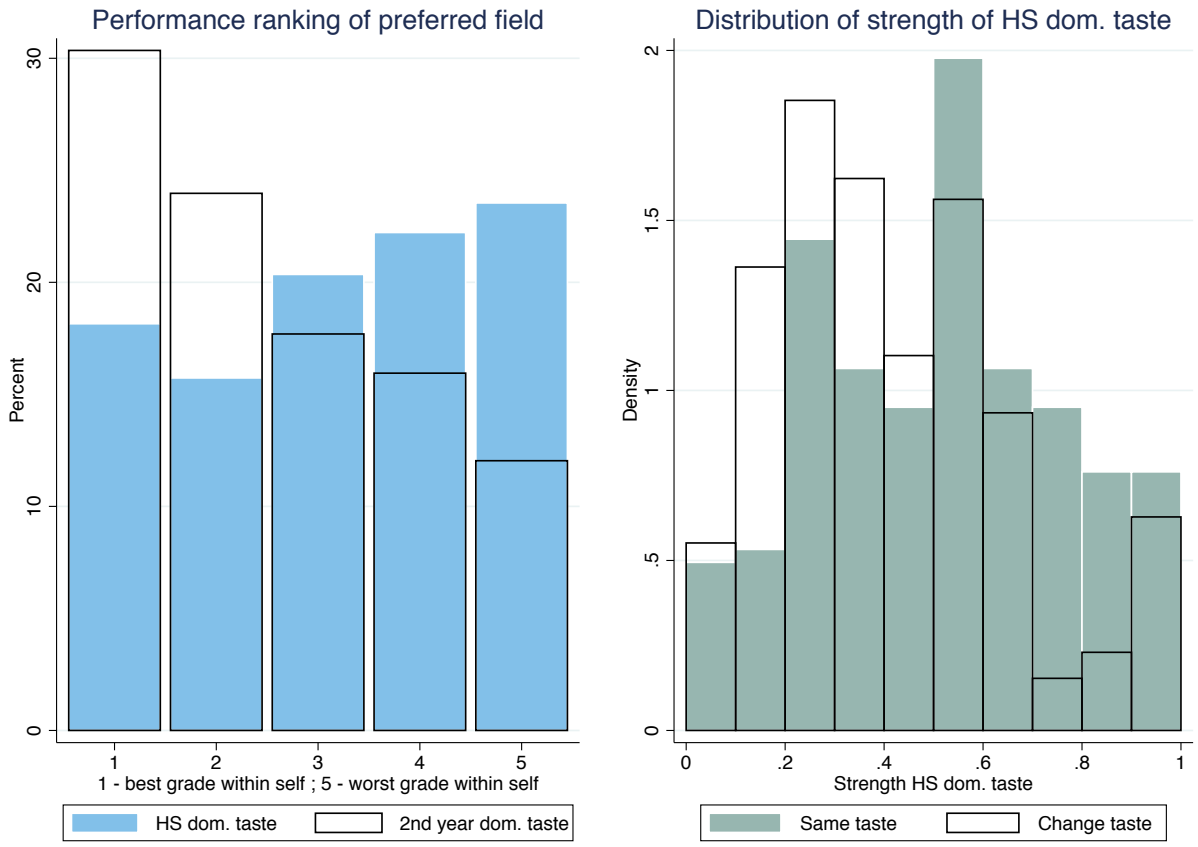


Figure 1.2 – Ability signal and strength of dominant high-school taste

Notes: The left-hand panel shows the ranking of the field that is the student's dominant high-school taste (blue bars) and dominant second-year taste (white bars). The bars are ordered from left to right from the best grade to the worst grade in the field of the dominant taste. The ranking is obtained by comparing the five grades obtained by each student at the end of the first year. These five grades are ordered from best (1) to worst (5), for each student. The right-hand panel shows the distribution of the strength of the dominant high-school taste, as measured in Section 1.4.1, starting from Equation (1.3). The green bars focus on the sample of students whose dominant taste changes between high school and the second year, and the white bars on those who maintain their initial taste. Students who do not have a clear dominant taste ("no taste" or "other taste") are not included.

Table 1.4 shows OLS regressions on how ability signals and the strength of the student's dominant taste in high school jointly drive taste updating. Columns 1 to 3 regress the binary variable for whether students' dominant taste changed between high school and their second year on the (standardized) course grade for the dominant high-school taste, a dummy variable for whether the student got her best grade in a field other than her favorite one, and the strength of the dominant high-school taste, in separate regressions. Columns 4 to 6 include these different variables jointly and progressively. Table 1.4 shows that ability signals are both drivers of stickiness when they are signals of good fit, and drivers of change when they are a signal that the student performs better in another field. Column 1 indicates that when a student obtains a grade that is one standard deviation higher in the field of her dominant high-school taste (about 1.8 points out of 20), she is on average 6.2 percentage points more likely to pursue her initial taste – priors are confirmed. On the contrary, having her best grade in another field is a strong driver of change in taste – leading to the updating of beliefs. Having one's best grade in another field increases the likelihood of a change by 21.6 percentage points (Column 2). Both types of ability signals are significant at the 1% level.

Columns 3 to 6 of Table 1.4 complement and qualify these results by showing that a strong

dominant high-school taste is also a significant driver of taste persistence. Going from a extremely weak dominant high-school taste (zero keywords cited) to an extremely strong dominant high-school taste (100% of academic keywords related to the favorite field), is associated with a 30 percentage-point decrease in the likelihood of changing taste, with this coefficient being significant at the 1% level (Column 3). When adding all three factors together in Column 6, we find that the coefficients of both ability signals decrease in magnitude, and sometimes in significance, which is expected given the collinearity of ability signals. The strength of dominant high-school taste remains a highly significant factor.

Table 1.4 – Determinants of change in taste: individual-level characteristics

	Dep. var. = 1 if change in dominant taste					
	(1)	(2)	(3)	(4)	(5)	(6)
Course grade (std) in HS dom. taste	-0.0621***			-0.0594***		-0.0326*
	(0.0158)			(0.0157)		(0.0167)
Best grade is not in HS dom. taste		0.216***			0.202***	0.168***
		(0.0443)			(0.0444)	(0.0480)
Strength HS dom. taste			-0.302***	-0.298***	-0.276***	-0.279***
			(0.0614)	(0.0613)	(0.0607)	(0.0610)
Constant	0.911	0.802	1.257**	1.055*	1.125**	1.013*
	(0.565)	(0.533)	(0.544)	(0.546)	(0.533)	(0.539)
Ind. Controls	Yes	Yes	Yes	Yes	Yes	Yes
RA FE	Yes	Yes	Yes	Yes	Yes	Yes
Campus FE	Yes	Yes	Yes	Yes	Yes	Yes
HS taste FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	906	917	913	902	913	902
R^2	0.108	0.119	0.113	0.130	0.138	0.145

Notes: OLS regressions where the dependent variable is binary and equal to 1 if the student's dominant taste changes between high school and the second year. Students who do not have a well-defined dominant taste in high school (no dominant taste or "other" dominant taste) are dropped. The explanatory variables are the standardized (at the field level) course grade in the student's dominant high-school taste, a dummy variable equal to 1 if the student's best (absolute) grade is in a field that is not her dominant high-school taste, and high-school taste intensity (measured by keyword shares). Individual controls include binary variables for gender, high socio-economic background, 3 baccalaureate dummies, 2 admission track dummies, 6 region of nationality dummies and the log of the number of words written in both letters of motivation. See Appendix table 3.E1 for a more detailed description of these variables. RA fixed effects are 3 dummy variables controlling for the differences in hand-labeling practices across our 4 research assistants. Campus FEs are 6 dummy variables controlling for the students' different campuses. HS taste FEs are 4 dummy variables for students' dominant high-school tastes. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

1.4.3 Heterogeneity in student characteristics

The fact that both (absolute) ability signals and high-school taste intensity matter implies the existence of heterogeneity. When faced with a trade-off, i.e. a situation where a student loves a field but receives a bad-fit signal, some students will follow the ability signal, and others will stick to their priors instead. Table 1.5 investigates the heterogeneity, in terms of gender and family background, associated with students' propensity to rely on ability signals or, on the contrary, maintain their initial vocation.

There is a large body of literature focusing on the differences in major choices according to gender and family background. It remains unclear whether such heterogeneity also extends to the importance students attach to the various factors leading to such choices. Astorne-Figari and Speer (2018) find that women are generally more likely to switch major, so we might expect the influence of the strength

of their high school taste to be weaker. Regarding ability signals, Astorne-Figari and Speer (2018) also show no evidence of gender-based sensitivity to poor performance in the decision to switch major or drop out. On the contrary, Ost (2010) find that female students are more sensitive to grades, but Elsner et al. (2018) find that male students pay more attention to their class rank than female students. Regarding family background, Hoxby and Turner (2015) and Guyon and Huillery (2019) find that low SES students tend to have less precise information about education and labor market outcomes. We would thus expect these students to be more reliant on the new informational signals they receive, especially ability signals. Guyon and Huillery (2019) further shows that low SES students tend to have low priors regarding their academic ability, so we might expect these students to be more performance-driven in the taste updating process.

Table 1.5 combines both ability-signal variables and the strength of dominant high-school taste with whether the student is female, and whether the student's family has a high socio-economic status. We do not find evidence of heterogeneity throughout these dimensions, even with respect to ability signals.

Although we do not find heterogeneity throughout these dimensions, there must be heterogeneity on other grounds. In Appendix 1.C we assess the relative importance of both forces in the taste updating trade-off, by defining a typology of students. Figure 1.C1 shows that performance-driven students (blue bars) dominate in relation to stubborn, "vocation-driven" students (orange bars), with the former encompassing 35% of the students, and the latter 17%.

1.4.4 No significant effect of relative ability signals

Table 1.6 shows the results of our analysis of the impact of relative ability signals on taste updating using the student's ranking within her local peer group (either tutorial group or campus) in her dominant high-school taste (columns 1 to 3). Columns 4 to 6 show the impact of being in the top 10% in the student's dominant high-school taste, on her campus. The first relative ability signal is noisy, in the sense that students only have access to incomplete information about their peers, while the second relative ability variable is known to students. Table 1.6 shows that relative ability correlates with changes in taste, but its coefficient becomes insignificant with the inclusion of the course grade in the dominant high-school taste. This result implies that students do not seem to use their relative standing as a relevant signal to update their tastes. They focus mostly on their own grades and especially on where they perform best across the five fields. This finding holds true both for noisy signals of relative ability (Column 2) and known relative ability (column 5). This result is also confirmed with a more conservative sample focusing on an arguably exogenous set of peers. Table 1.D5 shows the same regressions focusing on students whose tutorial group is known, with dual bachelor's degree students being dropped. The coefficient of relative ability also becomes smaller and non-significant when the course grade in the dominant high-school taste is added to the regression. Table 1.D6 further shows that this null result does not hide some heterogeneity according to gender or family background.

1.5 Conclusion

We elicit students' beliefs regarding their academic preferences – academic tastes – using the content of letters of motivation that students have to write at two points in time, and then document the taste-updating process that takes place while students experiment with courses, receive grades and

Table 1.5 – Determinants of change in taste: individual-level characteristics – heterogeneity by gender and socio-economic status

	Dep. var. = 1 if change in dominant taste					
	(1)	(2)	(3)	(4)	(5)	(6)
Course grade (std) in HS dom. taste	-0.0576**	-0.0624*				
	(0.0231)	(0.0323)				
Course grade (std) in HS dom. taste * female	-0.00787					
	(0.0307)					
Course grade (std) in HS dom. taste * High SES		0.000297				
		(0.0358)				
Best grade is not in HS dom. taste			0.230***	0.172**		
			(0.0643)	(0.0847)		
Best grade is not in HS dom. taste * female			-0.0245			
			(0.0842)			
Best grade is not in HS dom. taste * High SES				0.0590		
				(0.0979)		
Strength HS dom. taste					-0.408***	-0.257**
					(0.0939)	(0.128)
Strength HS dom. taste * female					0.188	
					(0.123)	
Strength HS dom. taste * High SES						-0.0593
						(0.146)
Female	0.0290	0.0287	0.0504	0.0303	-0.0510	0.0294
	(0.0307)	(0.0307)	(0.0782)	(0.0301)	(0.0563)	(0.0303)
High SES	0.0782*	0.0782*	0.0469	0.000501	0.0553	0.0805
	(0.0401)	(0.0407)	(0.0397)	(0.0896)	(0.0395)	(0.0699)
Constant	0.916	0.911	0.788	0.834	1.301**	1.239**
	(0.566)	(0.566)	(0.537)	(0.534)	(0.545)	(0.546)
Ind. Controls	Yes	Yes	Yes	Yes	Yes	Yes
RA FE	Yes	Yes	Yes	Yes	Yes	Yes
Campus FE	Yes	Yes	Yes	Yes	Yes	Yes
HS taste FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	906	906	917	917	913	913
R^2	0.108	0.108	0.119	0.120	0.115	0.113

Notes: OLS regressions where the dependent variable is binary and equal to 1 if the student's dominant taste changes between high school and the second year. Students who do not have a well-defined dominant taste in high school (no dominant taste or "other" dominant taste) are dropped. The explanatory variables are the standardized (at the field level) course grade in the student's dominant high-school taste, a dummy variable equal to 1 if the student's best (absolute) grade is in a field that is not her dominant high-school taste, and high-school taste intensity (measured by keyword shares). Individual controls include binary variables for gender, high socio-economic background, 3 baccalaureate dummies, 2 admission track dummies, 6 region of nationality dummies and the log of the number of words written in both letters of motivation. See Appendix table 3.E1 for a more detailed description of these variables. RA fixed effects are 3 dummy variables controlling for the differences in hand-labeling practices across our 4 research assistants. Campus FEs are 6 dummy variables controlling for the students' different campuses. HS taste FEs are 4 dummy variables for students' dominant high-school tastes. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.6 – Determinants of change in taste: performance position among peers

	Noisy relative ability signal			Known relative ability		
	(1)	(2)	(3)	(4)	(5)	(6)
Student's rank in HS dom. taste	-0.183***	-0.00923	-0.0544			
	(0.0530)	(0.108)	(0.107)			
Top 10% within campus in HS dom. taste				-0.121**	-0.0313	-0.0274
				(0.0472)	(0.0556)	(0.0543)
Course grade (std) in HS dom. taste		-0.0597*	-0.0178		-0.0559***	-0.0273
		(0.0321)	(0.0324)		(0.0186)	(0.0194)
Best grade is not in HS dom. taste			0.170***			0.167***
			(0.0481)			(0.0480)
Strength HS dom. taste			-0.279***			-0.279***
			(0.0609)			(0.0610)
Constant	1.034*	0.910	1.036*	0.810*	0.910	1.016*
	(0.548)	(0.556)	(0.542)	(0.463)	(0.553)	(0.539)
Ind. Controls	Yes	Yes	Yes	Yes	Yes	Yes
RA FE	Yes	Yes	Yes	Yes	Yes	Yes
Campus FE	Yes	Yes	Yes	Yes	Yes	Yes
HS taste FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	906	906	902	1074	906	902
R^2	0.105	0.108	0.145	0.118	0.108	0.145

Notes: OLS regressions where the dependent variable is binary and equal to 1 if the student's dominant taste changes between high school and the second year. Students who do not have a well-defined dominant taste in high school (no dominant taste or "other" dominant taste) are dropped. The explanatory variables are the student's rank within her tutorial group in the field of her dominant high-school taste (in columns 1 to 3 only), a dummy variable for whether the student is within the top 10% of her campus in the field of her dominant high-school taste, (columns 4 to 6), the standardized (at the field level) course grade in the student's dominant high-school taste, a dummy variable equal to 1 if the student's best (absolute) grade is in a field that is not her dominant high-school taste, and high-school taste intensity (measured by keyword shares). Individual controls include binary variables for gender, high socio-economic background, 3 baccalaureate dummies, 2 admission track dummies, 6 region of nationality dummies and the log of the number of words written in both letters of motivation. See Appendix Table 3.E1 for a more detailed description of these variables. RA fixed effects are 3 dummy variables controlling for the differences in hand-labeling practices across our 4 research assistants. Campus FEs are 6 dummy variables controlling for the students' different campuses. HS taste FEs are 4 dummy variables for students' dominant high-school tastes. Robust standard errors in parentheses. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

are matched with new peers. Our main findings are that the tastes of the vast majority of students change between high school and the second year of their bachelor studies. And second, the taste-updating is driven by two opposite forces: the strength of the students' dominant high-school taste and the performance signals revealing the quality of the match between students and their tastes. The second effect tends to dominate the first one, but only student-specific ability signals matter; relative ability signals do not seem relevant in our context. Weighting the relative importance of performance-driven vs stubborn students, we show that the former are twice as numerous as the latter. We do not find evidence of heterogeneity across gender or family socio-economic status in the determinants of taste updating.

The present research cannot determine which student profiles (stubborn or performance-driven students) lead to more or less favorable educational and labor market outcomes. Monitoring this cohort of students after they enter the labor market could ultimately enable us to determine whether updating tastes according to ability signals actually improves the welfare of students.

Our results nonetheless suggest that many students decide to switch fields when they are able to do so. The fact that 75% of students' dominant tastes change demonstrates that students' prior beliefs concerning their academic preferences are still largely undetermined in high school. This result indicates that a flexible higher education system, which gives students the opportunity to take courses in different fields before choosing a major, could help students make more informed choices.

Most universities in France run a system that is not flexible, with many students choosing their majors when they are still in high school. Our results suggest that a more flexible system could potentially reduce the large drop-out rate of students in the country's higher educational system. Indeed, a survey by the French Ministry of Higher Education shows that only 56% of students who graduated from high school in 2014 (same cohort as our students) passed their first year of undergraduate studies in the field they had originally chosen (MESRI-SIES, 2017).[21] The survey further shows that 38% of students who quit their initial track state that the reason for doing so was a mistake in their original choice. The students in this situation explain that they learned about the mistake they had made by receiving bad grades, obtaining new information about the field of study, or that they changed their mind about the job they wanted to enter after graduation. In contrast, in the context of our study, very few students dropped out despite the fact that many students updated their beliefs about their academic preferences. This result is most likely explained by the fact that students were able to switch fields very easily.

[21] The note summarizing the results of the survey can be found here: https://www.enseignementsup-recherche.gouv.fr/cid122940/les-bacheliers-2014-ou-en-sont-ils-a-la-rentree-2015.html

Bibliography

Altonji, Joseph G, Eric Blom, and Meghir Costas. Heterogeneity in Human Capital Investments: High School Curriculum, College Major, and Careers. *Annu. Rev. Econ.*, 4(1):185–223, 2012.

Altonji, Joseph G, Peter Arcidiacono, and Arnaud Maurel. The analysis of field choice in college and graduate school: Determinants and wage effects. *Handbook of the Economics of Education*, 5: 305–396, 2016.

Arcidiacono, Peter. Ability sorting and the returns to college major. *Journal of Econometrics*, 121: 343–375, 2004.

Astorne-Figari, Carmen and Jamin D Speer. Drop out , switch majors , or persist ? The contrasting gender gaps. *Economics Letters*, 164:82–85, 2018.

Beaman, Lori, Esther Duflo, Rohini Pande, and Petia Topalova. Female leadership raises aspirations and educational attainment for girls: A policy experiment in India. *Science*, 335(6068):582–586, 2012.

Beffy, Magali, Denis Fougere, and Arnaud Miguel. Choosing the Field of Study in Postsecondary Education: Do Expected Earnings Matter? *The Review Of Economics And Statistics*, 94(1):334–347, 2012.

Boring, Anne and Jen Brown. Gender, Competition and Choices in Higher Education. 2020.

Carlana, Michela, Eliana La Ferrara, and Paolo Pinotti. Goals and Gaps: Educational Careers of Immigrant Children *. 2017.

Chevalier, Arnaud, Steve Gibbons, Andy Thorpe, Martin Snell, and Sherria Hoskins. Students' academic self-perception. *Economics of Education Review*, 28:716–727, 2009.

Delavande, Adeline and Basit Zafar. University Choice: The Role of Expected Earnings, Nonpecuniary Outcomes, and Financial Constraints. *Journal of Political Economy*, 2019.

Elsner, Benjamin and Ingo E Isphording. A Big Fish in a Small Pond: Ability Rank and Human Capital Investment. *Journal of Labor Economics*, 35(3):787–828, 2017.

Elsner, Benjamin and Ingo E Isphording. Rank, Sex, Drugs, and Crime. *Journal of Human Resources*, 52(2):356–381, 2018.

Elsner, Benjamin, Ingo E Isphording, and Ulf Zölitz. Achievement Rank Affects Performance and Major Choices in College. *University of Zurich, Department of Economics, Working Paper*, 2018.

Filippin, Antonio and Andrea Ichino. Gender wage gap in expectations and realizations. *Labour Economics*, 12:125–145, 2005.

Gentzkow, Matthew and Jesse M Shapiro. What Drives Media Slant ? Evidence from U . S . Daily Newspapers. *Econometrica*, 78(1):35–71, 2010. doi: 10.3982/ECTA7195.

Gentzkow, Matthew, Bryan T. Kelly, and Matt Taddy. Text As Data. 2017.

Goux, Dominique, Marc Gurgand, and Eric Maurin. Adjusting Your Dreams? High School Plans and Dropout Behaviour. *Economic Journal*, 127(602):1025–1046, 2017.

Guyon, Nina and Elise Huillery. Biased aspirations and social inequality at school: evidence from French teenagers. 2019.

Hansen, Stephen, Michael McMahon, and Andrea Prat. Transparency and Deliberation Within the FOMC: a Computational Linguistics Approach. *Quarterly Journal of Economics*, (February):801–870, 2018.

Hoxby, Caroline M. and Sarah Turner. What High-Achieving Low-Income Students Know About College. *American Economic Review*, 105(5):514–517, 2015.

Landaud, Fanny, Son-thierry Ly, and Eric Maurin. Competitive Schools and the Gender Gap in the Choice of Field of Study. *Journal of Human Resources*, 0617-8864R(2018).

Malamud, Ofer. Discovering one's talent: Learning from academic specialization. *Industrial and Labor Relations Review*, 64(2):375–405, 2011.

MESRI-SIES, . Les bacheliers 2014, où en sont-ils à la rentrée 2015 ? *Note d'information du SIES*, Novembre, 2017.

Montmarquette, Claude, Kathy Cannings, and Sophie Mahseredjian. How do young people choose college majors ? *Economics of Education Review*, 21(2002):543–556, 2002.

Murphy, Richard and Felix Weinhardt. Top of the Class: The Importance of Ordinal Rank. *NBER Working Paper 24958*, 2018.

Ost, Ben. The role of peers and grades in determining major persistence in the sciences. *Economics of Education Review*, 29(6):923–934, 2010.

Rask, Kevin. Attrition in STEM fields at a liberal arts college : The importance of grades and pre-collegiate preferences. *Economics of Education Review*, 29(6):892–900, 2010.

Speer, Jamin D. The gender gap in college major : Revisiting the role of pre-college factors. *Labour Economics*, 44(August 2016):69–88, 2017.

Stinebrickner, Ralph and Todd Stinebrickner. Academic Performance and College Dropout : Using Longitudinal Expectations Data to Estimate a Learning Model. *Journal of Labor Economics*, 32 (3), 2014a.

Stinebrickner, Ralph and Todd R. Stinebrickner. A major in science? Initial beliefs and final outcomes for college major and dropout. *Review of Economic Studies*, 81(1):426–472, 2014b.

Wiswall, Matthew and Basit Zafar. Determinants of college major choice: Identification using an information experiment. *Review of Economic Studies*, 82(2):791–824, 2015.

Zafar, Basit. How do College Students Form Expectations? 2011.

Zafar, Basit. College Major Choice and the Gender Gap. *Journal of Human Resources*, 48(3):545–595, 2013.

Appendices

1.A Letters of motivation

1.A.1 Guidelines for letters of motivation

There are three admission procedures for Sciences Po: a standard procedure for French students, an international procedure for students studying abroad in high school, and an affirmative-action program. All students have to write a letter of motivation in their applications. Only students who are in their final year of high school are allowed to apply to the university's undergraduate program. The guidelines for letters of motivation at the admission level depend on the admission procedure. We set out these guidelines below:

- **French procedure**[22]: *This space is provided to enable us to better understand your motivations for applying, not only to the Undergraduate College but also to the specific programmes you have chosen. In your letter, you also have the opportunity to explain how and why Sciences Po's programmes (Bachelor or Master) will help you to realise your personal and professional goals. In addition, do not hesitate to let us know how you intend to participate in campus life and to provide any other information that you consider relevant to your application. Do not exceed 1000 words.*

- **International procedure:** *The motivation letter is an opportunity for you to freely express yourself and to provide any information you consider relevant. You should introduce yourself, provide your reasons for applying to your specific programmes, explain how and why Sciences Po's programmes will help you realize your personal and professional goals, and tell us how you intend to contribute to campus life. Be as specific as you can and do not exceed 1000 words.*

- **Conventions Education Prioritaire:** For the affirmative-action procedure, applicants receive more guidance about the structure of the letter and must answer four questions. We use only the first three questions as the last one is not about the applicant.

  1. Is there any information you wish to bring to the attention of Sciences Po?
  2. Why have you decided to apply to Sciences Po?

[22] This is the official Sciences Po translation of the following French guidelines:

– **French procedure**: *Cet espace est destiné à nous permettre de mieux comprendre vos motivations, non seulement pour le Collège universitaire, mais aussi pour les programmes auxquels vous vous portez candidat. Au travers de votre lettre, vous avez également la possibilité de nous expliquer comment et pourquoi les programmes de Sciences Po (Bachelor et même Master) vous aideront à réaliser vos projets personnels et professionnels. Enfin, n'hésitez pas à développer la manière dont vous entendez participer à la vie de campus et à nous fournir des éléments que vous jugerez pertinents pour votre candidature. Veillez à ne pas dépasser les 1000 mots.*

– **International procedure:** *La lettre de motivation est un espace où vous pouvez vous exprimer librement et nous fournir toute information qui vous paraît pertinente. Ainsi, vous pouvez vous présenter, nous faire part des raisons qui vous ont poussé à postuler au(x) programme(s) de votre choix. Vous pouvez également expliquer comment et pourquoi les programmes de Sciences Po vous aideront à réaliser vos projets professionnels et personnels. Enfin, vous pouvez aussi nous dire comment vous entendez participer à la vie universitaire. Veillez à être précis et à ne pas dépasser les 1000 mots.*

3. Have you already given any thought to the career path that interests you or to the field in which you wish to work in the future?
4. Is there a future for politics?

The letter of motivation for the third year abroad is written during the first semester of the second year and must be submitted in December. The letter must contain a justified ranking of six universities at which students would like to enrol for the next year. The administration reviews these letters alongside the students' transcripts, resumes and official language tests. No official guidelines have been published for the letter of motivation for the third year abroad, but students must attend several information meetings during their second year in preparation for it. They are encouraged to first decide upon the geographical area of their choice and choose their six universities within this area. For instance, if they target North America, all their choices should be either in the US or Canada. They must then rank these six universities, providing as much detail as possible on what they would do in each one of them, the courses they want to take, the professors they would like to meet etc. The letter must not exceed two pages. See **?** for a description of the allocation process of students to universities.

1.A.2 Hand-labeling

We asked four research assistants to hand-label the data. They had to read the letters and fill in an Excel spreadsheet detailing the characteristics of these letters. They had to determine the top four academic fields mentioned by the students. They had no restriction on the items they could enter, although a pre-existing list with the most common fields and all Sciences Po master's degrees were supplied to them. They had the opportunity to enter additional items if they felt the need to do so, and they did this in practice. Overall, across all letters, RAs identified 73 unique fields, from which we identified the five dominant tastes that match Sciences Po's first year curriculum as shown in Table 1.A1.

Table 1.A1 – Hand-labeled field categories

Final categories	**Hand-labeled fields**
Economics/Finance/Business	Economics, Finance, Business
History	history
Law	law
Political Science and international relations	political science, politics, institutions, international relations, geography, geopolitics, african geopolitics, development, international development, energy studies, environment, sustainable development, security, war studies
Sociology	sociology
Other	accounting, african studies, anthropology, archeology, architecture, art, asian studies, biology, chemistry, cinema, civilization, cognitive sciences, communication, computer sciences, criminology, cultural studies, culture, digital studies, education, egyptology, entrepreneurship, ethics, ethnology, european affairs, fashion, gender studies, hard sciences, humanitarian studies, health, human resources, jewish studies, journalism, language, literature, management, marketing, mathematics, media studies, middle eastern studies, oceanography, philosophy, physics, planning, public policies, public affairs[23], psychology, religion, social sciences, social work, statistics, theater, urbanism, writing

A narrow definition of dominant taste would only consider the first-ranked field variable, but this would come at the cost of dropping many observations, given the wealth of tastes described by students. We use a more flexible definition: if the first-ranked taste is not one of the five fields of interest (e.g. anthropology), we move down and use the second-ranked taste. We repeat this procedure down to the fourth-ranked taste (the last hand-coded variable for preferred fields). If none of the five fields is ever mentioned by the student, but the student nonetheless has another identified taste (for psychology for example), then we code the dominant taste as an "other taste". If the student does not mention any academic fields – which can happen if the students focuses exclusively on extra-curricular achievements for instance – we code the dominant taste as "no dominant taste". In the admission-related letters, 8.86% of students have another dominant taste and 6% have "no dominant taste". A flexible approach enables the reduction of the "other taste category" that would otherwise be too large (34% in high school and 59% in the second year[24]), if we were considering only first-ranked tastes.

Tables 1.A2 and 1.A3 show how far we need to move down in rank in order to identify one of the five fields. For high-school letters (Table 1.A2), we see that in more than 90% of the cases, simply moving down to the second-ranked taste is sufficient to find one of the five fields. For second-year letters (Table 1.A3), moving to the second-ranked taste is sufficient in 77% of the cases.

[24]Students have a wider variety of tastes at that stage of their studies. Moreover, studying abroad allows students to take courses in fields that are not available at Sciences Po.

Table 1.A2 – Dominant taste and hand-labeled rank in high-school letters

	Hand-labeled ranking (1 = most dominant taste)				
Dominant taste in high school	**1**	**2**	**3**	**4**	**Total**
Economics/Finance/Business	138	42	9	1	190
(row %)	72.6%	22.1%	4.7%	0.5%	100.0%
History	298	93	37	6	434
(row %)	68.7%	21.4%	8.5%	1.4%	100.0%
Law	68	16	5	0	89
(row %)	76.4%	18.0%	5.6%	0.0%	100.0%
Poli Sci/Inter. Rel.	116	45	13	2	176
(row %)	65.9%	25.6%	7.4%	1.1%	100.0%
Sociology	22	7	2	0	31
(row %)	71.0%	22.6%	6.5%	0.0%	100.0%
Total	642	203	66	9	920
(row %)	69.8%	22.1%	7.2%	1.0%	100.0%

Notes: The table shows at which stages the dominant tastes of the fields in the rows are identified. Column (1) indicates the frequencies and share of students for whom the dominant taste is identified as the top field mentioned in the high-school motivation letter. Column (2) indicates the frequencies and share of students for whom the dominant taste is identified as the second-to-top field cited in the letter, and so on. For instance, 72.6% of the students whose dominant taste is for Economics/Finance/Business in high school have their taste ranked first by our RAs in the letters of motivation (this represents 138 students out of a total of 190 students whose dominant taste is for Economics/Finance/Business)

Table 1.A3 – Dominant taste and hand-labeled rank in second-year letters

	Hand-labeled ranking (1 = most dominant taste)				
Dominant taste in 2nd year	**1**	**2**	**3**	**4**	**Total**
Economics/Finance/Business	109	83	59	24	275
(row %)	39.6%	30.2%	21.5%	8.7%	100.0%
History	70	41	14	3	128
(row %)	54.7%	32.0%	10.9%	2.3%	100.0%
Law	89	27	9	3	128
(row %)	69.5%	21.1%	7.0%	2.3%	100.0%
Poli Sci/Inter. Rel.	123	153	76	15	367
(row %)	33.5%	41.7%	20.7%	4.1%	100.0%
Sociology	47	24	15	3	89
(row %)	52.8%	27.0%	16.9%	3.4%	100.0%
Total	438	328	173	48	987
(row %)	44.4%	33.2%	17.5%	4.9%	100.0%

Notes: The table shows at which stages the dominant tastes of the fields in rows are identified. Column (1) indicates the frequencies and share of students for which the dominant taste is identified as the top field mentioned in the second-year letter of motivation. Column (2) indicates the frequencies and share of students for which the dominant taste is identified as the second-to-top field mentioned in the letter, and so on. For instance, 39.6% of the students whose dominant taste is for Economics/Finance/Business in high school have their taste ranked first by our RAs in the letters of motivation (this represents 109 students out of a total of 275 students whose dominant taste is for Economics/Finance/Business).

1.A.3 Keywords

Chosen keywords

To measure the strength of students' priors, i.e. the intensity of their dominant high-school taste, we searched for a number of keywords that can be associated with the five fields. We first pre-process the letters by putting all words in lowercase, removing accents, punctuation, and common stopwords.[35] Letters could be written either in French or in English, so we used both French and English keywords rather than translating the letters. In several cases, the French and English words are the same (such as "finance" and "business"). In some cases, we use the root word that is common to both English and French if it is not ambiguous to other meanings. We count the number of times keywords related to a particular field are mentioned in the letter. If the same keyword is mentioned several times, we count all of its occurrences. Table 1.A4 shows which keywords we chose and how they were associated with each academic taste. These keywords are intended to cover both the fields and the master's degrees at Sciences Po related to each topic. Note that we do not tokenize (separate in lists of words and n-grams) the letters before searching for the terms. This implies that our keywords can be embedded in longer words. For instance, the keyword "econom" would also pick "microeconomics". But this also

[35] We choose not to stem words, i.e. reducing all the same versions of a word to a common base, because the risk of false positives was too large. A word like "droit", which means "law" in French, can be the root word to many other words with a completely different meaning. The same applies to "finance".

Table 1.A4 – Keywords

Academic taste	Keywords	Master's degrees
Economics/Finance/Business	econom, finance[25], business	**At the Doctoral School:** master in economics. **At the school of Management, Organization and Finance:** master in economics and business, economics and public policy, finance and strategy, financial regulation and risk management
History	histoire, history	**At the Doctoral School:** master in history
Law	droit[26], law, judiciaire, juridique, judicial, legal[27]	**At the Doctoral School:** master in law[28]. **At the Law School:** master carrières judiciaires et juridiques[29], master in economic law[30]
Political Science and international relations	political science, science politique, sciences politique[31], affaire international, affaires international[32], psia, international affair, relation international, relations international, international relation, international security, international public management, international economic policy, environmental policy, environmental sciences policy, international development, human rights humanitarian action, international energy, development practice[33], journalism and international affairs	**At the Doctoral School**: master in political science. **At the Paris School of International Affairs (PSIA)** : master in international security, international public management, international economic policy, environmental policy, environmental sciences and policy, international development, human rights and humanitarian action, international energy, development practice, journalism and international affairs
Sociology	sociolog[34]	**At the Doctoral School:** master in sociology

means that long keywords (several words in a row) need to appear with the exact same sequence in the letter. So the keyword "political science" will not count the words "political" or "science" occurring separately, but the plural will be picked up (political sciences). In addition to keywords related to the field names, we include keywords related to the master's degrees delivered by Sciences Po. This is why the taste for political science and international relations has more keywords than the others; the Paris School of International Affairs at Sciences Po simply has more master's degrees available than the other schools, as the last column of Table 1.A4 shows (see Table 1.D2 for the distribution of students at Sciences Po Schools and in its master's degrees). However, we are somewhat conservative since we only count full master's degree names. For instance, only the exact term "human right and humanitarian action" would count, and not simply "human right". Note that we only use the keywords listed in the Keywords columns, not the ones in the master's degree column.

Table 1.A5 – Average number of keywords per letter

Variable	Obs	Mean	Std. Dev.	Min	Max
High school letters					
Economics/Finance/Business	1078	2.3	2.5	0	20
History	1078	1.8	1.8	0	14
Law	1078	.8	1.5	0	14
Political Science and International Relations	1078	1.3	1.5	0	9
Sociology	1078	.3	.6	0	6
Second year letters					
Economics/Finance/Business	1080	7.3	11.2	0	68
History	1080	2.9	4.8	0	56
Law	1080	3.3	8.5	0	68
Political Science and International Relations	1080	4.6	5.5	0	31
Sociology	1080	1	2.7	0	36

Table 1.A5 shows the average number of keywords per letter for each academic taste, listed separately for admission letters (written in high school), and second-year letters.[36] The average number of keywords increases substantially between both time periods. This is consistent with the fact that second-year letters are typically longer (828 words for high-school letters and 1,383 words for second-year letters, on average), and students are encouraged to mention courses they would like to attend during their academic visit, which tends to increase word counts significantly.

Are keywords accurate predictors of hand-labeled dominant tastes?

Using keywords as a measure of dominant taste intensity makes the assumption that keywords accurately reflect the hand-labeled coded dominant taste, and provide information about the intensive margin. In this section we show further evidence for the link between keywords and hand-labeled data by asking whether keywords would perform well at the extensive margin, i.e. are keywords accurate predictors of hand-labeled dominant tastes? We provide cross-tabulations summarizing the average number of keywords in high-school letters (Table 1.A6) and second-year letter (Table 1.A7), according to hand-labeled dominant tastes. We clearly see that the number of keywords is larger on the diagonal, which means that (hand-labeled) dominant tastes attract more keywords than other tastes.

[36] We do not use data on second-year letter keywords in this paper, but we include these statistics here for interested readers.

Table 1.A6 shows that the diagonal effect persists even when comparing high-school letter keywords with dominant second-year tastes.

Table 1.A6 – Average number of keywords in high-school letters per (hand-labeled) dominant taste

	Mean keyword count in high school letter				
	Econ	History	Law	Poli Sci	Sociology
Dominant taste in high school					
Economics/Finance/Business	4.1	1.3	0.7	1.2	0.4
History	2.0	2.8	0.7	1.3	0.4
Law	1.9	1.1	3.2	1.1	0.4
Poli Sci/Inter. Rel.	2.0	1.4	0.4	1.9	0.2
Sociology	2.7	1.7	0.7	1.3	1.5
Other taste	1.2	0.8	0.1	0.7	0.2
No expressed taste	1.1	0.5	0.3	0.8	0.0
Total	2.3	1.8	0.8	1.3	0.3
Dominant taste in 2nd year					
Economics/Finance/Business	3.1	1.5	0.6	1.2	0.3
History	2.2	3.0	0.4	1.3	0.4
Law	2.1	1.7	1.7	1.3	0.3
Poli Sci/Inter. Rel.	1.9	1.7	0.7	1.5	0.3
Sociology	1.7	1.6	0.7	0.9	0.5
Other taste	1.8	1.9	0.8	1.3	0.3
No expressed taste	2.5	1.2	0.8	1.4	0.2
Total	2.3	1.8	0.8	1.3	0.3

Notes: We read the table as follows: on average, students who have a dominant taste for Economics/Finance/Business in **high school** mention 4.1 keywords related to Economics/Finance/Business, 1.3 related to History, 0.7 related to law, 1.2 related to Political Science and International Relations and 0.4 related to Sociology, **in their high-school motivation letter**. Students who have a dominant taste for Economics/Finance/Business in their**second year** cited, on average, 3.1 keywords related to Economics/Finance/Business, 1.5 related to History, 0.6 related to Law, 1.2 related to Political Science and International Relations and 0.3 related to Sociology, **in their high-school motivation letter**.

Table 1.A7 – Average number of keywords in second-year letters per (hand-labeled) dominant taste

	Mean keyword count in 2nd year letter				
	Econ	History	Law	Poli Sci	Sociology
Dominant taste in high school					
Economics/Finance/Business	10.5	2.1	3.2	3.8	0.8
History	6.8	3.8	3.4	4.6	1.1
Law	5.8	2.2	8.8	4.7	1.2
Poli Sci/Inter. Rel.	7.0	2.6	1.5	5.6	1.1
Sociology	4.7	1.9	2.2	4.6	1.8
Other taste	6.4	1.7	2.5	3.2	0.9
No expressed taste	5.9	2.4	2.1	5.4	1.2
Total	7.3	2.9	3.3	4.5	1.0
Dominant taste in 2nd year					
Economics/Finance/Business	21.3	1.1	0.9	1.9	0.4
History	1.6	10.6	1.0	4.0	0.8
Law	3.8	1.5	18.9	4.8	0.8
Poli Sci/Inter. Rel.	2.8	2.4	1.7	8.0	0.6
Sociology	1.9	2.3	0.8	3.3	6.2
Other taste	0.8	1.6	0.6	0.6	0.6
No expressed taste	1.0	0.8	1.2	4.2	0.0
Total	7.3	2.9	3.3	4.6	1.0

Notes: We read the table as follows: on average, students who have a dominant taste for Economics/Finance/Business in **high school** mention 4.1 keywords related to Economics/Finance/Business, 1.3 related to History, 0.7 related to law, 1.2 related to Political Science and International Relations and 0.4 related to Sociology, **in their high-school motivation letter**. Students who have a dominant taste for Economics/Finance/Business in their**second year** cited, on average, 3.1 keywords related to Economics/Finance/Business, 1.5 related to History, 0.6 related to Law, 1.2 related to Political Science and International Relations and 0.3 related to Sociology, **in their high-school motivation letter**.

Another measure of the predictive power of keywords is to recreate a keyword-based measure of dominant taste (extensive margin measure). We define a keyword-based dominant taste as the field (among the five selected ones) that has the highest number of keywords. Keywords act like a "classifier", as used in machine-learning terminology, and we can use three measures to assess the performance of keyword-labeled tastes relative to hand-labeled ones. This procedure assumes that hand-labeled tastes reflect the "true" tastes. (**?**).

- Accuracy: proportion of correct predictions
- Sensitivity: proportion of correct predictions among true positives
- Specificity: proportion of correct predictions among true negatives

Accuracy is an intuitively appealing measure but it becomes less informative when outcomes are not evenly distributed. Consider the binary outcome "having a dominant taste for sociology". Only 2.86% of the students have this dominant taste in high school. The accuracy of a purely random guess should be 50%. However, a better guess (although still uninformed), would be to predict that students never have a dominant taste for sociology. In that case, the accuracy would rise to 97.14%. So, when positives or negatives are rare, accuracy should be accompanied by sensitivity and specificity rates.

In the previous example, the sensitivity rate of such an extreme guess would be 0% (none of the true positives were correctly identified), and the specificity rate would be 100%.

Figure 1.A1, 1.A2 and 1.A3 show the results for the accuracy, sensitivity, and specificity rates, respectively. Regarding accuracy, the grey bar in Figure 1.A1 shows the accuracy rate of keywords in predicting the student's exact dominant taste. Since there are seven dominant taste categories (the five fields and the "other" and "no expressed taste" categories), a random guess would give a 14% accuracy rate. Keywords perform almost three times better for high-school letters (41% accuracy) and 4.5 times better for second-year letters (65% accuracy). The rise in accuracy is consistent with the fact that second-year letters are typically longer, more detailed, and more academically focused than admission-level letters. Regarding binary outcomes, i.e. whether or not the student has a dominant taste in each field, accuracy is also high compared to a 50% random guess.

As expected, Figures 1.A2 and 1.A3 show that keywords perform much better at predicting negatives (high specificity rate), but sometimes struggle to predict positives, especially when they are rare. For sociology, for instance, one third and one half of positives are correctly identified as such by keywords, in high-school letters and third-year letters, respectively.

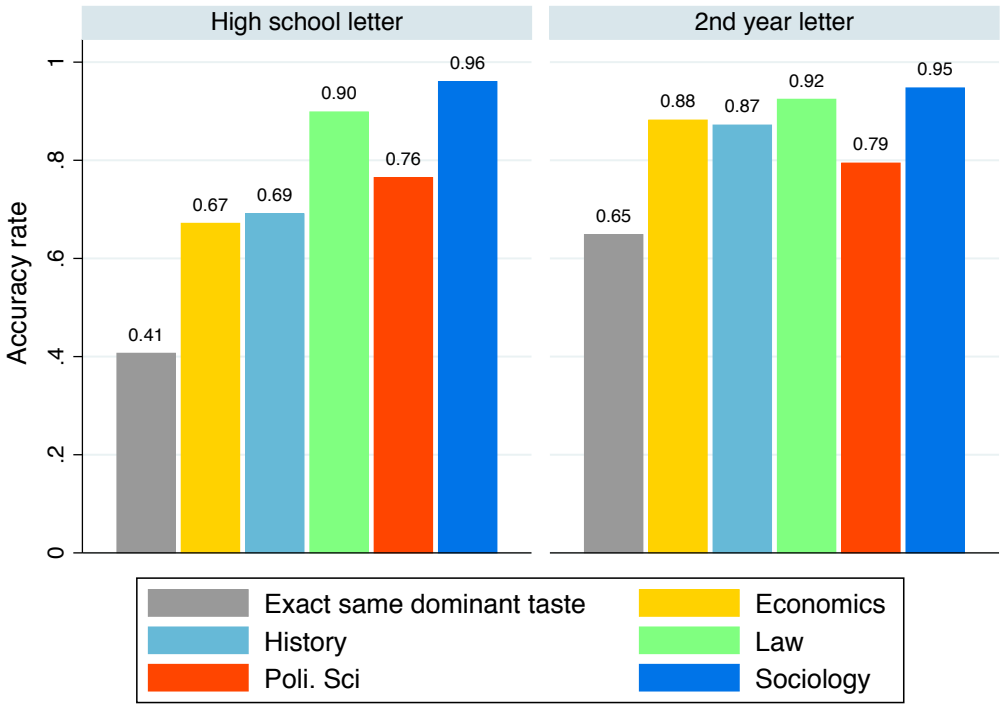


Figure 1.A1 – Proportion of match between keyword-based and hand-labeled dominant taste (**accuracy rate**)

Notes: The gray bars on the far left report the accuracy rate of keywords in predicting the student's exact dominant taste. The colored bars display the accuracy rate of keywords in predicting whether or not the student has a (hand-labeled) dominant taste in each of the five fields. We read that, for 67% of the students, keywords and hand-labeled methods give the same value (either 0 or 1) for having a dominant taste for economics/finance/business in high school (yellow bar). The left-hand panel shows the accuracy rate using high-school-letter data only, and the right-hand panel uses second-year letters.

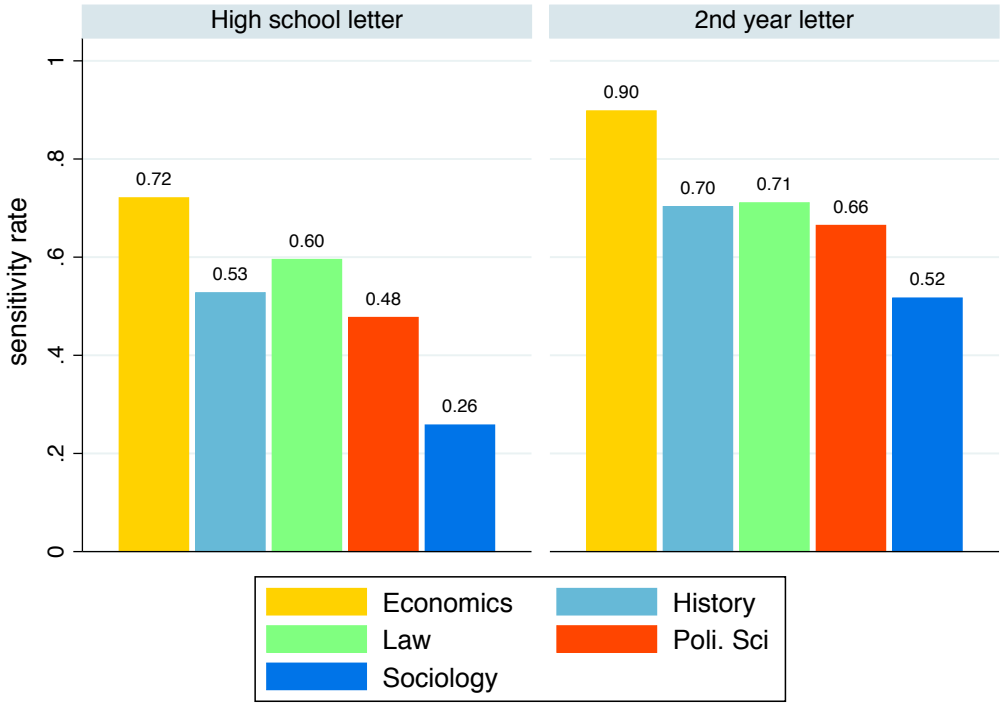


Figure 1.A2 – Proportion of correctly identified (hand-labeled) positives using keywords (**sensitivity rate**)

Notes: The graph reports the sensitivity rate, i.e. the proportion of actual positives (hand-labeled dominant taste) that were correctly identified as such by keyword-based dominant taste. For instance, there are 434 students that have a dominant taste for history in high school, according to hand-labeled data. These are the "actual" or "true" positives. Keywords correctly identify 229 of these students as having a dominant taste for history. This implies a sensitivity rate of 53% for history. The left-hand panel shows the accuracy rate using high-school-letter data only and the right-hand panel uses second-year letters.

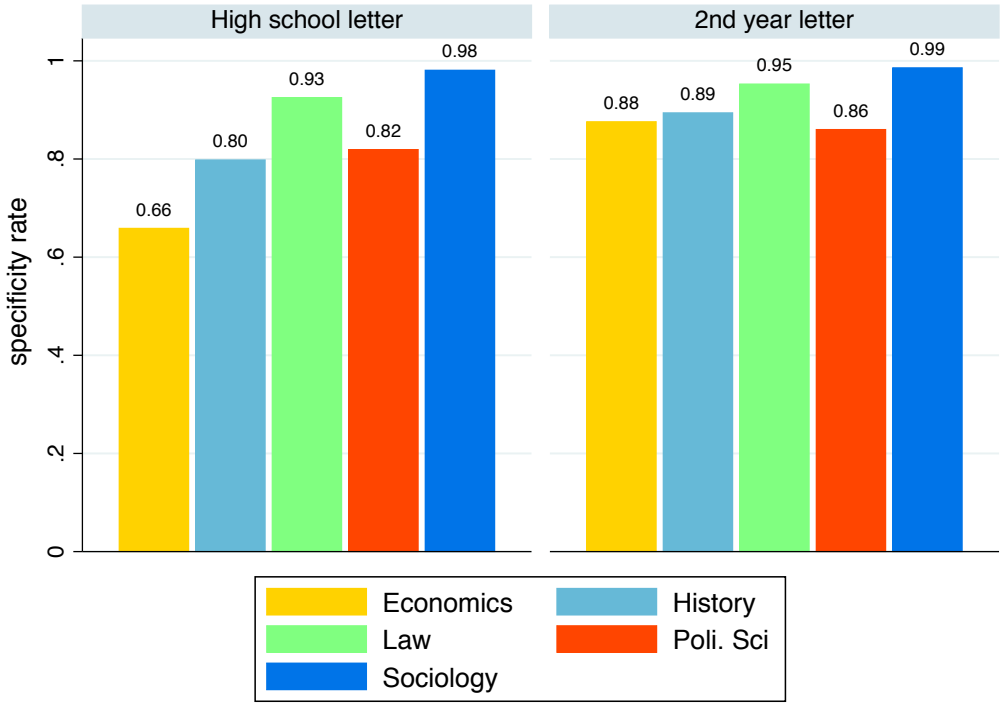


Figure 1.A3 – Proportion of correctly identified (hand-labeled) negatives using keywords (**specificity rate**)

Notes: The graph reports the specificity rate, i.e. the proportion of actual negatives (hand-labeled data indicates that the field is *not* a dominant taste) that were correctly identified as such by keyword data. For instance, there are 989 students who *do not* have a dominant taste for law in high school, according to hand-labeled data. This are the "actual" or "true" negatives. Keywords correctly identify 915 of these students as *not* having a dominant taste for law. This implies a specificity rate of 93% for law. The left-hand panel shows the accuracy rate using high-school-letter data only and the right-hand panel uses second-year letters.

1.B Peer groups

We collected data on the tutorial groups, which are formed in August and last for the entire academic year. They only exist on the Paris and Nancy campuses and are composed of approximately 15 students, (minimum 10 and maximum 20 students), as shown in the left-hand part of Figure 1.B1. There are 41 such tutorial groups in Paris and 6 in Nancy. Nine of them are composed of students enrolled in dual degrees with other Paris universities, which enable them to earn a bachelor's degree in mathematics, physics or philosophy, on top of the Sciences Po bachelor's degree. Students who applied and were admitted to these dual degrees are likely to be self-selected on the basis of their tastes and grades, so we cannot assume that peers are random in this case. These students constitute 13% of our sample according to Table 3.E1. The remaining students are quasi-randomly allocated to tutorial groups and spend the tutorial hours of the six main courses with their classmates (we describe this allocation process in the next paragraph). This implies that they spend six tutorial hours together per week. For the other campuses, tutorial groups are formed in a decentralized way and no records are associated with this allocation process. In that case, "tutorial group peers" are considered to be the other students of the same cohort who study on the same campus. There are 65 students in Dijon, 66 in Le Havre, Menton and Poitiers, and 140 in Reims, who belong to the 2014 cohort. The allocation of classmates is less likely to be random for those latter campuses since students can choose a particular campus based on their tastes. However, the relevant tastes for the choice of campus are mostly culture and language-based since campuses specialize in geographical areas. For instance, Nancy specializes in German culture and Reims in North-American culture. There is no specialization based on the five tastes we focus on (economics, history, law political science and sociology); therefore, choosing a campus is very different from being enrolled in a dual degree and we do not expect sorting to occur on the basis of these five tastes across the campuses. Nevertheless, we show results focusing on Paris and Nancy campuses in the form of robustness checks.

On the Paris and Nancy campuses, students are allocated to tutorial groups in the following manner. Students choose their group number during the registration period. A few days in advance, Sciences Po provides the complete list of tutorial groups alongside the name of the three teachers for the first semester as well as the corresponding timetable. No information is provided about the teachers and timetable for the second semester. Students are then invited to log on to the registration website and select the tutorial group they would like to be enrolled in. There is no add-and-drop period and when the website opens, all groups are filled within a minute. Usually, the tutorial groups listed first are the first to be filled. However, students may develop two kinds of strategies that could bias our estimates. First, they can coordinate with friends to choose the same tutorial group. Since friends are more likely to have similar tastes and ability, this may form clusters of students sharing similar preferences. Several items of evidence suggest that this channel is unlikely to be a large source of bias. First, the tutorial group registration occurs *before* the integration week and the start of the academic year, so no sorting is possible on the basis of integration group friendships. However, sorting on the basis of high-school acquaintances may still be possible, since every year, several high schools manage to have more than one student admitted.

A second source of bias in the allocation of students to tutorial groups is that students may acquire information about teachers prior to registration. This means that they could select a tutorial group with a better teacher in their favorite field. Anecdotal evidence suggests that this kind of strategy

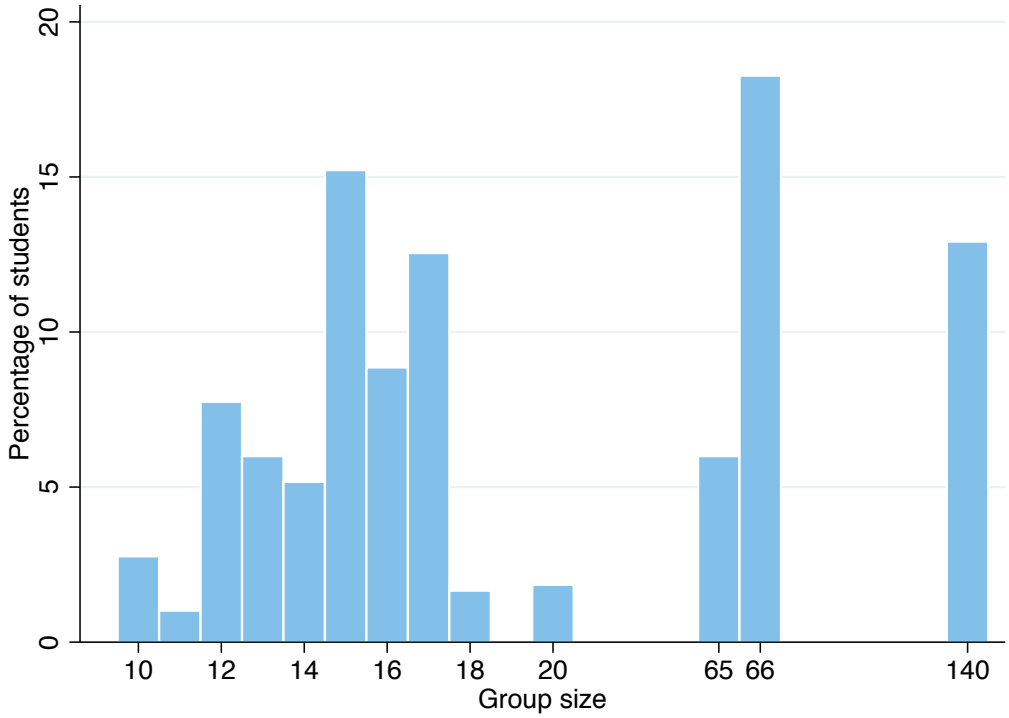


Figure 1.B1 – Distribution of local peer group sizes

Notes: Distribution of tutorial group sizes in Paris and Nancy campuses (up to 20 students) on the left-hand side of the graph. The right-hand side shows the sizes of the other campuses on which there are no fixed tutorial groups. There are 65 students in Dijon, 66 in Le Havre, Menton and Poitiers, and 140 in Reims.

is widely used. Students share information on Facebook groups and have access to a website called *Le Confoscope* for teacher ratings.[37] This is also likely to cluster students with similar preferences together, and additionally, it could assign the best teachers to the students who like their field the most.

Despite these concerns, our data shows that such strategies did not create taste clusters, and that students are not systematically matched on the basis of their observed characteristics. The general strategy we use is to follow **?** and check for selection on observables. More precisely, we run the following regressions:

$$Y_i = a + bPeer's\ HS\ taste_i + u_i \tag{1.4}$$

Where the dependent variable Y_i is a student characteristic. We use the student's dominant high-school taste, as we want to test for the tendency of students with similar academic preferences to cluster together into the same tutorial groups. We also use variables for gender, family background, type of baccalaureate, whether the student is French, and the length of each letter of motivation. Unfortunately, we don't have data on students' grades in high school, prior to joining Sciences Po. The explanatory variable is the share of classmates having a dominant high-school taste for a field j. The main coefficient of interest b should be close to zero if students do not select their study groups on the basis of their peers' tastes. We run regressions for each field separately.

Table 1.B1 shows how the student's dominant tastes and those of her local peers are correlated.

[37] This website is currently unavailable: http://www.lapeniche.net/confoscope/

The odd-numbered column focuses on the entire sample except for students enrolled in dual degrees. The even-numbered column additionally restricts the sample to the Paris and Nancy campuses. We do not find significant positive correlations except for dominant taste for Political Science/International Relations. However, we consider that this correlation is obtained by chance since students had no means of coordinating vis-à-vis the Political Science teachers. Indeed, this course is taught during the second semester, so students could not yet know which Political Science teachers would be assigned to each tutorial group.

Table 1.B1 – Correlation between the students' and their tutorial group peers' dominant high-school tastes

	Econ/Finance/Business		History		Law		Poli Sci/Inter. Rel.		Sociology	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Peers' taste in Econ	-0.140	-0.267*								
	(0.140)	(0.159)								
Peers' taste in History			0.172	0.140						
			(0.139)	(0.149)						
Peers' taste in Law					-0.0650	-0.130				
					(0.165)	(0.179)				
Peers' taste in Poli. Sci.							0.326**	0.203		
							(0.133)	(0.161)		
Peers' taste in Sociology									0.0729	-0.0513
									(0.124)	(0.132)
Observations	936	533	936	533	936	533	936	533	936	533
R^2	0.001	0.004	0.002	0.002	0.000	0.001	0.007	0.004	0.000	0.000

Notes: OLS regressions of a binary variable for whether the student has a dominant taste for field j and the share of tutorial group peers having the same dominant taste. In both cases, dominant tastes are measured in high school, prior to any contact between the student and her peers. The odd-numbered column focuses on the entire sample except for students enrolled in dual degrees. The even-numbered column additionally restricts the sample to the Paris and Nancy campuses, on which tutorial groups are known.

We next turn to the other students' characteristics, which we use as our dependent variable Y_i. Figures 1.B2 and 1.B3 plot the coefficients b for each dependent variable, for each of the 5 academic tastes. Figure 1.B2 drops students enrolled in dual degrees and 1.B3 additionally retains students on the Paris and Nancy campuses only. Although some selection on observables seems to take place in Figure 1.B2, once we focus on the Paris and Nancy campuses, the coefficients are closer to zero and mostly non-significant.

Figure 1.B4 shows the variations in dominant-taste distributions across local peer groups. Both figures show large naturally occurring variations in the distribution of tastes, which are permitted thanks to the relatively small sizes of the peer groups.

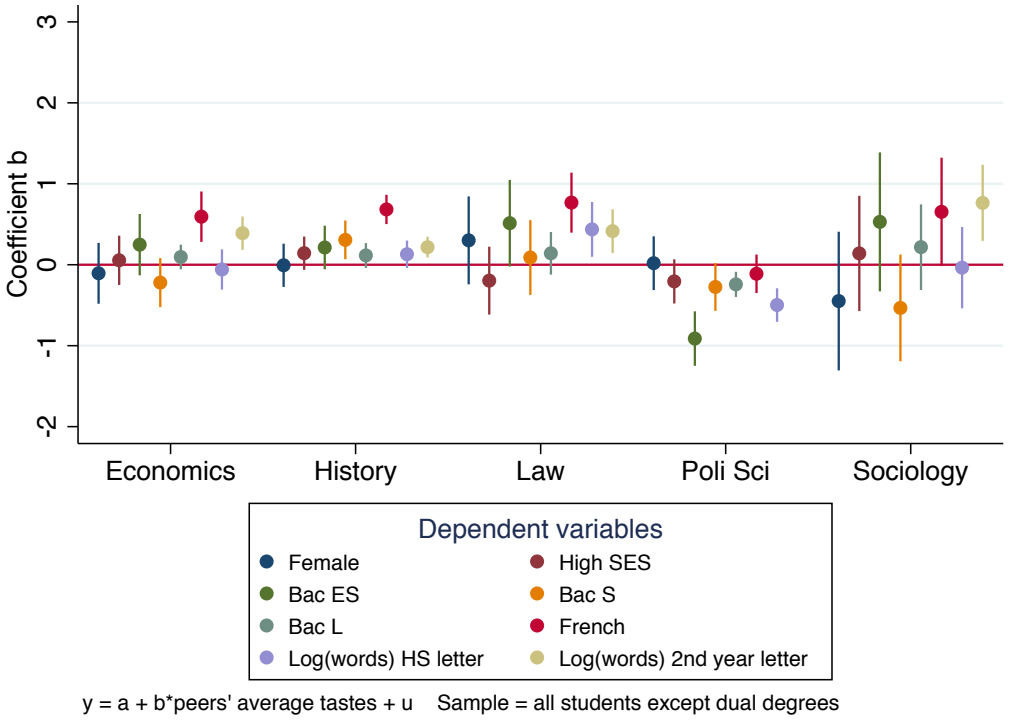


Figure 1.B2 – Randomization tests for tutorial groups

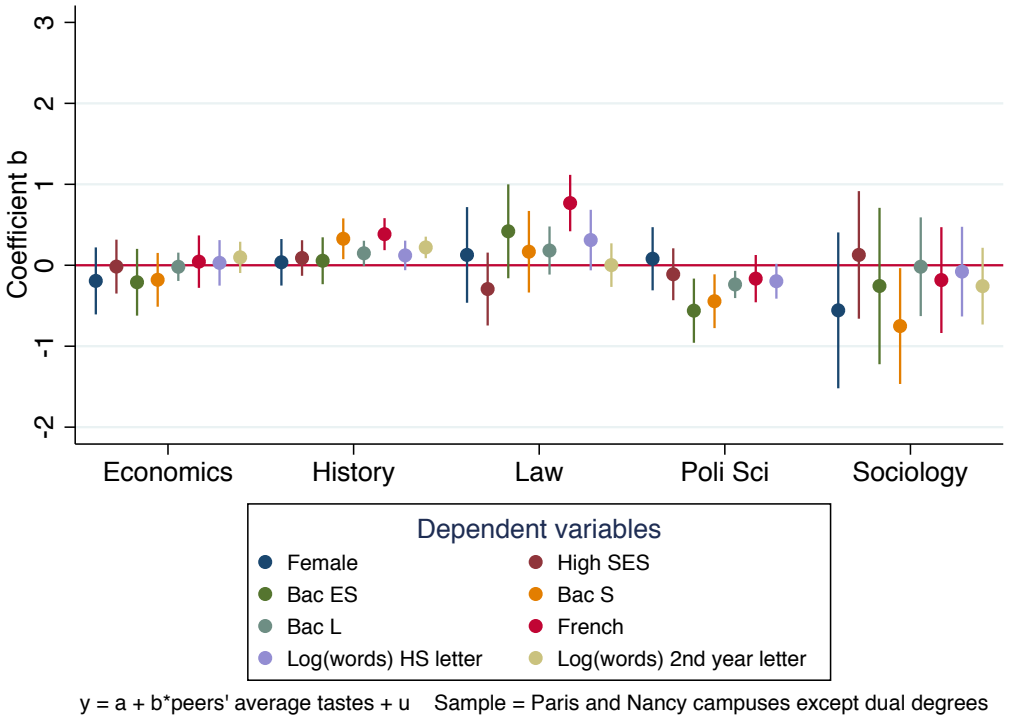


Figure 1.B3 – Randomization tests for tutorial groups – Paris and Nancy campuses

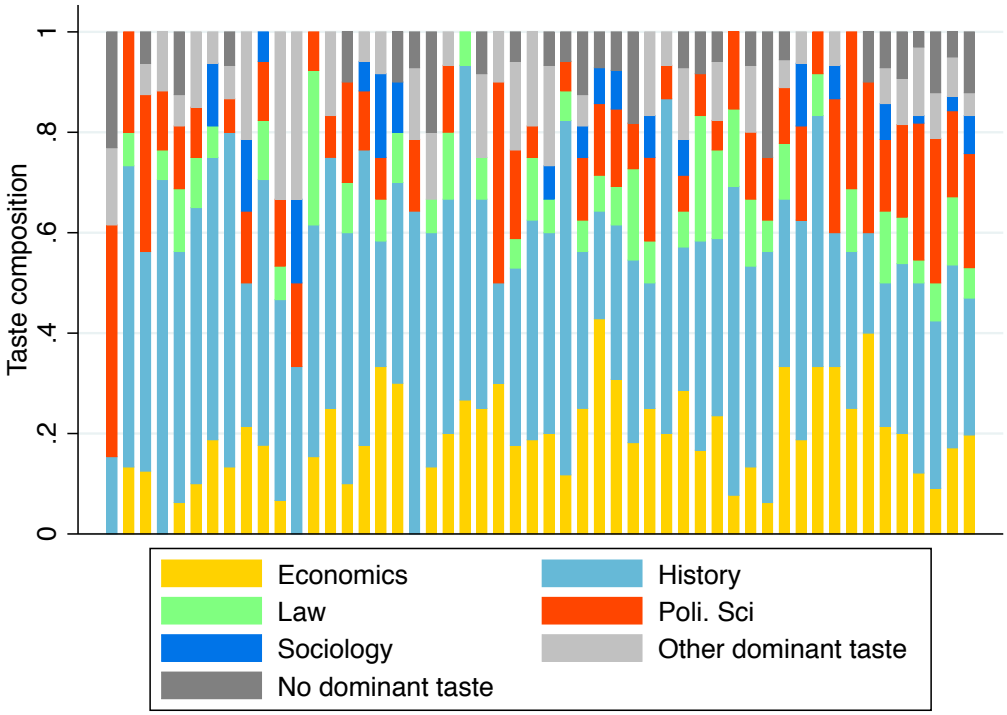


Figure 1.B4 – Composition of dominant high-school tastes in peer groups

Notes: Each bar correspond to one peer group (either a tutorial group for Paris and Nancy or the entire campus). It shows the distribution of dominant high-school tastes of the students composing each group.

1.C Typology of student heterogeneity

We define a typology of students to determine the relative importance of ability-driven vs stubborn students. We define four types. Students can be: (i) performance-driven, (ii) consistent, (iii) stubborn or (iv) inconsistent, with respect to the evolution of their dominant tastes and their grades. Within each type, we define two subcategories to allow for intensive margin differences. Table 1.C1 shows the definition of these categories. We show the distribution of these types in our sample in Figure 1.C1. Blue bars correspond to performance-driven students. Strongly performance-driven students change taste for the field in which they obtain their best grade. Mildly performance-driven individuals change taste for the field in which they obtain their second-best grade.[38] In other words, mildly performance-driven students only partially follow ability signals, while strongly performance-driven students strictly follow the signals. The "strongly consistent" students are represented in green: for these students, there is no trade-off between the ability signals and their high school taste, since they obtain their best grade in their favorite field and stick to this initial taste. Then, in orange, we show the stubborn students: they maintain their initial taste despite obtaining better grades elsewhere. "Mildly consistent/Mildly stubborn" students maintain their high school taste, which is also where they achieve their second-best performance. "Strongly stubborn" students maintain their high school taste despite it being their third, fourth or even fifth-ranked performance. The dark grey bar represents "inconsistent" students, who change taste despite obtaining their best grade in their dominant high-school taste (dark grey bar). Finally, the light grey bar stands for a residual category, which consists of the rest of the students, i.e. students who did not have a clear dominant taste in high school ("other" or "no taste" categories), or students taking other factors into account when updating their taste instead of ability signals and their dominant high-school taste. Examples of such other motives include students who are inspired by a particularly enthusiastic professor, students who follow a friend's academic tastes, or who are motivated by updated expectations or beliefs regarding labor market outcomes (sectoral differences in average starting salaries or employment rates, etc.). Figure 1.C1 shows the distribution of such types for the whole sample. The figure shows that performance-driven students (blue bars) dominate in relation to stubborn, "vocation-driven" students (orange bars), with the former encompassing 35% of the students, and the latter 17%.

[38] We do not include students who performed best in their dominant high-school taste, since those students are in the "inconsistent" category.

Table 1.C1 – Typology – category definitions

Category	**Change taste**	**Degree of adaptation to ability signal**
Strongly performance-driven	Yes	Best performance field = new dominant taste
Mildly performance-driven	Yes	Second-best performance field = new dominant taste
Strongly consistent	No	Best performance field = initial dominant taste
Mildly consistent/ Mildly stubborn	No	Second-best performance field = initial dominant taste
Strongly stubborn	No	Third, fourth or worst performance field = initial dominant taste
Inconsistent	Yes	Best performance field = initial dominant taste
Other motives		Rest of the students

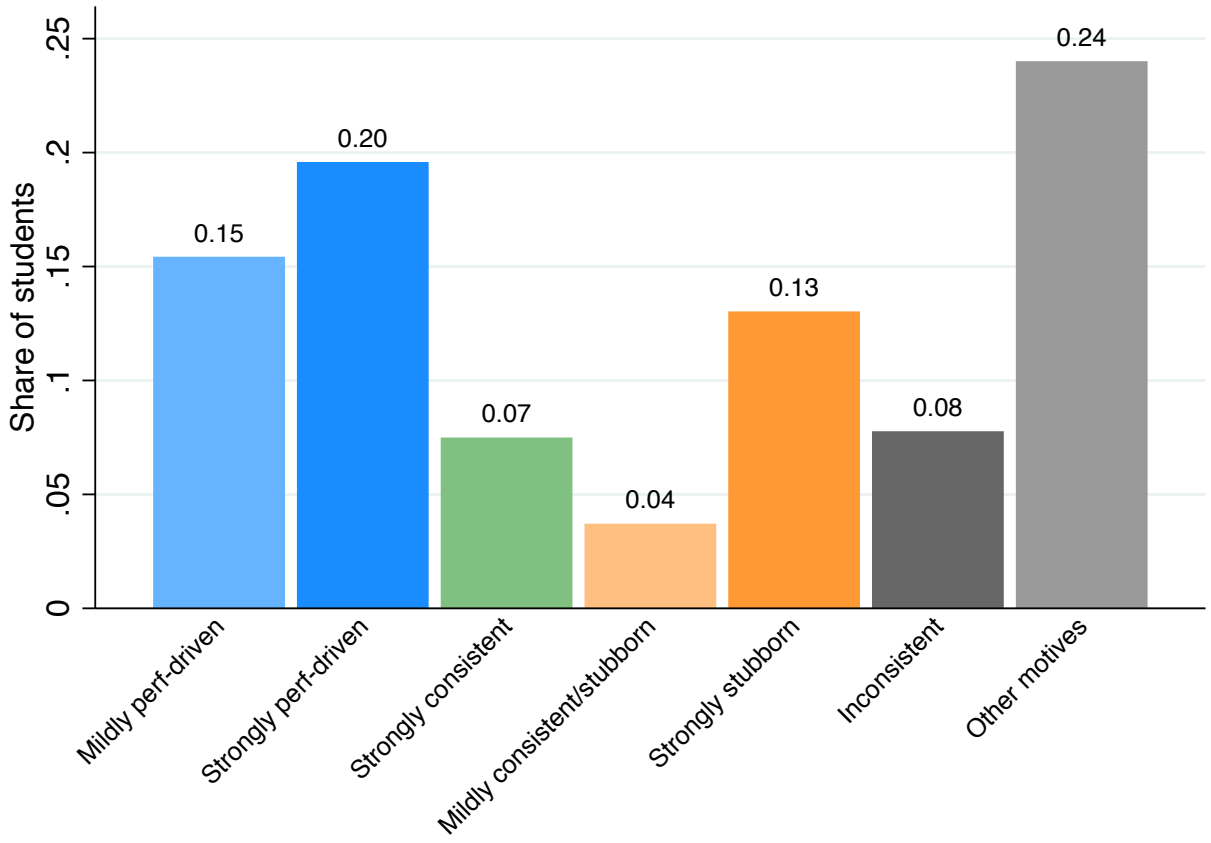


Figure 1.C1 – Typology of students

Notes: The bars display the distribution of students across taste updating types. Mildly performance-driven students change taste for the field in which they obtained their second-best grade and in which they performed worse than in their dominant high-school taste. Strongly performance-driven individuals change taste for the field in which they obtained their best grade. Strongly consistent students do not experience any trade-off between the ability signals and their dominant high-school taste since they obtained their best grade in their favorite field and stuck to this initial taste. Mildly consistent/Mildly stubborn students maintain their high school taste, in which they also achieved their second-best performance. Strongly stubborn students maintain their high school taste despite it being their third, fourth or even fifth-ranked performance. Inconsistent students change taste despite their best grade being in their dominant high-school taste. Finally, the light grey bar represents the rest of the students who do not fall into any category.

1.D Appendix Tables and Figures

Table 1.D1 – Students characteristics

Variable	Obs	Mean	Std. Dev.	Min	Max
Female	1084	.58	.494	0	1
High SES	1084	.789	.408	0	1
Bac ES	1084	.438	.496	0	1
Bac S	1084	.283	.451	0	1
Bac L	1084	.078	.269	0	1
Other high school diploma	1084	.2	.4	0	1
Domestic Admission Procedure	1084	.591	.492	0	1
CEP Admission Procedure	1084	.084	.277	0	1
International Admission Procedure	1084	.325	.468	0	1
Dual BA degree	1084	.13	.34	0	1
French Nationality	1084	.785	.411	0	1
Region nationality = Africa	1084	.033	.179	0	1
Region nationality = Asia or Oceania	1084	.025	.156	0	1
Region nationality = Europe except France	1084	.129	.336	0	1
Region nationality = Latin America	1084	.017	.128	0	1
Region nationality = Middle East	1084	.004	.061	0	1
Region nationality = North America	1084	.007	.086	0	1
Length (in words) High school letter	1078	829	201	112	2103
Length (in words) 2nd year letter	1080	1383	354	88	2851
Econ grade	1079	13.89	2.11	3.2	18.8
History grade	1082	12.95	1.86	6	18.2
Law grade	1069	13.42	1.81	7.4	18.6
Political Science grade	1070	13.54	1.59	0	18
Sociology grade	1068	14.26	1.87	0	19.3
Change taste between high school and 2nd year	1084	0.75	0.44	0	1

*Notes:*High SES is equal to 1 if at least one of the parents has (or had, if retired) a high socio-economic status profession, i.e. private- or public-sector managers, company directors, engineers, higher education professors and scientific professions, independent professions (lawyers, doctors etc.), retired managers and intermediary professions. Self-declared information from the admission files. Variations in the availability of grade data comes from the fact that some students received a letter instead of a numeric grade. Students that were declared to be "failing" (because they failed to attend a sufficient number of tutorials), are assigned a grade of 0. When students did not pass the course (grade lower than 10), they had to retake the exam. In that case, we focused on the first grade they received only.

Table 1.D2 – Sciences Po Schools and master's degrees

	N	Col %
School of Public Affairs		
Affaires européennes	9	2.9%
Corporate & Public Management	21	6.8%
European Affairs	18	5.8%
Politiques publiques	184	59.5%
Public Policy	77	24.9%
Total	309	100.0%
Paris School of International Affairs		
Environmental Policy	22	10.7%
Human Rights and Humanitarian action	17	8.3%
International Development	28	13.7%
International Economic Policy	16	7.8%
International Energy	9	4.4%
International Pub. Management	39	19.0%
International Security	72	35.1%
Sc. et pol. de l'environnement	2	1.0%
Total	205	100.0%
Journalism school		
Journalisme	10	100.0%
Total	10	100.0%
Communication school		
Comm. Media and Crea. Indus	4	7.1%
Comm. Médias et Indus. Créa	37	66.1%
Marketing et Etudes	15	26.8%
Total	56	100.0%
School of Management, Organization and Finance		
Economics and Business	18	15.5%
Finance et stratégie	88	75.9%
Organisations & Management RH	10	8.6%
Total	116	100.0%
Law school		
Carrières jud et juridiques	31	23.7%
Droit économique	100	76.3%
Total	131	100.0%
Urban school		
Governing the large metropolis	23	31.9%
Stratégies territ. & urbaines	49	68.1%
Total	72	100.0%
Doctoral school		
Economie	10	17.5%
Histoire	9	15.8%
Science Politique	26	45.6%
Sociologie	12	21.1%
Total	57	100.0%
TOTAL	956	

Table 1.D3 – Choice of specialist courses in the second year and dominant high-school tastes

	(1) Econ/Finance/Business	(2) History	(3) Law	(4) Poli Sci/Inter. Rel.	(5) Sociolgy
Econ/Finance dominant taste in HS	0.134***				
	(0.0355)				
History dominant taste in HS		0.0822***			
		(0.0254)			
Law dominant taste in HS			0.197***		
			(0.0513)		
Poli Sci/ Inter. Rel. dominant taste in HS				0.0105	
				(0.0347)	
Sociology dominant taste in HS					-0.0242
					(0.0798)
Ind. Controls	Yes	Yes	Yes	Yes	Yes
RA FE	Yes	Yes	Yes	Yes	Yes
Campus FE	Yes	Yes	Yes	Yes	Yes
Observations	1074	1074	1074	1074	1074
Pseudo R^2					

Notes: OLS regressions where the dependent variable is binary, and equal to 1 if the student chooses a specialist course in the second year in Economics or Finance (columns 1 and 2), Law (columns 3 and 4) or Political Science/International Relations (columns 5 and 6). The explanatory variables are binary variables for whether the student has a dominant taste in the corresponding field in high school (HS) or in the second year. Note that it is possible for students to choose several specialist courses, not necessarily in the same field. The explanatory variables are binary variables for whether the student has a dominant taste in the corresponding field in high school (HS) or in the second year. Individual controls include binary variables for gender, high socio-economic background, 3 baccalaureate dummies, 2 admission track dummies, 6 region of nationality dummies and the log of the number of words written in both letters of motivation. See Appendix table 3.E1 for a more detailed description of these variables. RA fixed effects are 3 dummy variables controlling for the differences in hand-labeling practices across our 4 research assistants. Campus FEs are 6 dummy variables controlling for the students' different campuses (Dijon, le Havre, Menton, Nancy, Paris, Poitiers and Reims). Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.D4 – Choice of elective courses in 2nd year and dominant high-school tastes

	(1) Econ/Finance/Business	(2) History	(3) Law	(4) Poli Sci/Inter. Rel.	(5) Sociolgy
Econ/Finance dominant taste in HS	0.112***				
	(0.0380)				
History dominant taste in HS		0.0765***			
		(0.0294)			
Law dominant taste in HS			0.0724		
			(0.0512)		
Poli Sci/ Inter. Rel. dominant taste in HS				0.0404	
				(0.0358)	
Sociology dominant taste in HS					0.00442
					(0.0644)
Constant	0.268	0.869*	0.913**	0.740*	0.283
	(0.475)	(0.495)	(0.411)	(0.442)	(0.367)
Ind. Controls	Yes	Yes	Yes	Yes	Yes
RA FE	Yes	Yes	Yes	Yes	Yes
Campus FE	Yes	Yes	Yes	Yes	Yes
Observations	1074	1074	1074	1074	1074
R^2	0.112	0.097	0.075	0.271	0.136

Notes: OLS regressions where the dependent variable is binary, and equal to 1 if the student chooses a specialist course in the second year in Economics or Finance (columns 1 and 2), Law (columns 3 and 4) or Political Science/International Relations (columns 5 and 6). The explanatory variables are binary variables for whether the student has a dominant taste in the corresponding field in high school (HS) or in the second year. Note that it is possible for students to choose several specialist courses, not necessarily in the same field. The explanatory variables are binary variables for whether the student has a dominant taste in the corresponding field in high school (HS) or in the second year. Individual controls include binary variables for gender, high socio-economic background, 3 baccalaureate dummies, 2 admission track dummies, 6 region of nationality dummies and the log of the number of words written in both letters of motivation. See Appendix table 3.E1 for a more detailed description of these variables. RA fixed effects are 3 dummy variables controlling for the differences in hand-labeling practices across our 4 research assistants. Campus FEs are 6 dummy variables controlling for the students' different campuses (Dijon, le Havre, Menton, Nancy, Paris, Poitiers and Reims). Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.D5 – Determinants of change in taste: performance position among peers – students with tutorial groups only and students in dual degrees are also dropped

	Noisy relative ability signal			Known relative ability		
	(1)	(2)	(3)	(4)	(5)	(6)
Student's rank in HS dom. taste	-0.200***	-0.0721	-0.103			
	(0.0718)	(0.139)	(0.138)			
Top 10% within campus in HS dom. taste				-0.133**	-0.0428	-0.0486
				(0.0613)	(0.0727)	(0.0718)
Course grade (std) in HS dom. taste		-0.0460	-0.0258		-0.0563**	-0.0441*
		(0.0413)	(0.0421)		(0.0253)	(0.0264)
Best grade is not in HS dom. taste			0.0620			0.0561
			(0.0673)			(0.0672)
Strength HS dom. taste			-0.283***			-0.283***
			(0.0836)			(0.0841)
Constant	1.310*	1.204	1.398*	1.139*	1.167	1.342*
	(0.762)	(0.770)	(0.760)	(0.609)	(0.759)	(0.747)
Ind. Controls	Yes	Yes	Yes	Yes	Yes	Yes
RA FE	Yes	Yes	Yes	Yes	Yes	Yes
Campus FE	Yes	Yes	Yes	Yes	Yes	Yes
HS taste FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	456	456	455	532	456	455
R^2	0.160	0.162	0.187	0.175	0.162	0.187

Notes: OLS regressions where the dependent variable is binary, and equal to 1 if the student's dominant taste changes between high school and the second year. Students who do not have a well-defined dominant taste in high school (no dominant taste or "other" dominant taste) are dropped. The explanatory variables are the student's rank within her tutorial group in the field of her dominant high-school taste (in columns 1 to 3 only), a dummy variable for whether the student is within the top 10% of her campus in the field of her dominant high-school taste, (columns 4 to 6), the standardized (at the field level) course grade in the student's dominant high-school taste, a dummy variable equal to 1 if the student's best (absolute) grade is in a field that is not her dominant high-school taste, and high-school taste intensity (measured by keyword shares). Individual controls include binary variables for gender, high socio-economic background, 3 baccalaureate dummies, 2 admission track dummies, 6 region of nationality dummies and the log of the number of words written in both letters of motivation. See Appendix table 3.E1 for a more detailed description of these variables. RA fixed effects are 3 dummy variables controlling for the differences in hand-labeling practices across our 4 research assistants. Campus FEs are 6 dummy variables controlling for the students' different campuses. HS taste FEs are 4 dummy variables for students' dominant high-school tastes. Robust standard errors in parentheses. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.D6 – Determinants of change in taste: performance position within peers – heterogeneity by gender and socio-economic status

	Noisy relative ability signal		Known relative ability	
	(1)	(2)	(3)	(4)
Student's rank in HS dom. taste	-0.147*	-0.138		
	(0.0791)	(0.123)		
Student's rank in HS dom. taste * female	-0.0640			
	(0.105)			
Student's rank in HS dom. taste * High SES		-0.0555		
		(0.135)		
Top 10% within campus in HS dom. taste			-0.124*	0.000194
			(0.0662)	(0.108)
Top 10% within campus in HS dom. taste * female			0.00481	
			(0.0927)	
Top 10% within campus in HS dom. taste * High SES				-0.144
				(0.119)
Female	0.0630	0.0287	0.0286	0.0282
	(0.0627)	(0.0309)	(0.0282)	(0.0270)
High SES	0.0764*	0.103	0.0512	0.0635*
	(0.0402)	(0.0735)	(0.0331)	(0.0342)
Constant	1.034*	1.009*	0.810*	0.794*
	(0.547)	(0.552)	(0.464)	(0.463)
Ind. Controls	Yes	Yes	Yes	Yes
RA FE	Yes	Yes	Yes	Yes
Campus FE	Yes	Yes	Yes	Yes
HS taste FE	Yes	Yes	Yes	Yes
Observations	906	906	1074	1074
R^2	0.105	0.105	0.118	0.119

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

Notes: OLS regressions where the dependent variable is binary and equal to 1 if the student's dominant taste changes between high school and the second year. Students who do not have a well-defined dominant taste in high school (no dominant taste or "other" dominant taste) are dropped. The explanatory variables are the student's rank within her tutorial group in the field of her dominant high-school taste (in columns 1 to 3 only), a dummy variable for whether the student is within the top 10% of her campus in the field of her dominant high-school taste, (columns 4 to 6), the standardized (at the field level) course grade in the student's dominant high-school taste, a dummy variable equal to 1 if the student's best (absolute) grade is in a field that is not her dominant high-school taste, and high-school taste intensity (measured by keyword shares). Individual controls include binary variables for gender, high socio-economic background, 3 baccalaureate dummies, 2 admission track dummies, 6 region of nationality dummies and the log of the number of words written in both letters of motivation. See Appendix Table 3.E1 for a more detailed description of these variables. RA fixed effects are 3 dummy variables controlling for the differences in hand-labeling practices across our 4 research assistants. Campus FEs are 6 dummy variables controlling for the different campuses students can live in. HS taste FEs are 4 dummy variables for students' dominant high-school tastes. Robust standard errors in parentheses. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Chapter 2

Preferences over Income Distribution: Evidence from a Choice Experiment

This chapter is a joint work with Max Lobeck, Claudia Senik and Thierry Verdier.

2.1 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., 2017). 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 paper 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 (2004) 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.[1] 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.

2.2 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.

2.2.1 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.[2]

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 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.[3] 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 2.2.3). All subjects also go through a comprehension test to show them examples of the choices they will have to make.[4]

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.

[1] 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).

[2] 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.*

[3] 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.

[4] 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 2.C5).

The choices we drop in the second part are Choice 4 in Table 2.A6 and the sanity check Choice 8.[5] 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.

2.2.2 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. Note that, throughout this paper, 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.[6]

Table 2.1 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.[7] 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 2.2 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

[5] See the end of Section 2.2.2 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 inventivized 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 2.B.

[6] 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.

[7] The screenshot of a choice, as presented in the laboratory, can be seen in Figure 2.C1.

Table 2.1 – Constant Efficiency Choices

	Choice 1		**Choice 2**		**Choice 3**		**Choice 4**	
	A	**B**	**A**	**B**	**A**	**B**	**A**	**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
Information below not shown to the subjects								
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

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.

Table 2.2 – 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
Information below not shown to the subjects						
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 2.C5 in the Appendix shows screenshots of the training choice and the questionnaire.

2.2.3 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.[8] 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 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. (2011) (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 2.C and is displayed in Figure 2.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.

2.2.4 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

[8] Before performing the real task, subjects carried out a 15-second training task.

variance (riskiness) of the lotteries are jointly increasing.[9] Second, we let subjects choose between two different lotteries, depicted in Table 2.A6 (Appendix 2.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.[10]

Table 2.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.

2.2.5 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, 2004), 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.[11] 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.

[9] The sixth lottery is an exception. Here, the variance is increasing compared to the fifth one, but not its expected value.

[10] 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 2.D for more details.

[11] This show-up fee was necessary to comply with the lab rules.

2.3 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.

2.3.1 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 2.1 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).

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 2.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.

Moving to the right panel of Figure 2.1, 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 2.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 2.3) 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

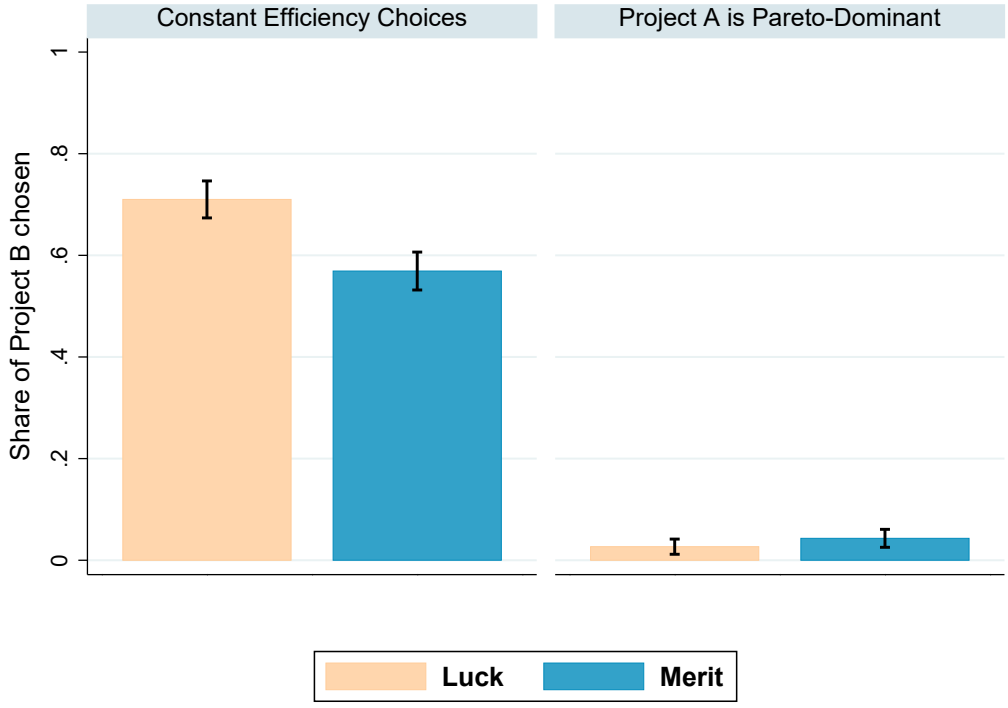


Figure 2.1 – Choices Behind the Veil of Ignorance by Treatment

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.

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).

2.3.2 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 2.2 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.[12]

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

[12]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 2.3 – Drivers of Inequality Aversion (Choice of Project B) - Logit Regressions on Pooled Data

	Veil of ignorance			Known position		
	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Var. = 1 if Project B (low inequality) is chosen						
Merit	-0.0732***	0.0718	0.0721	-0.00601	0.0260	0.0303
	(0.0227)	(0.0646)	(0.0594)	(0.0202)	(0.0339)	(0.0335)
Constant Efficiency	0.566***	0.648***	0.634***	0.173***	0.203***	0.203***
	(0.0194)	(0.0474)	(0.0430)	(0.0185)	(0.0300)	(0.0296)
Merit * Constant Efficiency		-0.161**	-0.156**		-0.0562	-0.0574
		(0.0695)	(0.0650)		(0.0422)	(0.0418)
Risk aversion (Eckel-Grossman)			0.0308***			0.000991
			(0.00663)			(0.00606)
(Payoff B - Payoff A)/Payoff A				0.526***	0.525***	0.527***
				(0.112)	(0.112)	(0.112)
Female dummy			0.0657***			0.0205
			(0.0206)			(0.0201)
Higher education dummy			-0.0481**			-0.0312
			(0.0234)			(0.0261)
Age			0.000693			0.000756
			(0.00217)			(0.00347)
Background in economics			-0.109***			-0.0112
			(0.0207)			(0.0204)
Currently employed			-0.0366			-0.0261
			(0.0301)			(0.0319)
Currently in a relationship			0.0480**			-0.00366
			(0.0224)			(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).

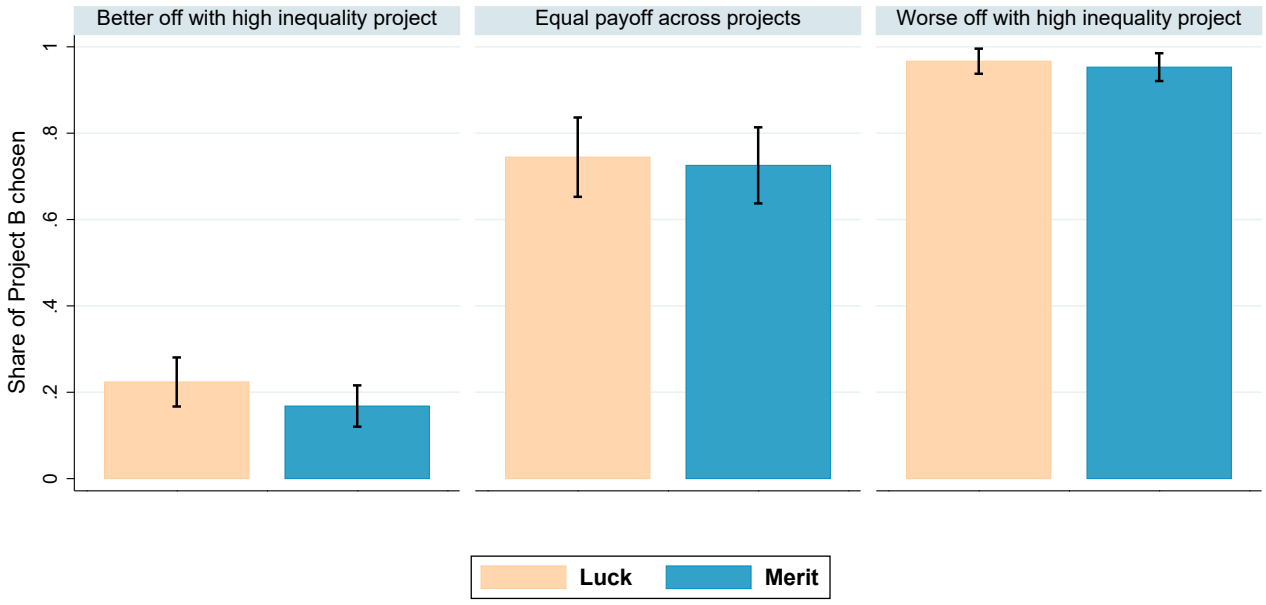


Figure 2.2 – Decisions with Known Position - Constant Efficiency Choices

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).

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 2.A3).

Instead of pooling the decisions by tradeoff type, one can also look at decisions for each choice (Figures 2.A5-2.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[13], 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[14].

Figure 2.3 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 2.A8-2.A10 display the distribution of choices made by subjects depending on their rank,

[13]Players 4 and 5 in Choices 1 and 2, and Player 5 in Choice 3
[14]Players 1 and 2 in Choice 1, players 1, 2 and 3 in Choice 2

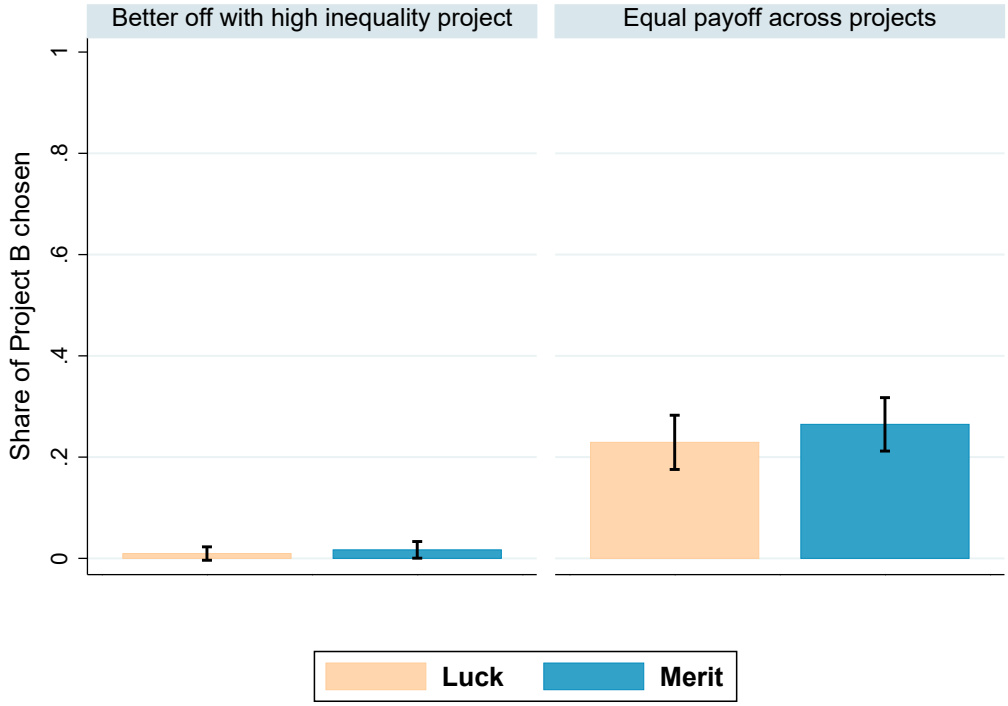


Figure 2.3 – Decisions with Known Position - Project A is Pareto-Dominant

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.

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 2.3 confirm these results. The only two significant predictors of 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 2.3, 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 2.A3 and 2.A2).

2.3.3 Observable heterogeneity

Table 2.3 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 2.A2 and 2.A3).

2.4 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 2.2, 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).

2.5 Robustness Checks

2.5.1 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 2.A3 and 2.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 2.A5 to 2.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 2.3.2. 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.

2.5.2 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 2.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 2.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 2.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.

2.5.3 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.[15] Table 2.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.[16]

2.6 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; Bjerk, 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).

[15] We did not ask subjects in the Luck treatment about their predicted position, in order to avoid confusion.
[16] Choice by choice regressions do not yield statistically significant results (regressions not show)

Bibliography

Alesina, Alberto and George-Marios Angeletos. Fairness and redistribution: US vs. Europe. *American Economic Review*, 95, 2005.

Alesina, Alberto and Eliana La Ferrara. Preferences for redistribution in the land of opportunities. *Journal of Public Economics*, 89(5):897–931, 2005.

Alós-Ferrer, Carlos, Jaume García-Segarra, and Alexander Ritschel. Performance curiosity. *Journal of Economic Psychology*, 64:1–17, 2018. ISSN 01674870. doi: 10.1016/j.joep.2017.08.002.

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. ISSN 01674870. doi: 10.1016/S0167-4870(03)00065-5.

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*, 2006. ISSN 00028282. doi: 10.1257/aer.96.5.1906.

Breza, Emily, Supreet Kaur, and Yogita Shamdasani. The morale effects of pay inequality. *The Quarterly Journal of Economics*, 2017.

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 Matthew Rabin. Understanding social preferences with simple tests. *The Quarterly Journal of Economics*, (3):817–869, 2002.

Charness, Gary and Marie Claire Villeval. Cooperation and competition in intergenerational experiments in the field and the laboratory. *American Economic Review*, 2009. ISSN 00028282. doi: 10.1257/aer.99.3.956.

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. ISSN 01674870. doi: 10.1016/j.joep.2017.11.001.

Dickinson, David L. An Experimental Examination of Labor Supply and Work Intensities. *Journal of Labor Economics*, 17(4):638–670, 1999. ISSN 0734-306X. doi: 10.1086/209934.

Durante, Ruben, Louis Putterman, and Joël Weelevan der . Preferences for Redistribution and Perception of Fairness: An Experimental Study. *Journal of the European Economic Association*, 12 (4):1059–1086, jul 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 Nikos Nikiforakis. Relative earnings and giving in a real-effort experiment. *American Economic Review*, 101(7):3330–3348, 2011. ISSN 00028282. doi: 10.1257/aer.101.7.3330.

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. An online recruitment system for economic experiments. 2004.

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. ISSN 09275371. doi: 10.1016/j.labeco.2018.04.006.

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. ISSN 01674870. doi: 10.1016/j.joep.2011.12.003.

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, Michał. 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, 2016. ISSN 00472727. doi: 10.1016/j.jpubeco.2016.08.012.

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. ISSN 01674870. doi: 10.1016/0167-4870(91)90006-F.

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. ISSN 01674870. doi: 10.1016/j.joep.2015.12.009.

Appendices

2.A Tables and Figures

2.A.1 Tables

Table 2.A1 – Balance Table

	(1) Luck	(2) Merit	(3) Overall	(4) Diff.	(5) 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
N	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 2.D.

Table 2.A2 – Who are the money burners?

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Var. = 1 if subject chooses to burn money						
Merit	0.0455	0.0465			0.0442	0.0529
	(0.0522)	(0.0523)			(0.0580)	(0.0545)
Risk aversion (Eckel-Grossman)	0.0115	0.0110	0.0171	0.0180	0.0143	0.00735
	(0.0152)	(0.0151)	(0.0234)	(0.0231)	(0.0174)	(0.0165)
Age	-0.00209	-0.00252	-0.0135	-0.0136	0.000596	-0.00597
	(0.00817)	(0.00825)	(0.00824)	(0.00872)	(0.00661)	(0.00520)
Female dummy	0.0546	0.0528	0.00259	-0.00599	0.0740	0.0760
	(0.0551)	(0.0550)	(0.0818)	(0.0819)	(0.0613)	(0.0577)
Currently employed	0.0327	0.0347	0.0704	0.0700	0.0408	0.0682
	(0.0531)	(0.0528)	(0.0759)	(0.0758)	(0.0675)	(0.0616)
Background in economics	-0.0261	-0.0244	0.0104	0.0130	-0.0507	-0.0754
	(0.0585)	(0.0586)	(0.0785)	(0.0784)	(0.0642)	(0.0608)
Currently in a relationship	0.0302	0.0306	0.113	0.114	-0.0130	0.0195
	(0.0816)	(0.0816)	(0.1000)	(0.100)	(0.0829)	(0.0782)
Higher education dummy	-0.0835	-0.0830	0.0455	0.0424	-0.119	-0.0855
	(0.0691)	(0.0690)	(0.104)	(0.105)	(0.0822)	(0.0701)
Rank		0.0168				
		(0.0239)				
Predicted Position			-0.0294			
			(0.0429)			
Rank - pred. pos.				-0.00441		
				(0.0306)		
Political Orientation					0.0417***	0.0131
					(0.0139)	(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
N	512	512	272	272	385	385
Pseudo-R^2	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 2.3). These decisions where 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 2.A3 – Who are the people that are willing to pay for more equality?

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Var. = 1 if subject chooses to reduce inequality at her own expense						
Merit	-0.0761	-0.0762	-0.0759	-0.0764	-0.0649	-0.0633
	(0.0423)	(0.0416)	(0.0424)	(0.0417)	(0.0438)	(0.0437)
Risk aversion (Eckel-Grossman)	-0.00981	-0.00738	-0.0109	-0.00844	-0.00313	-0.00319
	(0.0129)	(0.0128)	(0.0127)	(0.0126)	(0.0136)	(0.0135)
Age	0.00474	0.00438	0.00510	0.00480	0.00464	0.00489
	(0.00647)	(0.00601)	(0.00656)	(0.00604)	(0.00678)	(0.00670)
Female dummy	0.0132	0.0128	0.0138	0.0143	0.0106	0.0118
	(0.0419)	(0.0414)	(0.0419)	(0.0414)	(0.0447)	(0.0448)
Background in economics	-0.0608	-0.0606	-0.0556	-0.0512	-0.0248	-0.0230
	(0.0458)	(0.0456)	(0.0466)	(0.0465)	(0.0486)	(0.0492)
Currently employed	-0.160**	-0.170**	-0.159**	-0.170**	-0.233***	-0.234***
	(0.0728)	(0.0717)	(0.0722)	(0.0710)	(0.0876)	(0.0882)
Currently in a relationship	-0.0410	-0.0346	-0.0406	-0.0319	-0.00222	-0.00538
	(0.0482)	(0.0478)	(0.0479)	(0.0478)	(0.0481)	(0.0492)
Higher education dummy	-0.00812	-0.00978	-0.00198	0.000466	-0.0262	-0.0293
	(0.0520)	(0.0506)	(0.0539)	(0.0521)	(0.0507)	(0.0508)
(Payoff B - Payoff A)/Payoff A		0.486		0.581**	0.633**	0.643**
		(0.280)		(0.272)	(0.288)	(0.291)
Rank			-0.0184	-0.0318		
			(0.0208)	(0.0232)		
Political Orientation					-0.0379***	-0.0401***
					(0.0127)	(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
N	448	448	448	448	361	361
Pseudo-R^2	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. *(Payoff B - Payoff A)/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 2.A4 – Is choice of Project B *only* explained by risk?

	(1)	(2)
Dep. Var. = 1 if Project B (low inequality) of Choice 1 is chosen		
Safe lottery is chosen	0.256***	0.218***
	(0.0447)	(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
N	320	320
Pseudo-R^2	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 2.A5 – Confidence and Over-Confidence do not Predict Choices

	Constant Efficiency (excl. Choice 3)			Choice 3		
	(1)	(2)	(3)	(4)	(5)	(6)
Dep. variable = 1 if Project B (low inequality) is chosen						
Predicts rank 2	0.0692		0.0746	0.000800		0.00733
	(0.0711)		(0.0718)	(0.0798)		(0.0797)
Predicts rank 3	0.0523		0.0629	0.348***		0.363***
	(0.0744)		(0.0734)	(0.0936)		(0.0904)
Predicts rank 4	-0.0578		-0.0609	0.103		0.107
	(0.165)		(0.163)	(0.174)		(0.179)
Rank		-0.0215	-0.0240		-0.0208	-0.0318
		(0.0203)	(0.0202)		(0.0246)	(0.0241)
Risk aversion (Eckel-Grossman)	0.0552***	0.0575***	0.0554***	0.0492**	0.0538**	0.0488**
	(0.0182)	(0.0175)	(0.0178)	(0.0205)	(0.0216)	(0.0199)
Female dummy	0.129**	0.141***	0.129**	0.0906	0.130*	0.0896
	(0.0530)	(0.0533)	(0.0531)	(0.0694)	(0.0685)	(0.0694)
Higher education dummy	-0.0971	-0.0886	-0.0874	-0.0454	-0.0282	-0.0344
	(0.0692)	(0.0694)	(0.0702)	(0.0813)	(0.0834)	(0.0814)
Age	-0.00251	-0.000655	-0.000202	-0.0136	-0.0113	-0.0106
	(0.00773)	(0.00778)	(0.00808)	(0.00890)	(0.00950)	(0.00951)
Background in economics	-0.196***	-0.193***	-0.194***	-0.176***	-0.214***	-0.173***
	(0.0550)	(0.0531)	(0.0551)	(0.0616)	(0.0653)	(0.0617)
Currently employed	-0.0638	-0.0884	-0.0745	0.124	0.115	0.108
	(0.0911)	(0.0915)	(0.0920)	(0.101)	(0.114)	(0.103)
Currently in a relationship	0.133**	0.135**	0.132**	0.0703	0.0669	0.0689
	(0.0563)	(0.0559)	(0.0559)	(0.0683)	(0.0723)	(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 2.A6 – Choice table to elicit risk aversion using Eckel-Grossman method (Lottery 1)

Lottery	Low gain	High gain	Mean	St. Dev.	r	Choice share
Lottery 1	140	140	140	0	$3.46 < r$	10.00%
Lottery 2	120	180	150	30	$1.16 < r < 3.46$	11.25%
Lottery 3	100	220	160	60	$0.72 < r < 1.16$	31.88%
Lottery 4	80	260	170	90	$0.5 < r < 0.72$	11.56%
Lottery 5	60	300	180	120	$0 < r < 0.5$	14.69%
Lottery 6	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 2.A7 – Choice between two lotteries

	Lottery A	Lottery 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.

2.A.2 Figures

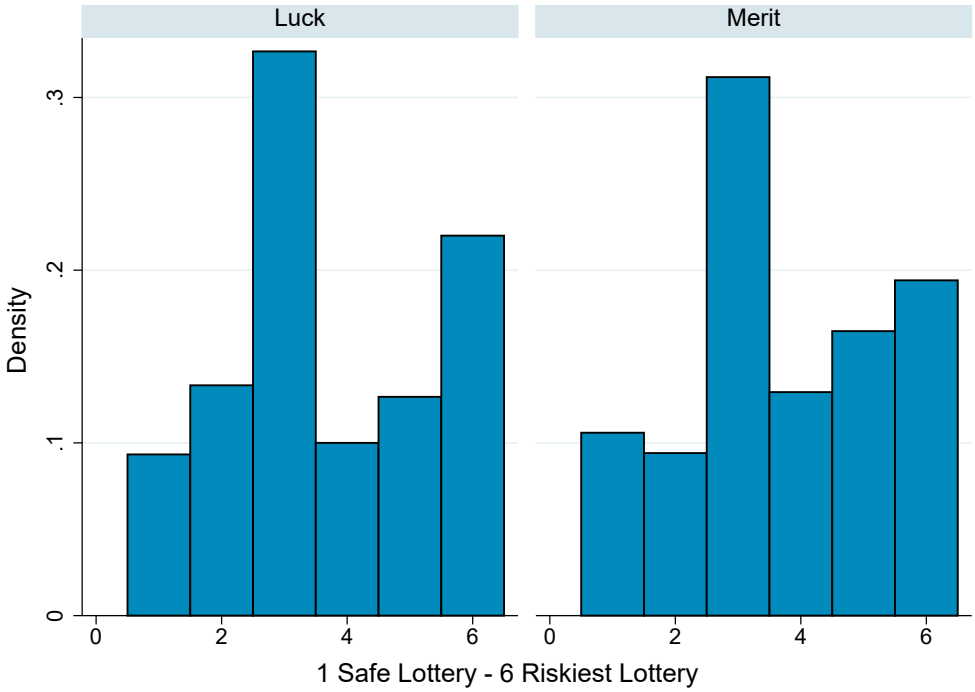


Figure 2.A1 – Distribution of the Choices Made for Lottery 1

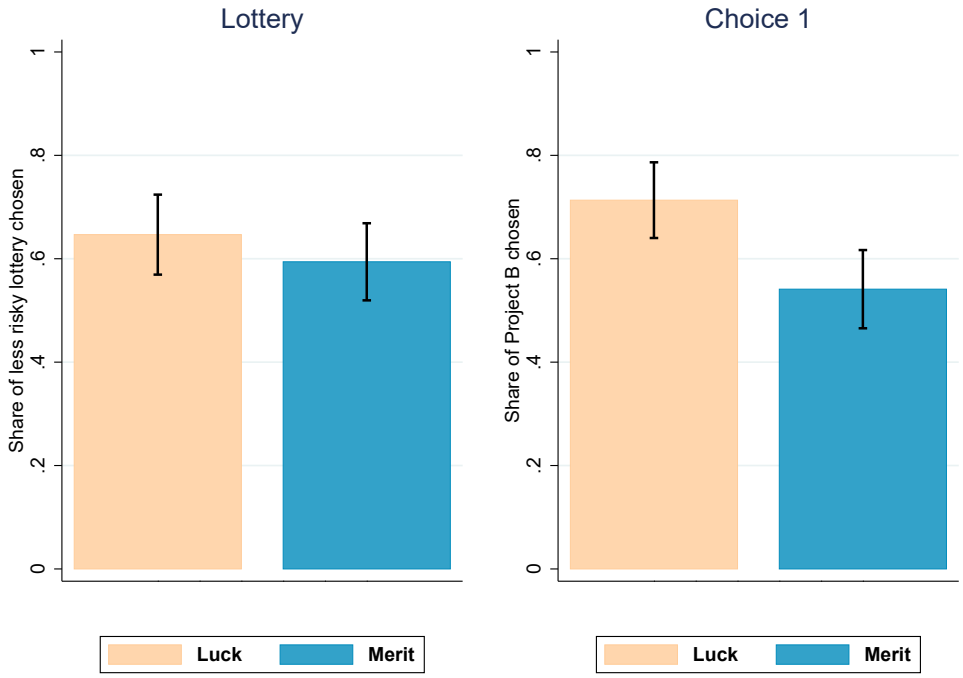
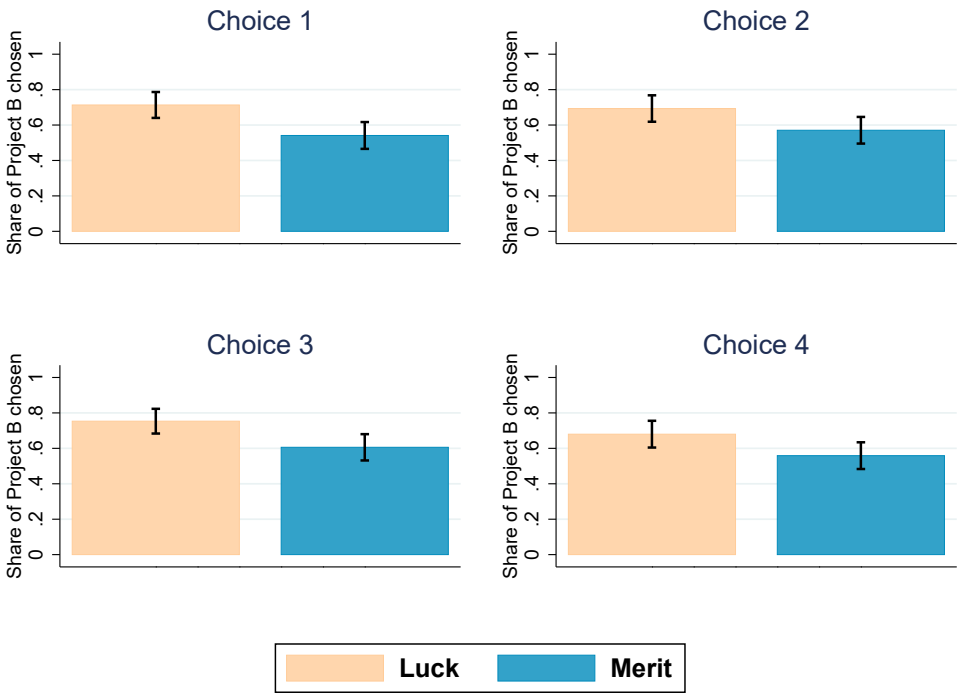


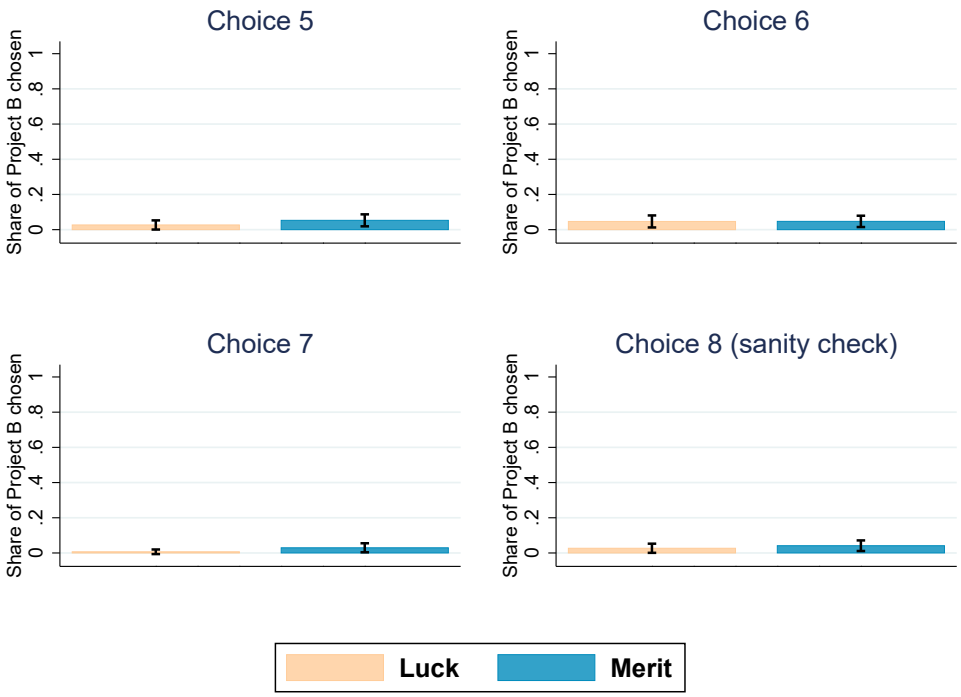
Figure 2.A2 – Differences between Choice 1 and the payoff-equivalent lottery
Notes: The lottery and Choice 1 are payoff-equivalent. In the lottery setting, subjects choose between two lotteries that have 5 equally likely payoffs. These payoffs are identical to the ones of Project A and Project B in Choice 1. We display 95% confidence intervals.

Figure 2.A3 – Distribution of the Choices Made Behind the Veil of Ignorance - **Constant Efficiency Choices**



Notes: We display 95% confidence intervals.

Figure 2.A4 – Distribution of the Choices Made Behind the Veil of Ignorance - **A Pareto-Dominant project**



Notes: We display 95% confidence intervals.

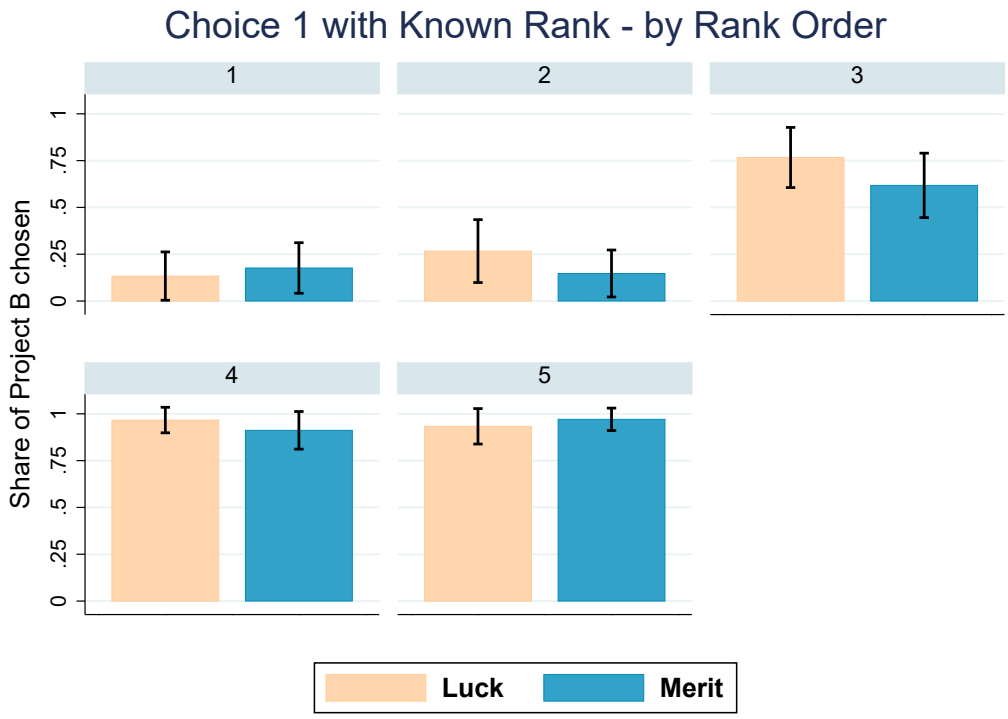
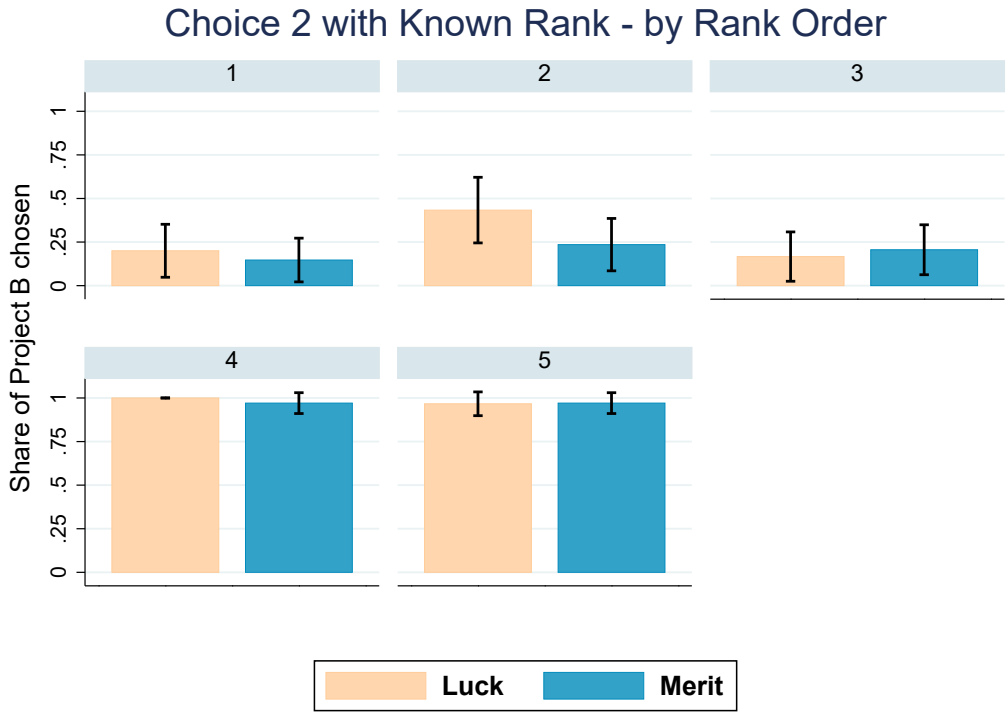


Figure 2.A5 – Choice 1 (Constant Efficiency Choice) with Known Rank - by Rank Order

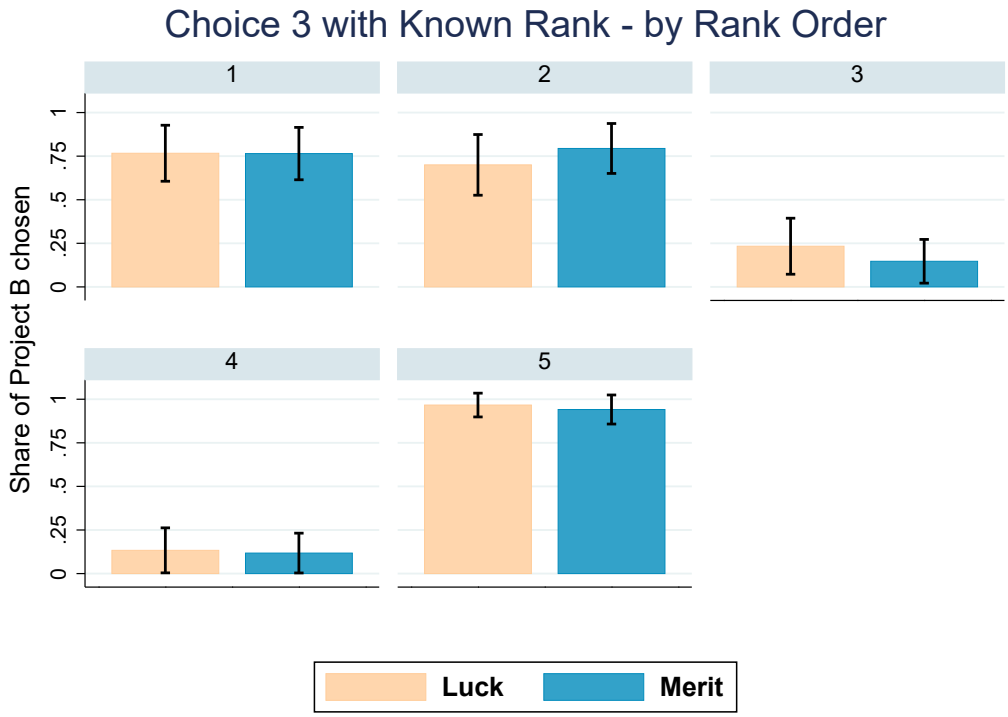
Notes: The graph shows the share of Project B chosen in Choice 1 depending on Treatment groups (Luck vs Merit) and the rank that the subject achieved after the real effort task, with 95% confidence intervals. The top left panel shows the choices made by players that achieved the top rank and the bottom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

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



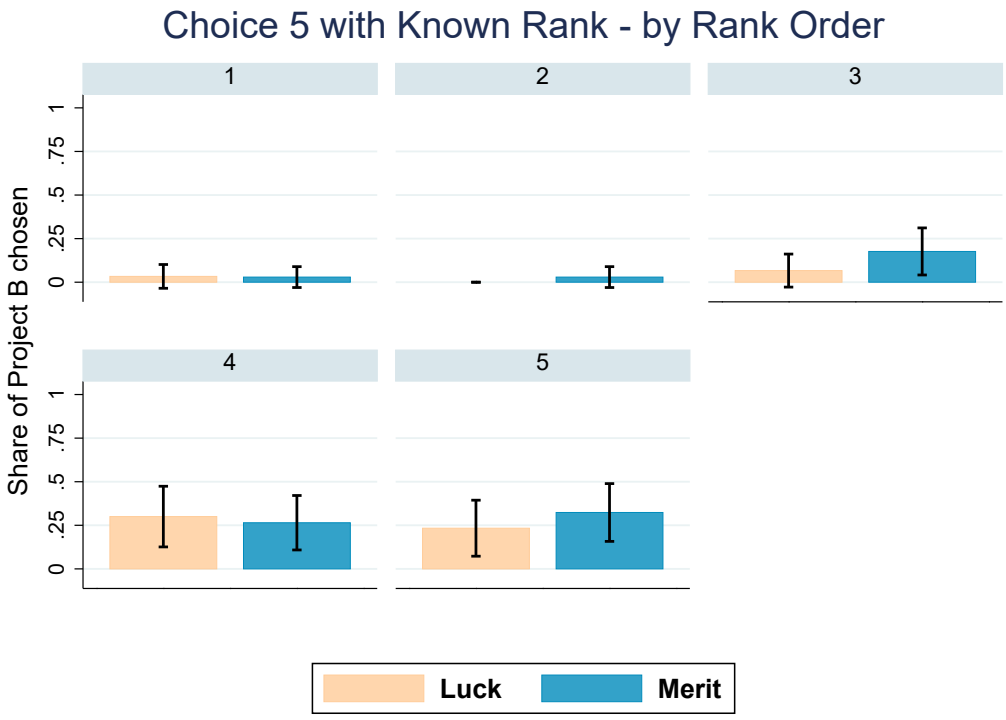
Notes: The graph shows the share of Project B chosen in Choice 2 depending on Treatment groups (Luck vs Merit) and the rank that the subject achieved after the real effort task, with 95% confidence intervals. The top left panel shows the choices made by players that achieved the top rank and the bottom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

Figure 2.A7 – Choice 3 (Constant Efficiency Choice) with Known Rank - by Rank Order



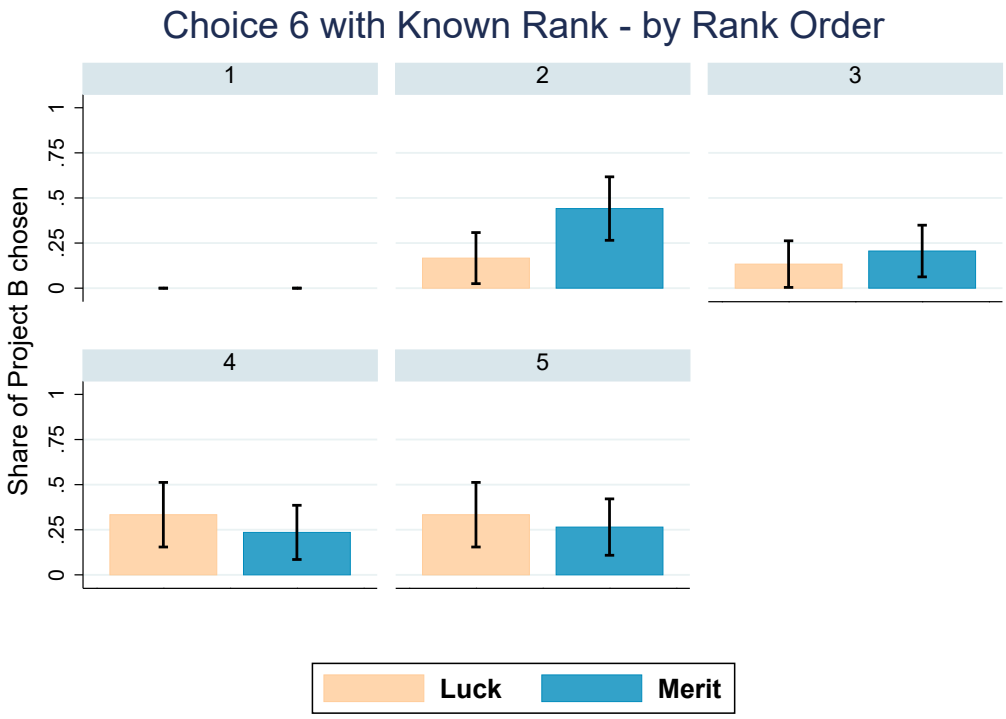
Notes: The graph shows the share of Project B chosen in Choice 3 depending on Treatment groups (Luck vs Merit) and the rank that the subject achieved after the real effort task, with 95% confidence intervals. The top left panel shows the choices made by players that achieved the top rank and the bottom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

Figure 2.A8 – Choice 5 (A Pareto-Dominant) with Known Rank - by Rank Order



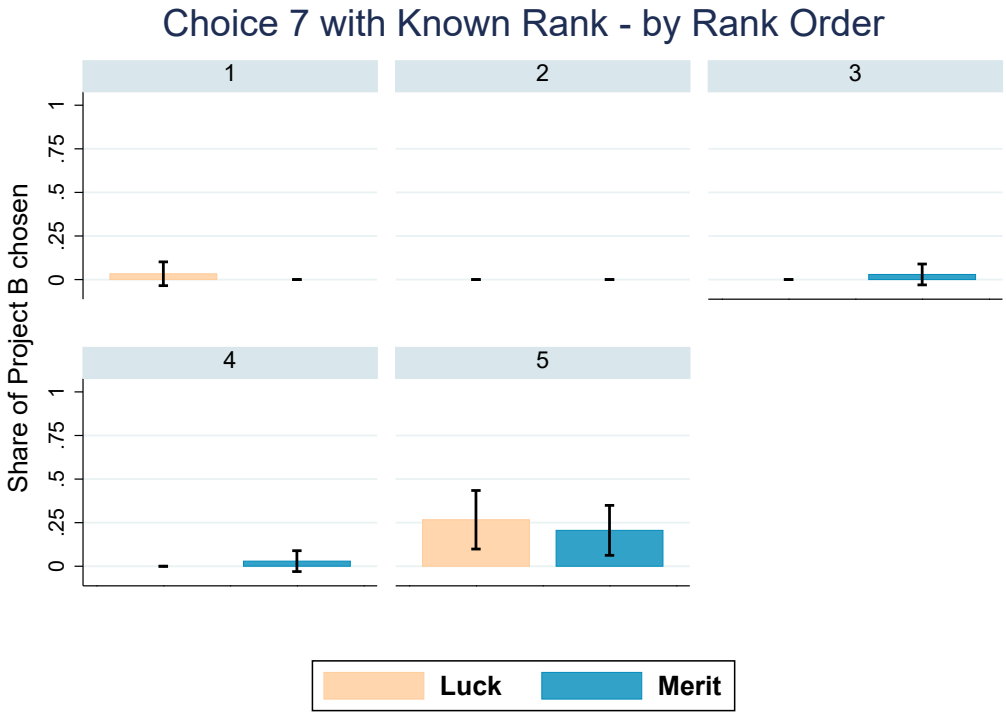
Notes: The graph shows the share of Project B chosen in Choice 5 depending on Treatment groups (Luck vs Merit) and the rank that the subject achieved after the real effort task, with 95% confidence intervals. The top left panel shows the choices made by players that achieved the top rank and the bottom 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 2.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 2.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.

2.B Hypothetical Decisions

Hypothetical choices are summarized in Table 2.B1. 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 2.B1 – Hypothetical situations

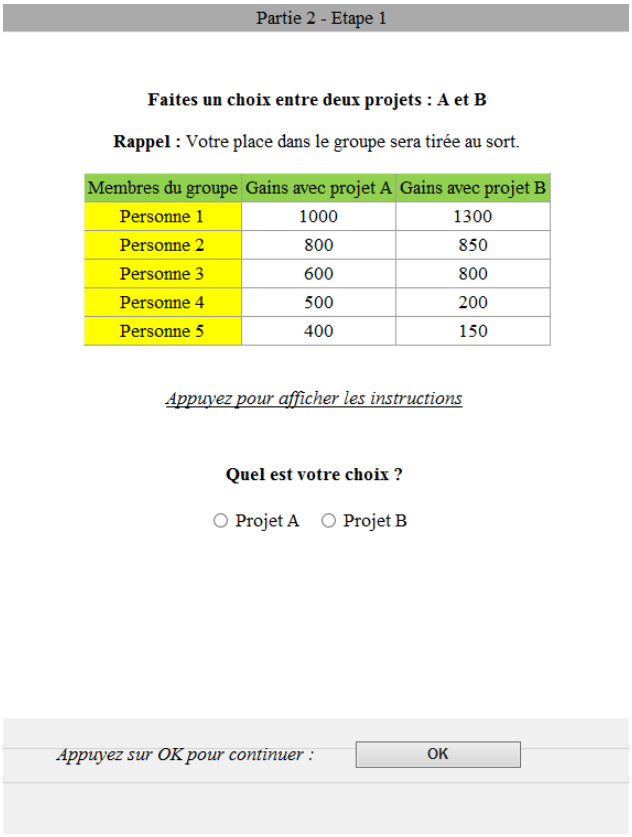
	Choice 4		**Choice 8**	
	A	**B**	**A**	**B**
Person 1	1400	1100	1200	**800**
Person 2	1200	800	1100	700
Person 3	**600**	700	1000	600
Person 4	350	650	900	500
Person 5	300	**600**	**800**	400
Information below not shown to subjects				
Total	3850	3850	4100	3300
Std. dev.	502	199	303	241
GINI	0.32	0.12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

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

2.C Screenshots of the Experiment

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

(a) Luck group



(b) Merit group

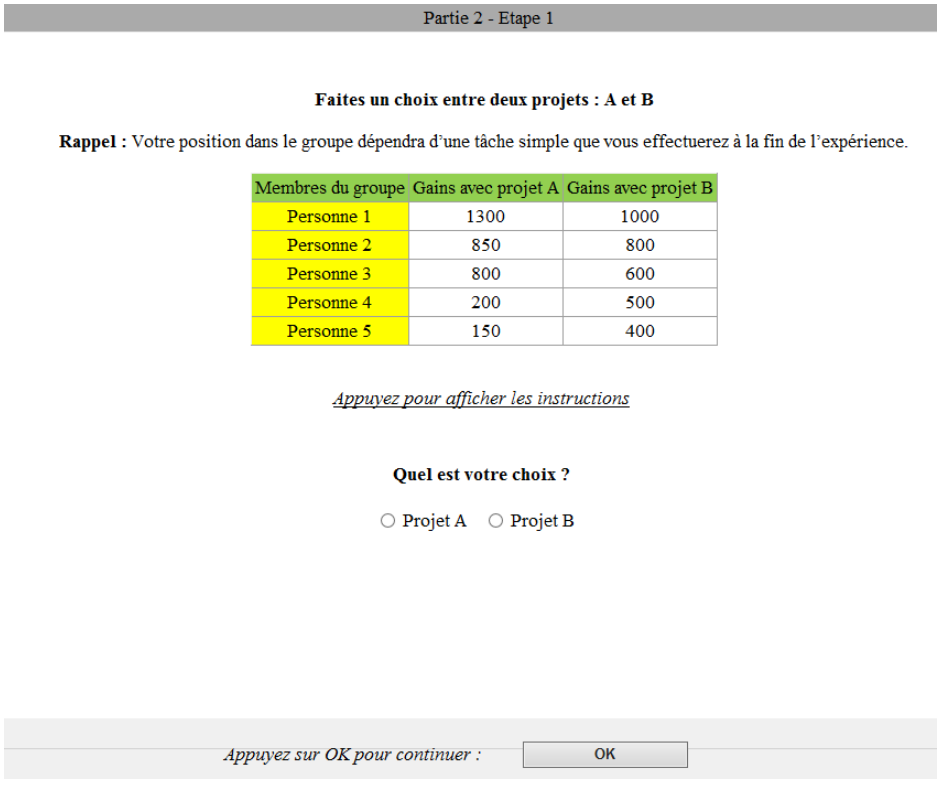
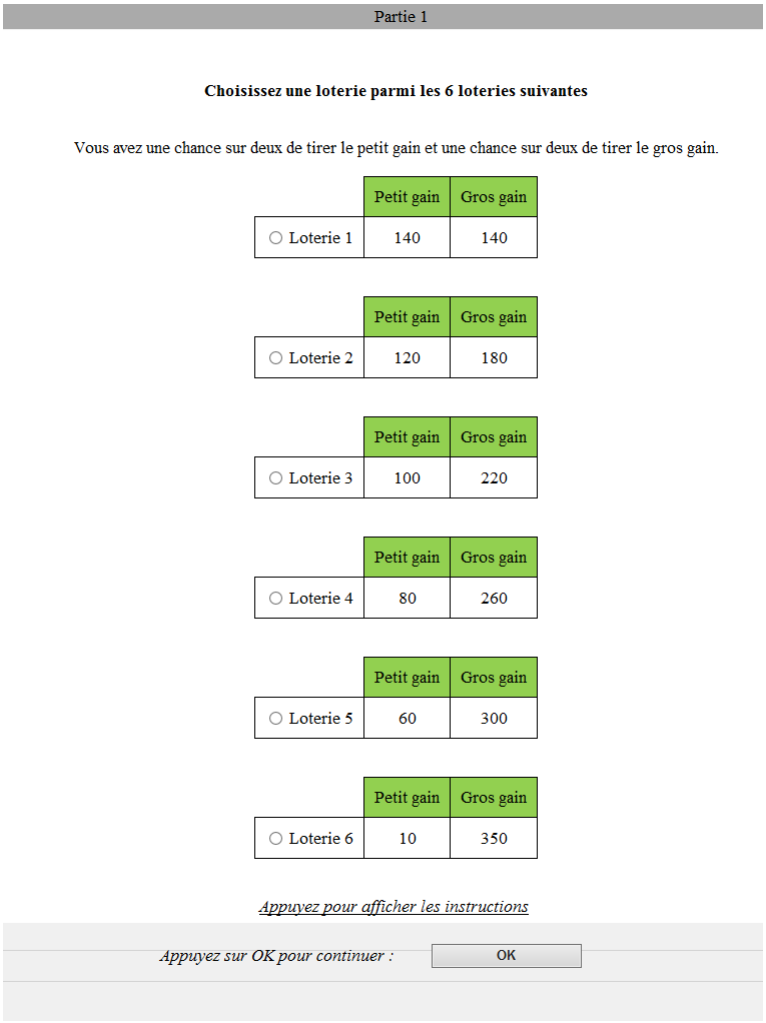


Figure 2.C2 – Screenshot of Risk Aversion Elicitation

(a) Eckel-Grossman Task



(b) Decision between two lotteries equivalent to Choice 1

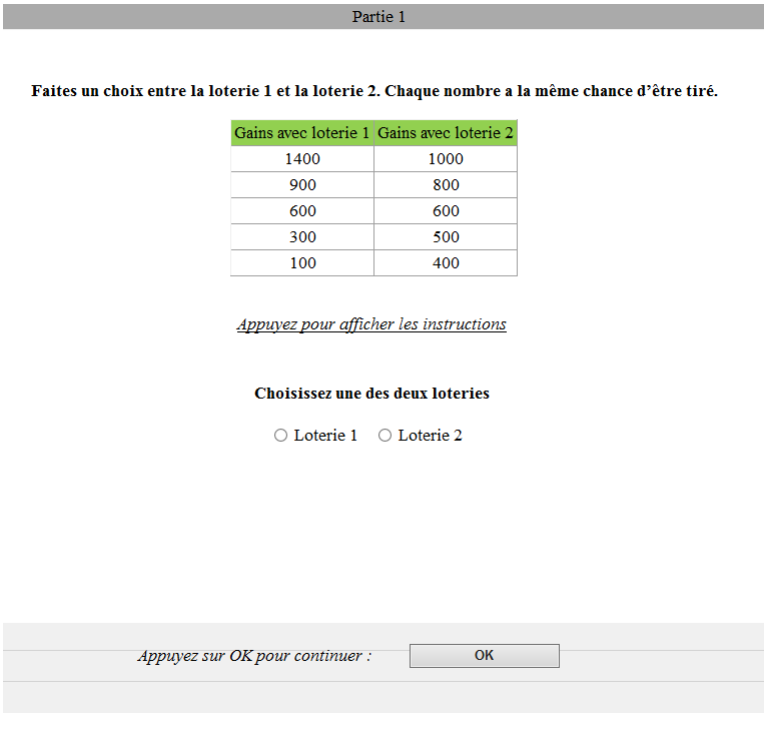


Figure 2.C3 – Screenshot of Choice with Known Position (Choice 6, Rank 2)

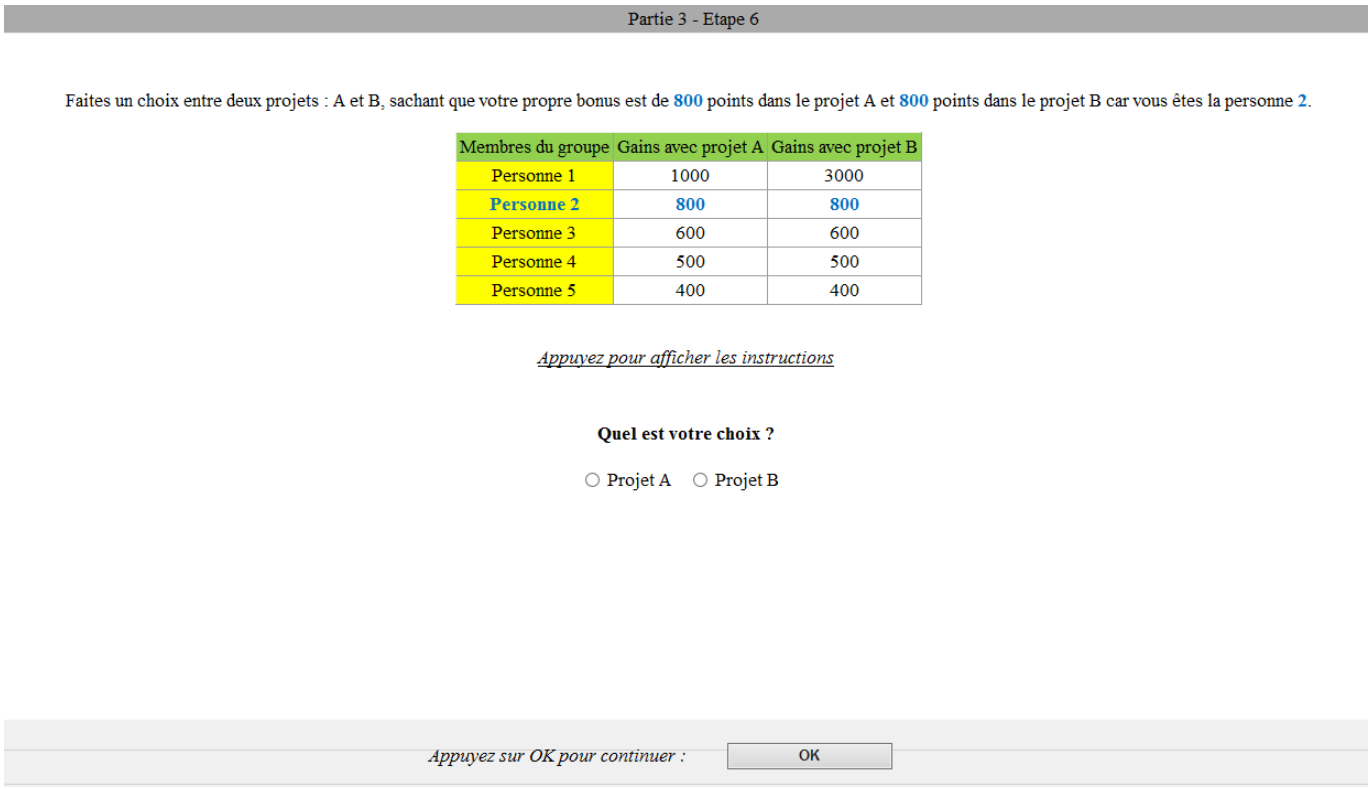
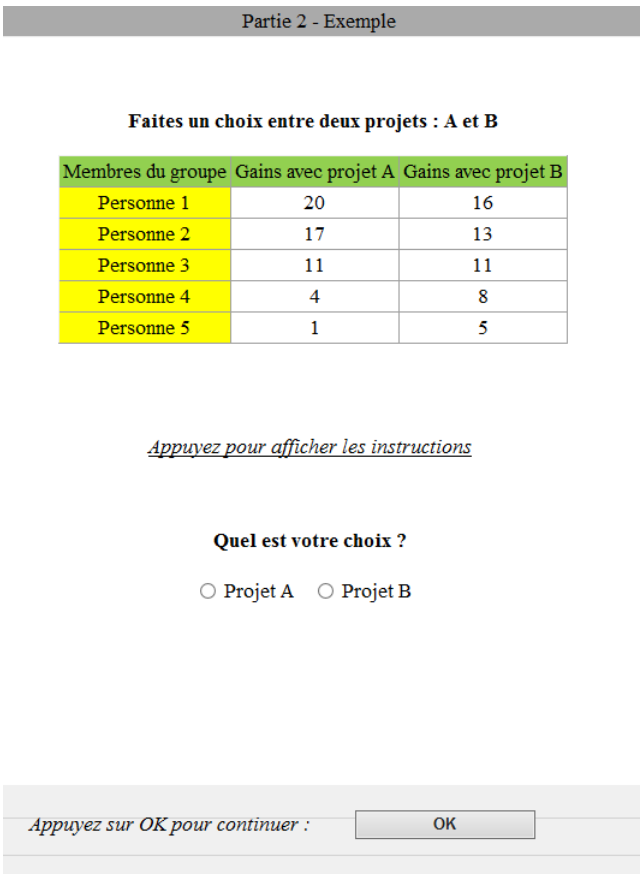


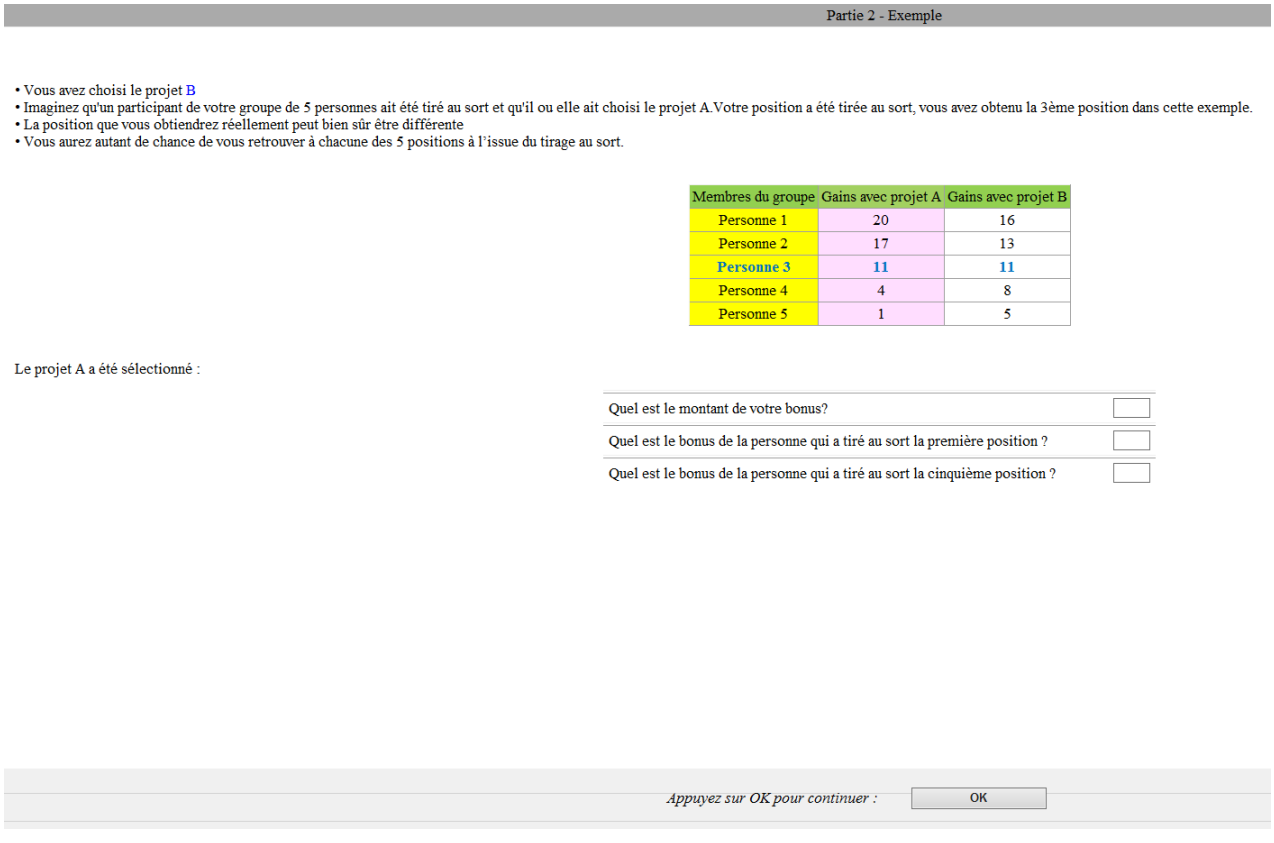
Figure 2.C4 – Screenshot of the Real Effort Task

Figure 2.C5 – Comprehension test

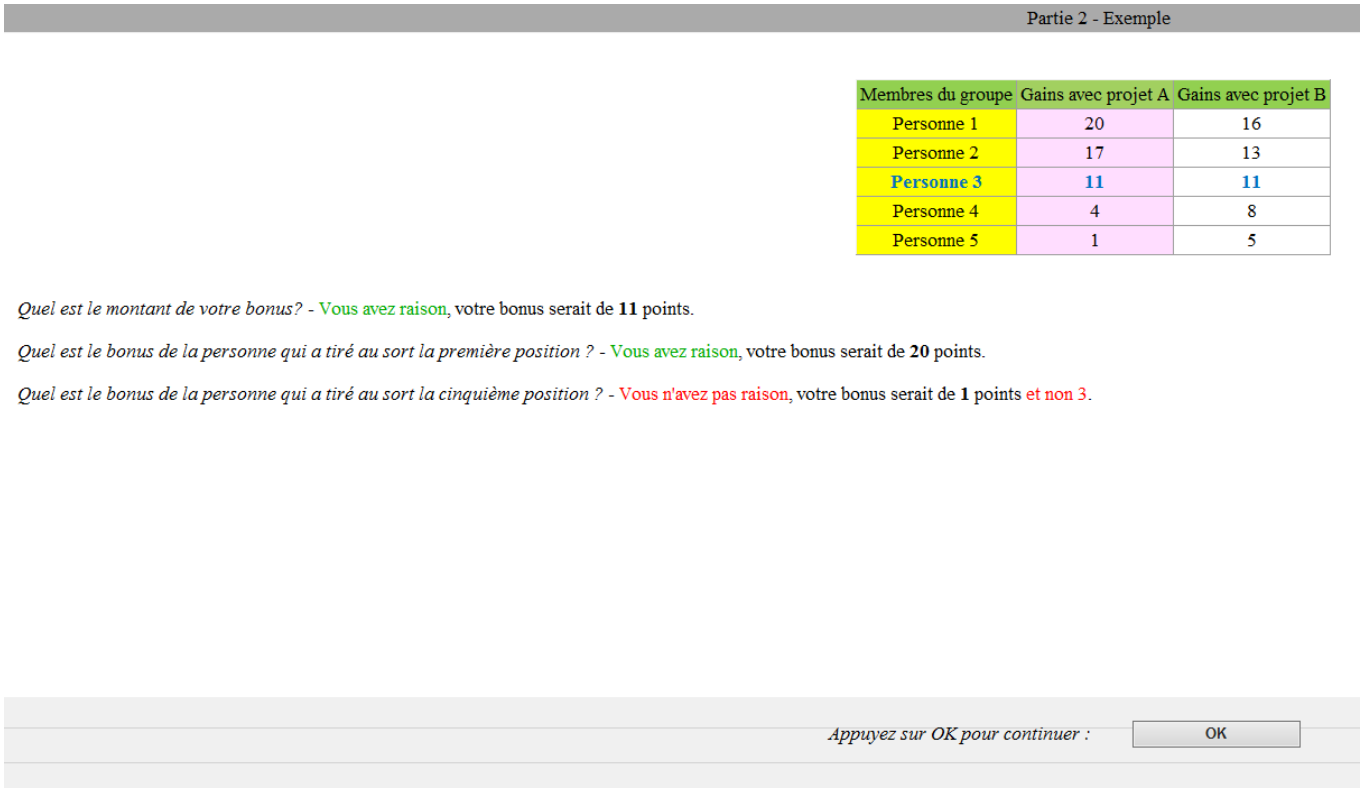
(a) Screen 1



(b) Screen 2



(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
Personne 3 | 11 | 11
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 points et non 3.
Appuyez sur OK pour continuer :
OK

2.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 was 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 paper. This difference in the post-experimental questionnaire across subjects could not affect the core results of this paper 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?

2.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.

2.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.

2.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 €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 2.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 2.C2b.

2.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 2.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 2.C1.
Text in Figure 2.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 2.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

2.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.

2.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 2.C3.

Text in Figure 2.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.

2.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 2.D.

Bibliography

Charness, Gary and Brit Grosskopf. Relative payoffs and happiness: an experimental study. *Journal of Economic Behavior & Organization*, 45(3):301–328, 2001.

Chapter 3

Principals' Distributive Preferences and the Incentivization of Agents

This chapter is a joint work with Max Lobeck.

3.1 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 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 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 efficiency incentives than others because their pay is indexed on 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. So 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, efficiency-minded, 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 efficiency concerns, 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 to 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 rest. This provides control over the efficiency-equality trade-off that principals believe they face before workers start working.

We find that despite the firm-like setting 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 efficiency 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 equality in procedure as such is not seen as a particularly attractive contract characteristic and principals are more interested in distributive 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. The structural estimates allow us to make counterfactual statements on how the implications of these preferences on firm performance change once we assume that agents hold horizontal social preferences.

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 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, 2004, 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 contribution to this literature is threefold. First, we show that principals' normative distributive preferences affect workers' incentive schemes, which in turn affect the firm's overall performance. Second, we establish that these preferences play a variable role in decision-making according to whether the principal has a stake in the firm's outcome. Finally, we take an ex-ante perspective, in the sense that principals choose incentive schemes *before* agents have exerted any effort. We thus take account of the fact that managers typically make decisions in an uncertain environment.

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., 2018). 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 (2010) 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., 2017; Cohn et al., 2014; Gross et al., 2015; Eisenkopf and Teyssier, 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 paper is structured as follows: section 3.2 presents survey evidence, section 3.3 introduces the design, section 4.3 presents our main results, section 3.5 presents the structural model, and section 4.5 concludes.

3.2 Managers' preferences: survey evidence

We use the two waves (2011 and 2017) of a French survey on Professional Relationships and Firm Negociations (REPONSE), that has been conducted every six years since 1993.[1] The survey was administered in 4,023 firms in 2011 and 4,364 in 2017[2] and three types of questionnaires are distributed, one for a representative of the executive managerial positions[3], one for a personnel representative[4], 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. [5] We use these outcome variables as indicators of whether the firm engaged in pay-for-performance and thus wage (or bonus) differentiation based on effort or ability, for both types of workers. In our experiment, we proxy such kinds 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 3.1 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

[1] We only had access to the last two waves via the *Réseau Quetelet*, as the earliest ones are no longer available.

[2] 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.

[3] Either CEO, Secretary General, Plant Director, Head of Human Resources, or another top managerial position.

[4] Either a union representative or a staff representative.

[5] *Cadres* and *Non Cadres* in French.

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 3.1 – Individualized wage raises and managers' distributive preferences

	White-collar workers			Blue-collar workers		
	(1)	(2)	(3)	(4)	(5)	(6)
Dep var = Did white/blue-collar workers benefit from individualized wage raises						
Individualized wage raises are unfair	-0.200***	-0.103***	-0.0830***	-0.209***	-0.138***	-0.119***
	(0.0127)	(0.0140)	(0.0151)	(0.0118)	(0.0130)	(0.0145)
Individualized wage raises create tension		-0.146***	-0.0993***		-0.116***	-0.0881***
		(0.0102)	(0.0110)		(0.0101)	(0.0113)
Individualized wage raises motivate		0.0678***	0.0579***		0.0425***	0.0317*
		(0.0165)	(0.0178)		(0.0156)	(0.0174)
Individualized wage raises are subjective		-0.111***	-0.0418***		-0.0926***	-0.0650***
		(0.0120)	(0.0131)		(0.0114)	(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 4.C1 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.[6] In columns (3) and (6), we also control for the socio-demographic characteristics of the manager answering the survey and a large set of firm-related controls, in order to minimize the self-selection issue.[7]

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.[8] Appendix Table 3.A2 shows the same regressions but focuses on the implementation of performance-based bonuses, rather than pay raises, as the dependent variable. Results

[6] Appendix Table 4.C1 describes all these variables, including an English translation of the original questions.

[7] Of course, we do not claim that such an empirical strategy is sufficient to establish a causal relationship.

[8] 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)

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. This survey indicates that managers in real situations are sometimes willing to abandon a (theoretically) efficient tool to increase production (performance pay) in order to avoid a conflict with their own normative distributive preferences. We use this evidence as a starting point to investigate the causal impact of principals' distributive preferences in a more controlled lab environment.

3.3 Experimental design

3.3.1 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".[9] We inform all participants that the currency used during the experiment is the ECU with the following conversion rate: 1€ = 10 ECU. The detailed instructions (translated from French to English) are included in Appendix 3.D.

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, 2004). 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.[10] 226 participants were randomly assigned to the worker role and 113 to the principal role.

3.3.2 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 α_i is the marginal productivity which varies across workers ($\alpha \in \{\alpha_H, \alpha_L\}$). In all sessions, we define $\alpha_H = 60$ and $\alpha_L = 40$.

α_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

[9] 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.

[10] 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.

comprehension test.[11] They are informed that performing better at the aptitude test will increase their chances of having higher productivity (a high α). Using an aptitude 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. We do not want principals to think that ability is completely arbitrary. 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.[12] Ability is determined at the pair level. We assign α_H to the worker with the best performance within the pair and α_L to the other one.

The cost function is constant across agents and it is convex in effort choices. Figure 3.A2 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.[13]

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 ($e \in \{0, 0.5, 1, ..., 5\}$). 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 $c(.)$ is the effort cost function.

A screenshot of agent B's decision can be found in the appendix, Figure 3.A3. 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 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.

[11] 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 3.E describes this test further and how the subjects performed.

[12] Appendix 3.F 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.

[13] 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.

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.

3.3.3 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 3.A2). Labels A and B are randomized and thus independent of the ability level.[14] 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.[15]

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€).[16] The drawing of the payoff-relevant piece rate in the belief elicitation is completely independent from 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.

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.

[14] 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

[15] 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 3.E for more details regarding the comprehension test.

[16] 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.

Table 3.2 – Set of decisions made by the principal assuming workers' choose effort to maximize their own income

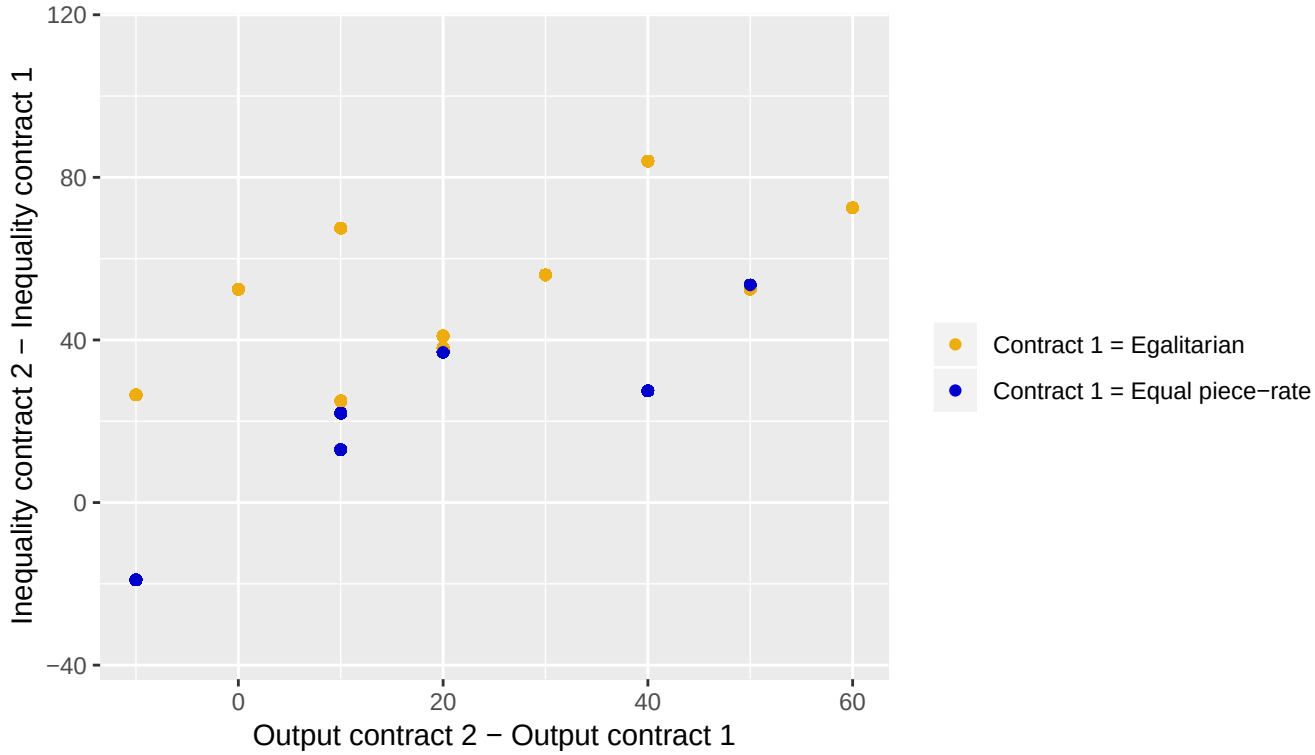
	piece rates		**Income Contract 1**		**Income Contract 2**		**Joint output**		
	Contract 1	**Contract 2**	$Worker_H$	$Worker_L$	$Worker_H$	$Worker_L$	**Contract 1**	**Contract 2**	**Δ-output**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Egalitarian VS output-maximizing choices									
N°									
1	0.4 – 0.6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

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, the 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 3.2, 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). The choices ask the principals to decide between an efficient (the total output is maximized) and an egalitarian contract ($Worker_L$ receives a higher piece rate compared to $Worker_H$), or an equal piece rate contract (both workers receive the same piece rate). The equal piece rate contract pays the workers in proportion to their output level, while the egalitarian contract assigns a higher piece rate to the lower-productivity worker to ensure that the workers are paid in proportion to their *effort level*. If workers best respond to wages, the egalitarian piece rate will either perfectly equalize income levels (Choices 1-3 and 8-10) or significantly decrease inequality (Choices 4-7). Note that we label Contract 2 as "output-maximizing", although it is not always efficient, even if workers best respond to it (see Choices 4 and 13). We also want to test also 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. For the sake of simplicity, we abstract from the two exceptions

and retain "output-maximizing" label when referring to Contract 2.

Figure 3.1 – 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.

The 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 3.1 shows how differences in inequality between Contract 2 and 1 (on the y-axis) and efficiency (difference in output between Contracts 1 and 2, on the x-axis) vary across choices. Blue dots represent each case in which Contract 1 is egalitarian and orange 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 -19 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 quickly 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 maximizing output and equality.

Figure 3.A4 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[17], 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 random order at the subject-level. Within a choice, the labeling of contracts as "Contract 1" or "Contract 2" is randomized. This implies that people cannot always choose Contract 2 to maximize their own income. On each occasion, they must check which contract maximizes income with a view to optimizing efficiency. 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 (and sessions), we will implement two treatments: (1) the ***spectator treatment*** and (2) the ***stakeholder 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 male a 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.

3.3.4 Summary statistics

Table 3.A3 shows the subjects' socio-demographic 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 3.A4 reports

[17]Note that they are not shown the production and cost of each worker for each effort level, 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.

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.

3.4 Results

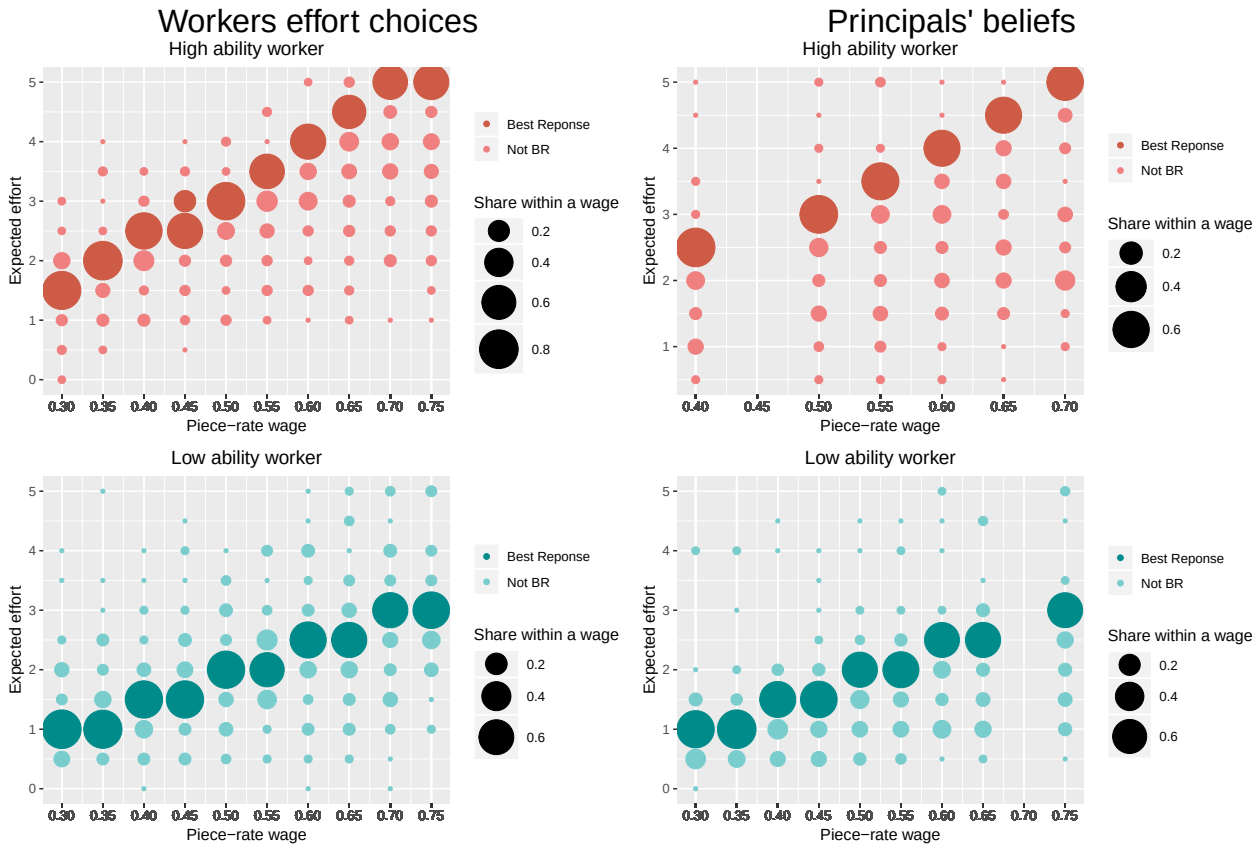
3.4.1 Effort choices and effort beliefs

We first describe, side-by-side, the effort levels chosen by workers for each piece rate wage and principals' corresponding beliefs. Figure 3.2 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 a 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 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.

3.4.2 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 3.3 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 unobserved characteristics. We thus present results using both the theoretical and the belief-based trade-offs to account for these two aspects.

Figure 3.2 – Workers' stated effort and principals' expected effort by piece rate wage



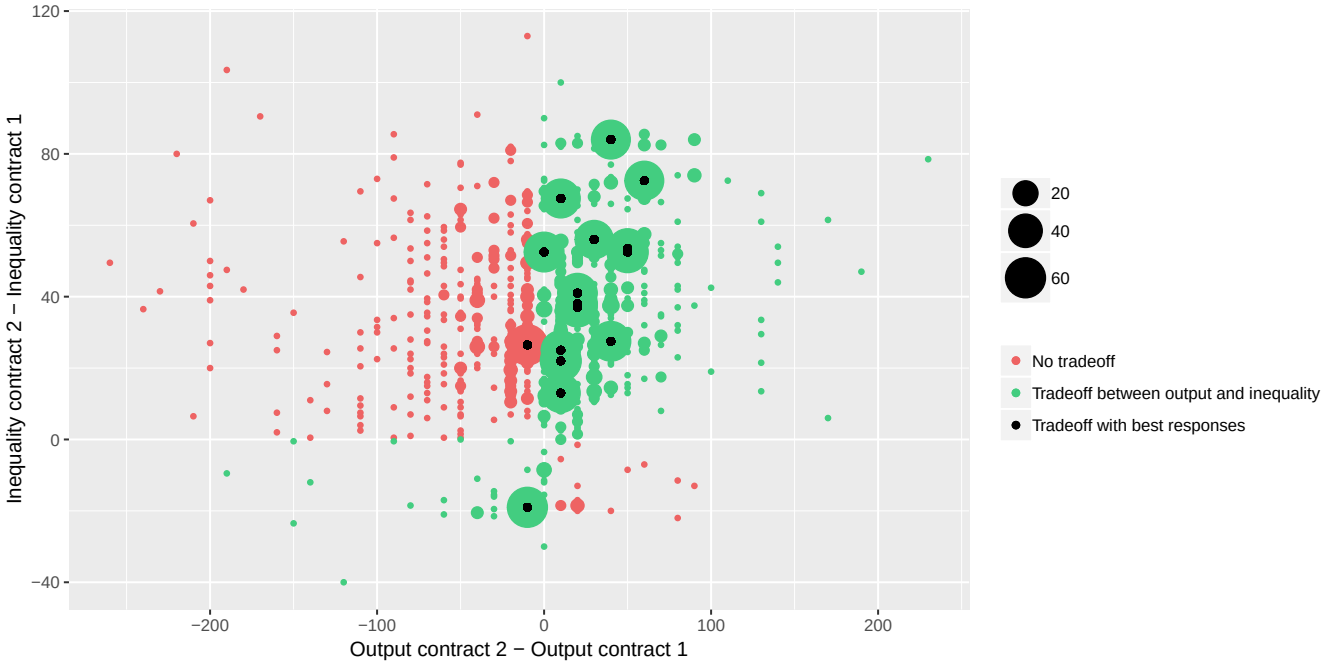
Notes: The figures on the left-hand side plot the workers' choices of effort level for each piece rate (on the x-axis) by ability type. The figures on the right-hand side plot principals' beliefs regarding the effort level chosen by workers for each piece rate. High-ability workers are in red and low-ability 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 3.2.

On the x-axis of Figure 3.3, 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 3.3 shows that many decisions are consistent with our theoretical trade-offs, as expected given the belief-elicitation results in Section 3.4.1.

We further classify tradeoffs 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 3.4.3 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.

Figure 3.3 – 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 3.1. 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).

3.4.3 Principals' choices

We now describe the pattern of choices across treatment groups. The y-axis of Figure 3.4 shows the share of cases in which the most egalitarian contract of the pair is selected. This corresponds to Contract 1 in all cases, except for Choice 13.[18] We plot this share by the size of the trade-off: Contract 2 increases in efficiency 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. The top panel shows the choices based on theoretical trade-offs (assuming workers' best responses) and the bottom panel focuses on belief-based trade-offs.

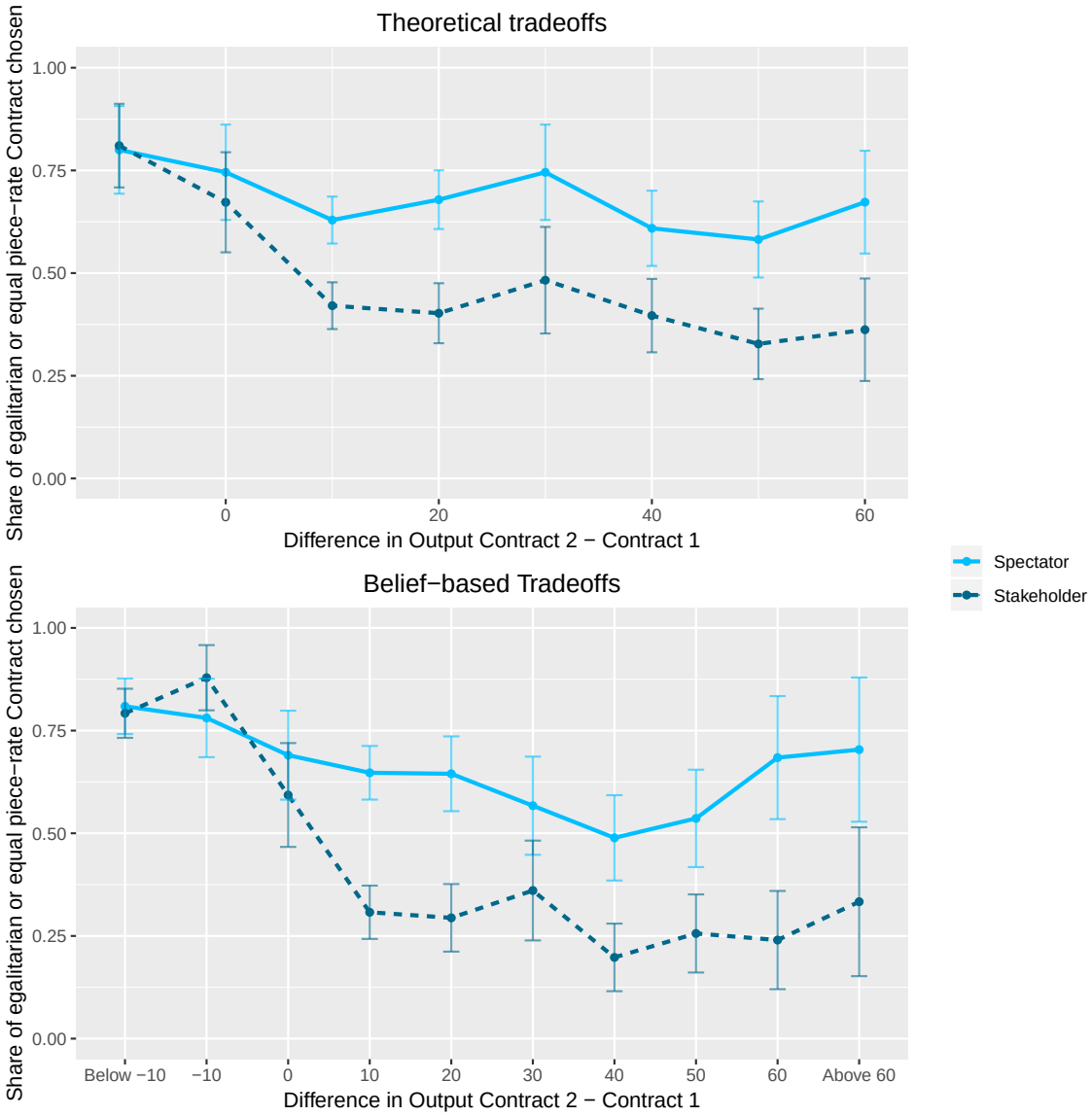
Overall, we find that, on average, both treatment groups compress wages to a certain extent, given that for all trade-offs, the share of Contract 1 decisions is always significantly different from 0. This confirms our hypothesis that, generally speaking, principals hold distributive preferences. Now 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 efficiency in relation to Contract 1. On the contrary, Spectators seem less sensitive to output differences.

Note that the outliers we described in Section 3.4.2 can only affect the first and end points of the

[18] For Choice 13, Contract 2 is the most egalitarian contract of the pair.

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 3.4 – 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). This corresponds to Contract 1 in all cases, except for Choice 13, where Contract 2 is the most egalitarian contract. We calculate these shares by output trade-off, i.e. the difference in output between Contract 2 and Contract 1 (except for Choice 13 where it is the difference in output between Contract 1 and Contract 2). 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 top panel shows the choices based on theoretical trade-offs (assuming workers' best responses) and the bottom panel focuses on belief-based trade-offs. We show 95% confidence intervals for the shares.

We can also use a regression analysis to characterize principals' choices. Table 3.3 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 two dummy variables: one for whether the observation is about Choice 1 and the other for whether it is about Choice 13. We include these dummies because Choices 1 and 13 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 3.3 calculate these trade-offs assuming that workers best respond (theoretical trade-offs), which can be interpreted as reduced-form estimates that are not biased by the heterogeneity in beliefs. The drawback of these measures is that they may be less precise given that principals may expect deviations from 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 3.3 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 3.4 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 3.4. 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. This effect becomes insignificant once we control explicitly for a choice of 1 or 13 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.

The alternative of 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

Table 3.3 – Regressions that characterize Contract decisions

	Theoretical trade-offs		Belief-based trade-offs	
	(1)	(2)	(3)	(4)
Dependent variable: Contract 2 (high inequality) was chosen				
Stakeholder	0.177**	0.177**	0.266***	0.265***
	(0.0837)	(0.0838)	(0.0771)	(0.0780)
$\frac{\Delta(\text{Output 2 and 1})}{10}$	0.0467***	0.0433***	0.0316***	0.0302***
	(0.0129)	(0.0132)	(0.00667)	(0.00653)
$\frac{\Delta(\text{Output 2 and 1})}{10}$ * Stakeholder	0.0300*	0.0300*	0.0339***	0.0339***
	(0.0173)	(0.0173)	(0.0113)	(0.0112)
$\frac{\Delta(\text{Inequality 2 and 1})}{10}$	-0.0483***	-0.0366**	-0.0218**	-0.0133
	(0.0138)	(0.0144)	(0.00994)	(0.00996)
$\frac{\Delta(\text{Inequality 2 and 1})}{10}$ * Stakeholder	-0.00171	-0.00178	-0.0180	-0.0178
	(0.0184)	(0.0184)	(0.0144)	(0.0145)
1 is equal piece rate	-0.0864*	-0.0461	-0.0263	0.00466
	(0.0451)	(0.0448)	(0.0415)	(0.0407)
1 is equal piece rate * Stakeholder	0.00669	0.00619	-0.00421	-0.00462
	(0.0637)	(0.0637)	(0.0543)	(0.0545)
Choice 1 = 1		0.123***		0.102**
		(0.0431)		(0.0468)
Choice 13 = 1		0.0691		0.0792*
		(0.0427)		(0.0463)
Constant	0.476***	0.409***	0.428***	0.375***
	(0.0873)	(0.0872)	(0.0793)	(0.0787)
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 current observation concerns Choice 1 (Choice 1 = 1) or Choice 13 (Choice 13 = 1). 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.

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 more interested in implementing their preferred outcome rather than treating both agents identically.

The only instance where equality in procedure seems to make a difference is when we ask subjects to choose *directly* between an egalitarian and equal piece rate contract. In this case, subjects are significantly more likely to choose the equal piece rate contract, as suggested by the positive and significant dummies of Choices 1 and 13 (significant on the 5 and 10 percent level). This shows that on average, subjects are more likely to choose an equal piece rate contract if it is posited directly against an egalitarian contract, even after taking contract characteristics into account.[19]

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.

Table 3.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.[20] Table 3.A7 replicates Table 3.3 but excludes observations where the characteristics of the contracts imply that Contract 2 yields a lower output relative to Contract 1, i.e. $\frac{\Delta(\text{Output 2 and 1})}{10} \leq 0$. This enables us to show that the intensive margin results do not only reflect a threshold effect, i.e. the point at which reducing inequality (choosing Contract 1) comes at a cost to output. We can see that the coefficient of $\frac{\Delta(\text{Output 2 and 1})}{10}$ is still positive and significant for both theoretical and belief-based tradeoffs. Finally, Table 3.A6 replicates Table 3.3 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.

3.5 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.

[19] The individual fixed effects regressions in Table 3.A5 suggest that this effect is mainly driven by stakeholders.

[20] 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.

3.5.1 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.[21] 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 $E\big(U(y_p, \pi(e_h(w_h), e_l(w_l), y_h(e_h(w_h)), y_l(e_l(w_l)), w_h, w_l)\big)$, where the principal's income is denoted by y_p and π 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 and Le Lec, 2010), because agents are not identical to begin with (different ability levels leading to different effort levels), which implies ex post inequality even when both agents are treated equally with equal piece rates.

3.5.2 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.

$$\begin{aligned} E(U) = E\Big(y_p + \gamma\pi + \big[\big(\alpha * \mathbb{1}(w_h \leq w_l) + \beta * \mathbb{1}(w_h > w_l)\big) y_l \\ - \big(\alpha * \mathbb{1}(w_h \leq w_l) + \beta * \mathbb{1}(w_h > w_l)\big) y_h\big]\Big) \end{aligned} \tag{3.1}$$

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

[21] 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.

- α 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.
- β 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.[22]

We can identify several cases:

1. **Output oriented** $\alpha = \mathbf{0}, \beta = \mathbf{0}$: This principal only cares about the maximization of output. The way income is distributed among workers is irrelevant.
2. **Equal procedure** $\alpha < \mathbf{0}, \beta > \mathbf{0}$: 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** $\alpha > \mathbf{0}, \beta > \mathbf{0}$: 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.[23]
4. **Inequality-targeting** $\alpha = \mathbf{0}, \beta > \mathbf{0}$: 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** $\alpha < \mathbf{0}, \beta < \mathbf{0}$: 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 3.4.3, we assume that preferences are defined over distributional outcomes, i.e. the workers' expected income.

[22] 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.

[23] Our experiment does not include cases where the low-ability worker receives *higher* ex post earnings than the high-ability worker.

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 α and β, 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 α and β. 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 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 γ We use our treatment variation to identify γ. Spectators' own income was kept constant but agents' joint output varied, while both dimensions were varied for stakeholders. The parameter γ 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.

$$\begin{aligned} E(U) = E\Big(\pi + \big[&(\alpha * \mathbb{1}(w_h \leq w_l) + \beta * \mathbb{1}(w_h > w_l))y_l \\ &- (\alpha * \mathbb{1}(w_h \leq w_l) + \beta * \mathbb{1}(w_h > w_l))y_h\big]\Big) \end{aligned} \tag{3.2}$$

In this case, the joint output π 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 α and β.

Table 3.4 – Results from a pooled specification

	Model based on equation (3.1)	Model based on equation (3.2)		
	(1) Interaction	(2) All	(3) Stakeholder	(4) Spectator
γ	0.311**	–	–	–
	(0.103)			
α	0.07	.095	.05	.21
	(0.055)	(0.04)	(0.09)	(.18)
β	0.17***	0.26***	0.1*	.55***
	(0.035)	(0.04)	(.05)	(0.13)
σ	0.08***	0.05***	.06***	.033***
	(0.013)	(0.005)	(0.012)	(0.008)
N	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 (3.1); column (2) reports parameters from equation (3.2). Columns (3) and (4) use the model based on equation (3.2) 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.

3.5.3 Pooled results

Table 3.4 focuses on average results for the entire population. Column (1) presents results from a conditional logit model that fits equation (3.1). The results mirror the results from Table 3.3, where we ran similar regressions but without assuming any underlying utility function. γ is significantly different from 0 ($p < 0.01$) but also significantly smaller than 1 (t-test, $p < 0.01$). The fact that γ 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 γ is above 0 shows that Spectators still care about output for intrinsic motives.

We can also see that, on average, α is not significantly greater than zero but β is ($p < 0.01$). This corresponds to the behavior outlined in case 4 (inequality-targeting principals). The non-significant α 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 (3.2) 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 β 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. α is non-significant for both samples, but the point estimate is larger for spectators. The point estimate is estimated relatively imprecisely.

3.5.4 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 efficiency 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 efficiency 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 (3.2) 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 (3.1) to be the underlying function because we would need within-principal variations in incentives.[24] 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 3.C, we detail the derivation of the type-specific conditional density $f_k(\theta, \sigma | X_1, X_2, Choice)$ following McFadden (1973). θ is the vector of parameters in the utility function (3.1), σ 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 θ and σ. It posits that the population can be categorized into K preference types, where each type has a distinct parameter vector (θ_k, σ_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

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

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

[24] 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.

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

$$ll(p_2, ..., p_K, \theta, \sigma | X_1, X_2, Choice) = \sum_{i=1}^{N} \log \Big(\sum_{k=1}^{K} p_k f_i(\theta_k, \sigma_k | X_1, X_2, Choice) \Big) \tag{3.3}$$

In our estimation of type-specific parameters of the utility function (3.2), we are interested 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.[25]

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.

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

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

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, 2004, 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

[25] 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.

models because gradient-based algorithms tend to suffer from convergence problems due to the non-linearity of the likelihood function. We follow McLachlan and Peel (2004), p.64, in their procedure for calculating standard errors by bootstrapping them parametrically using 1000 iterations and clustering at the individual level.

3.5.5 Results from the finite mixture model

The FMM characterizes heterogeneity using the value function specified in equation (3.2). 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 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.[26] Figure 3.A1 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 3.5.

Table 3.5 – Results from the finite mixture model with three classes

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

Bootstrapped 95% confidence intervals in squared brackets clustered at the individual level using 1000 iteration (McLachlan and Peel, 2004, 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 3.5.2. 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 3.B1 replicates this table using the full sample.

The model yields three classes that can be easily interpreted. The first class attaches no importance

[26] Results for the specifications with K=2 or K=4 are available on request.

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.[27]

The second class of subjects (Intermediate type) has a positive and significant β. 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. α 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 framework, this boils down to a model in which the principal has strong redistributive concerns and wants to minimize inequality as far as possible.[28]

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.

3.5.6 Counterfactual analyses

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.[29] Agents hold a power-cost function (see e.g.

[27] We can make this statement because none of the Spectators are sorted into this group.

[28] 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.

[29] 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 $w_h = w_l \longrightarrow \infty$ 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.

DellaVigna and Pope, 2018) $c(e) = \frac{ke^{1+s}}{1+s}$, 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.[30]

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.

Including social comparisons by agents

Social comparisons among agents matter in the field (Breza et al., 2017; 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 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 performs 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.[31]

We follow the general framework laid out by Breza et al. (2017). 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

$$V(y_i, y_R, w_i, w_r, e_i) = y_i(e_i, w_i) + M(w_i, w_R, y_i(e_i, w_i), y_R(e_R, w_R))e_i \tag{3.4}$$

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. $M(.)$ 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 $M(.)$ 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:

[30] 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.

[31] 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.

$$M(w_i, w_R) = -\alpha_a(w_R - w_i|w_i < w_R) - \beta_a(w_i - w_R|w_R \leq w_i) \tag{3.5}$$

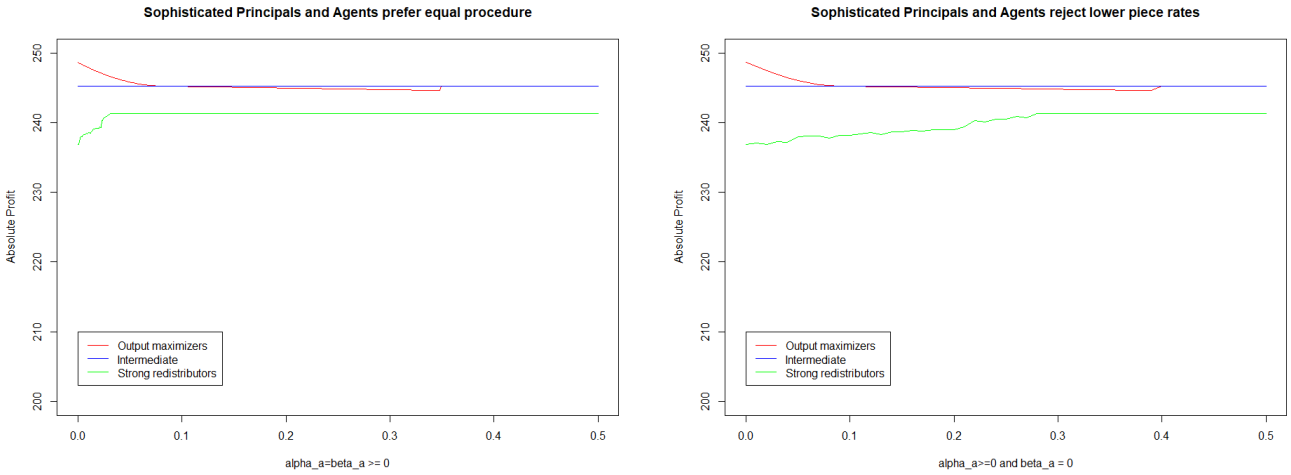
For a given level of α_a and β_a they will reduce their effort level if the dispersion of piece rates become too high.[32] Indeed their optimal effort decision becomes:

$$e = \left[\frac{w_i - \alpha_a(w_R - w_i|w_i < w_R) - \beta_a(w_i - w_R|w_R \leq w_i)}{k}\right]^{1/s_i} \tag{3.6}$$

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 α_a and β_a we have one 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 3.5 – 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 (3.5). The y-axis displays absolute profit. The x-axis displays variation in α_a and β_a. Figure 3.5a simulates agents with $\alpha_a = \beta_a \geq 0$, while Figure 3.5b simulates agents with $\alpha_a \geq 0$ and $\beta_a = 0$. Principals correctly anticipate their agents' behavior.

Sophisticated Principals Figure 3.5 plots the profits associated with each distributive type for different values of α_a and β_b. Here, we make the assumption that principals correctly anticipate the agents' social preferences. Figure 3.5a 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

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

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 α_a and β_a the gap in profits between intermediates and output maximizers shrinks as α_a and β_a increase. For high levels of α_a and β_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 α_a and β_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 α_a and β_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.

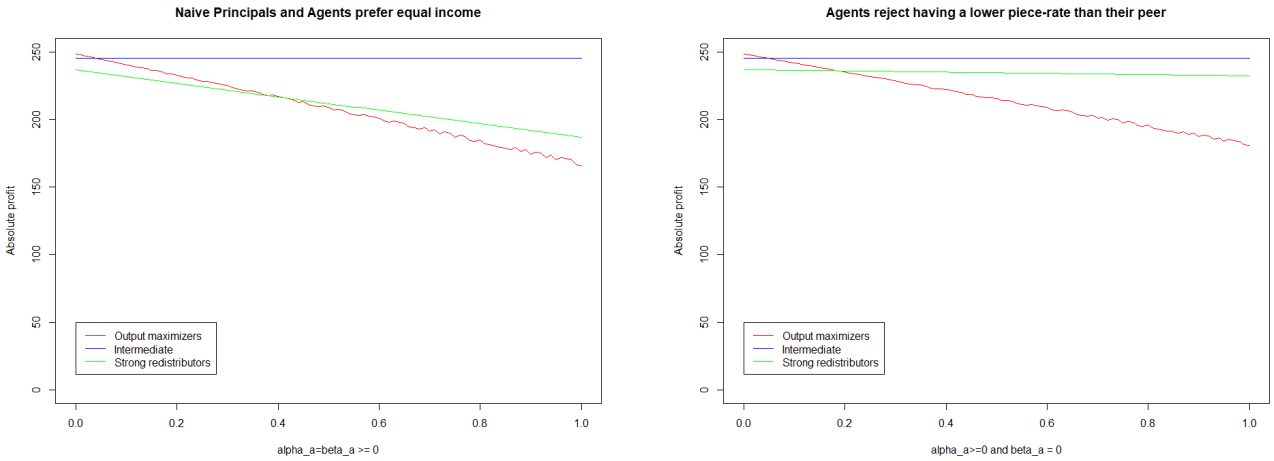
Figure 3.5b examines the case in which agents dislike receiving a lower piece rate than their co-worker but do not mind receiving a higher piece rate. In this case, convergence is slower. This stems from the fact that the *high-ability* agent does not reject this contract whereas the low-ability agent 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 3.6 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 3.5. Figure 3.6a 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.

Turning to Figure 3.6b, 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

Figure 3.6 – 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 (3.5). The y-axis displays absolute profit. The x-axis displays the variation in α_a and β_a. Figure 3.6a simulates agents with $\alpha_a = \beta_a \geq 0$, while Figure 3.6b simulates agents with $\alpha_a \geq 0$ and $\beta_a = 0$. Principals believe that agents do not hold any social preferences.

$(y_i(w_i, e_i^*), y_R(w_R, e_R^*))$:[33]

$$M(w_i, w_R) = -\alpha_a(y_R(w_R, e_R^*) - y_i(w_i, e_i^*)|y_i^* < y_R^*) - \beta_a(y_i(w_i, e_i^*) - y_R(w_R, e_R^*)|y_R^* \leq y_i^*) \quad (3.7)$$

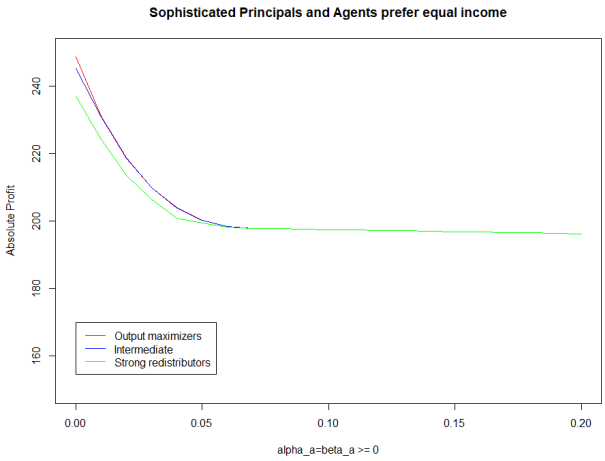
For a given level of α_a and β_a agents will reduce their effort level if the dispersion of *potential* income becomes too high. Indeed their optimal effort decision becomes

$$e_i = \left[\frac{w_i - \alpha_a(y_R(w_R, e_R^*) - y_R(w_i, e_i^*)|y_i^* < y_R^*) - \beta_a(y_i(w_i, e_i^*) - y_R(w_R, e_R^*)|y_R^* \leq y_i^*)}{k}\right]^{1/s_i} \quad (3.8)$$

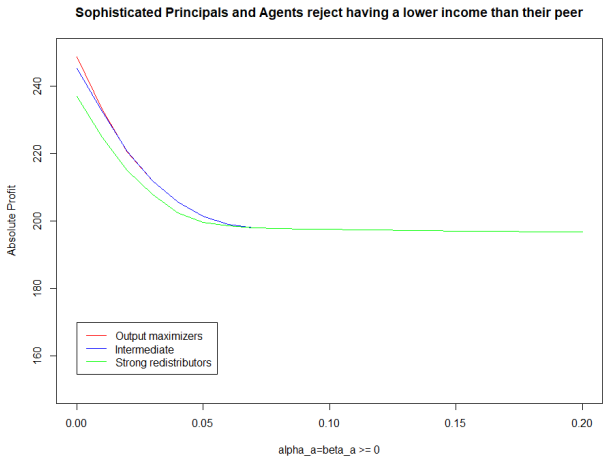
We will further examine principals' decisions when they correctly anticipate the agents' behavior, and when they are naive about agents' other-regarding preferences.

[33] 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 3.7 – Firm's relative profit if agents care about potential ex-post income



(a) Agents reject inequality in potential income

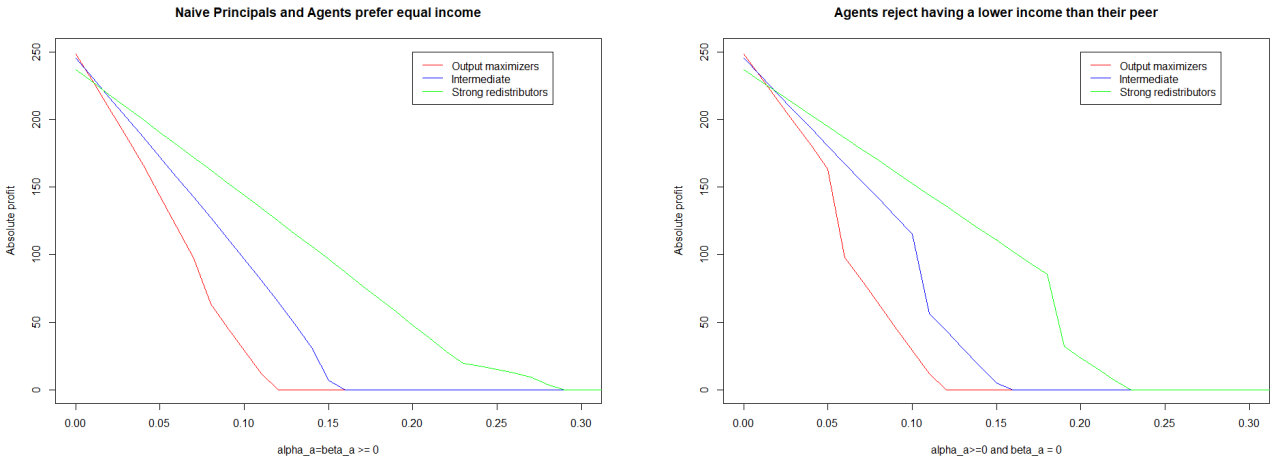


(b) Agents dislike having lower potential income and principal anticipates correctly

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 (3.7). The y-axis displays the absolute profit.The x-axis displays the variation in α_a and β_a. Figure 3.7a simulates agents with $\alpha_a = \beta_a \geq 0$, while Figure 3.7b simulates agents with $\alpha_a \geq 0$ and $\beta_a = 0$. Principals correctly anticipate their agents' behavior.

Sophisticated Principals For each distributive type, Figure 3.7 plots the absolute profits for different values of α_a and β_b, now assuming that agents hold preferences as in equation (3.7). Figure 3.7a 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 α_a and β_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 3.7b, 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 3.8 – Firm's relative profit if agents care about potential ex post income



(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 (3.7). The y-axis displays absolute profit.The x-axis displays the variation in α_a and β_a. Figure 3.8a simulates agents with $\alpha_a = \beta_a \geq 0$, while Figure 3.8b simulates agents with $\alpha_a \geq 0$ and $\beta_a = 0$. Principals believe that agents do not hold any social preferences.

Naive Principals Figure 3.8 simulates profits for naive principals and inequality-averse agents. Figure 3.8a 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.[34] Hence, agents with high α_a and α_b will eventually retaliate in response to even a small gap in potential income.

Figure 3.8b 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 α_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. However, 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

[34]The egalitarian principals that we identified in our data are not "perfect" egalitarians and still face a residual trade-off.

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.

3.6 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 highly encouraging 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.

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.

Bibliography

Abeler, Johannes, Steffen Altmann, Sebastian Kube, and Matthia Wibral. Gift exchange and workers' fairness concerns: When equality is unfair. *Journal of the European Economic Association*, 8(6): 1299–1324, 2010. ISSN 15424766. doi: 10.1162/jeea_a_00026.

Alekseev, Aleksandr, Gary Charness, and Uri Gneezy. Experimental methods: When and why contextual instructions are important. *Journal of Economic Behavior and Organization*, 2017. ISSN 01672681. doi: 10.1016/j.jebo.2016.12.005.

Almås, Ingvild, Alexander W Cappelen, and Bertil Tungodden. Cutthroat capitalism versus cuddly socialism: Are Americans more meritocratic and efficiency-seeking than Scandinavians? *Mimeo*, 2018.

Ashraf, Nava and Oriana Bandiera. Social Incentives in Organizations. *Annual Review of Economics*, 2018.

Balafoutas, Loukas, Martin G. Kocher, Louis Putterman, and Matthias Sutter. Equality, equity and incentives: An experiment. *European Economic Review*, 2013. ISSN 00142921. doi: 10.1016/j.euroecorev.2013.01.005.

Bandiera, O, I Barankay, and I Rasul. Incentives for Managers and Inequality among Workers: Evidence from a Firm-Level Experiment. *The Quarterly Journal of Economics*, 122(2):729–773, 2007. ISSN 0033-5533. doi: 10.1162/qjec.122.2.729. URL https://academic.oup.com/qje/article-lookup/doi/10.1162/qjec.122.2.729.

Bandiera, Oriana, Iwan Barankay, and Imran Rasul. Social Preferences and the Response to Incentives: Evidence from Personnel Data. *Quarterly Journal of Economics*, 120(3):917–962, 2005.

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*, 2011. ISSN 10586407. doi: 10.1111/j.1530-9134.2011.00310.x.

Bellemare, Charles and Bruce Shearer. Gift giving and worker productivity: Evidence from a firm-level experiment. *Games and Economic Behavior*, 2009. ISSN 08998256. doi: 10.1016/j.geb.2008.12.001.

Bertrand, Marianne and Antoinette Schoar. Managing with style: The effect of managers on firm policies. *Quarterly Journal of Economics*, 2003. ISSN 00335533. doi: 10.1162/003355303322552775.

Bloom, Nicholas, Renata Lemos, Raffaella Sadun, Daniela Scur, and John Van Reenen. Jeea-fbbva 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. Distributional concerns in managers' compensation schemes for heterogeneous workers: Experimental evidence. 2019.

Breza, Emily, Supreet Kaur, and Yogita Shamdasani. The morale effects of pay inequality. *The Quarterly Journal of Economics*, 2017.

Brock, J Michelle, Andreas Lange, and Erkut Y Ozbay. Dictating the risk: Experimental evidence on giving in risky environments. *American Economic Review*, 103(1):415–37, 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*, 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. ISSN 00028282. doi: 10.1257/aer.100.5.2261.

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*, 2007. ISSN 0734-306X. doi: 10.1086/519540.

Clark, Andrew E, David Masclet, and Marie Claire Villeval. Effort and comparison income: Experimental and survey evidence. *ILR 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.

Cronqvist, Henrik and Frank Yu. Shaped by their daughters: Executives, female socialization, and corporate social responsibility. *Journal of Financial Economics*, 2017. ISSN 0304405X. doi: 10.1016/j.jfineco.2017.09.003.

Croson, Rachel and Uri Gneezy. Gender differences in preferences. *Journal of Economic literature*, 47 (2):448–474, 2009.

DellaVigna, Stefano and Devin Pope. What motivates effort? evidence and expert forecasts. *The Review of Economic Studies*, 85(2):1029–1069, 2018.

DellaVigna, Stefano, John A List, Ulrike Malmendier, and Gautam Rao. Estimating social preferences and gift exchange at work. *Mimeo*, 2016.

Eisenkopf, Gerald and Sabrina Teyssier. Envy and loss aversion in tournaments. *Journal of Economic Psychology*, 34:240–255, 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*, 1993. ISSN 0033-5533. doi: 10.2307/2118338.

Fehr, Ernst and Klaus M Schmidt. Fairness and incentives in a multi-task principal–agent model. *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. ISSN 13864157. doi: 10.1007/s10683-006-9159-4.

Fisman, Raymond, Shachar Kariv, and Daniel Markovits. Individual preferences for giving. *American Economic Review*, 2007. ISSN 00028282. doi: 10.1257/aer.97.5.1858.

Gagnon, Nickolas, Kristof Bosmans, and Arno Riedl. The effect of unfair chances and gender discrimination on labor supply. *Available at SSRN 3519540*, 2020.

Greiner, Ben. An online recruitment system for economic experiments. 2004.

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. ISSN 01672681. doi: 10.1016/j.jebo.2015.06.009.

Kocher, Martin G., Ganna Pogrebna, and Matthias Sutter. Other-regarding preferences and management styles. *Journal of Economic Behavior and Organization*, 2013. ISSN 01672681. doi: 10.1016/j.jebo.2013.01.004.

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

Krawczyk, Michal and Fabrice Le Lec. 'give me a chance!'an experiment in social decision under risk. *Experimental economics*, 13(4):500–511, 2010.

McFadden, D. Conditional logit analysis of qualitative choice behaviour. In Zarembka, P., editor, *Frontiers in Econometrics*, pages 105–142. Academic Press New York, New York, NY, USA, 1973.

McLachlan, Geoffrey and David Peel. *Finite mixture models*. John Wiley & Sons, 2004.

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. ISSN 01672681. doi: 10.1016/j.jebo.2017.12.007. URL http://dx.doi.org/10.1016/j.jebo.2017.12.007.

von Siemens, Ferdinand A. Heterogeneous social preferences, screening, and employment contracts. *Oxford Economic Papers*, 63(3):499–522, 2010.

Appendices

3.A Tables and Figures

Table 3.A1 – Description of the main variables used in the REPONSE survey

Variable name	Original question in the survey	Scale used in the analysis
Dependent variables		
White-collar individualized wage raise	*Did white-collar workers benefited from individualized wage raises*	0 = No; 1 = Yes
White-collar performance-based bonus	*Did white-collar workers benefited from bonuses related to individual performance?*	0 = No; 1 = Yes
Blue-collar individualized wage raise	*Did non-white collar workers benefited from individualized wage raises*	0 = No; 1 = Yes
Blue-collar performance-based bonus	*Did non-white collar workers benefited from bonuses related to individual performance?*	0 = No; 1 = Yes
Main explanatory variables	*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*	
Individualized wage raises are unfair.	*They are fairer than undifferentiated increases. 1= Completely agree; 2=Somewhat agree; 3=Somewhat disagree; 4=Completely disagree*	1=Disagree ; 0 = Agree
Individualized wage raises create tension	*They create tensions that could undermine collective functioning. 1= Completely agree; 2=Somewhat agree; 3=Somewhat disagree; 4=Completely disagree*	1=Disagree ; 0 = Agree
Individualized wage raises are subjective	*They cannot be based on objective criteria. 1= Completely agree; 2=Somewhat agree; 3=Somewhat disagree; 4=Completely disagree*	1=Disagree ; 0 = Agree
Individualized wage raises motivate	*They motivate employees. 1= Completely agree; 2=Somewhat agree; 3=Somewhat disagree; 4=Completely disagree*	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).[35]
- **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%).[36]

Table 3.A2 – Performance-based bonuses and managers' distributive preferences

	White-collar workers			Blue-collar workers		
	(1)	(2)	(3)	(4)	(5)	(6)
Dep var = Did white/blue-collar workers benefited from bonuses based on individual performance?						
Individualized wage raises are unfair	-0.181***	-0.111***	-0.0834***	-0.131***	-0.0810***	-0.0606***
	(0.0135)	(0.0149)	(0.0160)	(0.0150)	(0.0163)	(0.0182)
Individualized wage raises create tension		-0.110***	-0.0613***		-0.0624***	-0.0445***
		(0.0110)	(0.0119)		(0.0117)	(0.0133)
Individualized wage raises motivate		0.0382**	0.0104		0.0421**	0.0298
		(0.0181)	(0.0190)		(0.0194)	(0.0213)
Individualized wage raises are subjective		-0.0940***	-0.0489***		-0.0712***	-0.0775***
		(0.0128)	(0.0136)		(0.0139)	(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 4.C1 for a description of all the variables.

[35] 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.

[36] 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 3.A3 – Summary statistics Agents vs principal

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

Table 3.A4 – Summary statistics principals Spectator vs Stakeholder

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

Table 3.A5 – Regressions that characterize contract decisions using belief-based trade-offs and individual fixed effects

	Stakeholders		Spectators	
	(1)	(2)	(3)	(4)
Dependent variable: Contract 2 (high inequality) was chosen				
A is equal piece rate	-0.0245	0.0204	-0.0310	0.000299
	(0.0343)	(0.0363)	(0.0451)	(0.0444)
$\frac{\Delta(\text{Expected Output 2 and 1})}{10}$	0.0700***	0.0674***	0.0299***	0.0283***
	(0.00854)	(0.00836)	(0.00794)	(0.00774)
$\frac{\Delta(\text{Expected Inequality 2 and 1})}{10}$	-0.0334***	-0.0194*	-0.0241**	-0.0150
	(0.00887)	(0.00997)	(0.0112)	(0.0115)
Choice 1 = 1		0.131**		0.107
		(0.0566)		(0.0644)
Choice 13 = 1		0.114**		0.0658
		(0.0545)		(0.0614)
Constant	0.609***	0.528***	0.410***	0.356***
	(0.0413)	(0.0495)	(0.0508)	(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. All columns include individual fixed effects.

Table 3.A6 – Regressions that characterize contract decisions

	Theoretical trade-offs		Belief-based trade-offs	
	(1)	(2)	(3)	(4)
Dependent variable: Contract 2 (high inequality) was chosen				
Stakeholder	0.190**	0.190**	0.228***	0.208***
	(0.0830)	(0.0831)	(0.0632)	(0.0675)
$\frac{\Delta(\text{Output 2 and 1})}{10}$	0.0452***	0.0419***	0.0194***	0.0180***
	(0.0127)	(0.0129)	(0.00655)	(0.00651)
$\frac{\Delta(\text{Output 2 and 1})}{10}$ * Stakeholder	0.0317*	0.0317*	0.0148	0.0151
	(0.0167)	(0.0167)	(0.0111)	(0.0109)
$\frac{\Delta(\text{Inequality 2 and 1})}{10}$	-0.0468***	-0.0356**	-0.0191**	-0.0112
	(0.0134)	(0.0138)	(0.00921)	(0.00954)
$\frac{\Delta(\text{Inequality 2 and 1})}{10}$ * Stakeholder	-0.00727	-0.00727	-0.00268	0.00182
	(0.0180)	(0.0180)	(0.00998)	(0.0111)
A is equal piece rate	-0.0876**	-0.0493	-0.0226	0.0118
	(0.0437)	(0.0436)	(0.0425)	(0.0429)
A is equal piece rate * Stakeholder	0.0105	0.0105	0.0283	0.0337
	(0.0621)	(0.0621)	(0.0519)	(0.0531)
Choice 1 = 1		0.117***		0.122**
		(0.0432)		(0.0497)
Choice 13 = 1		0.0673		0.110**
		(0.0415)		(0.0549)
Constant	0.483***	0.419***	0.433***	0.379***
	(0.0856)	(0.0855)	(0.0836)	(0.0864)
Control variables	Yes	Yes	Yes	Yes
Observations	1808	1808	1808	1808
R^2	0.100	0.103	0.121	0.125

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 sample includes all observations. 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) the difference variables are computed using principals' beliefs. Columns (2) and (4) add controls for whether the current observation is about Choice 1 (Choice 1 = 1) or Choice 13 (Choice 13 = 1). 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.

Table 3.A7 – Regressions that characterize contract decisions

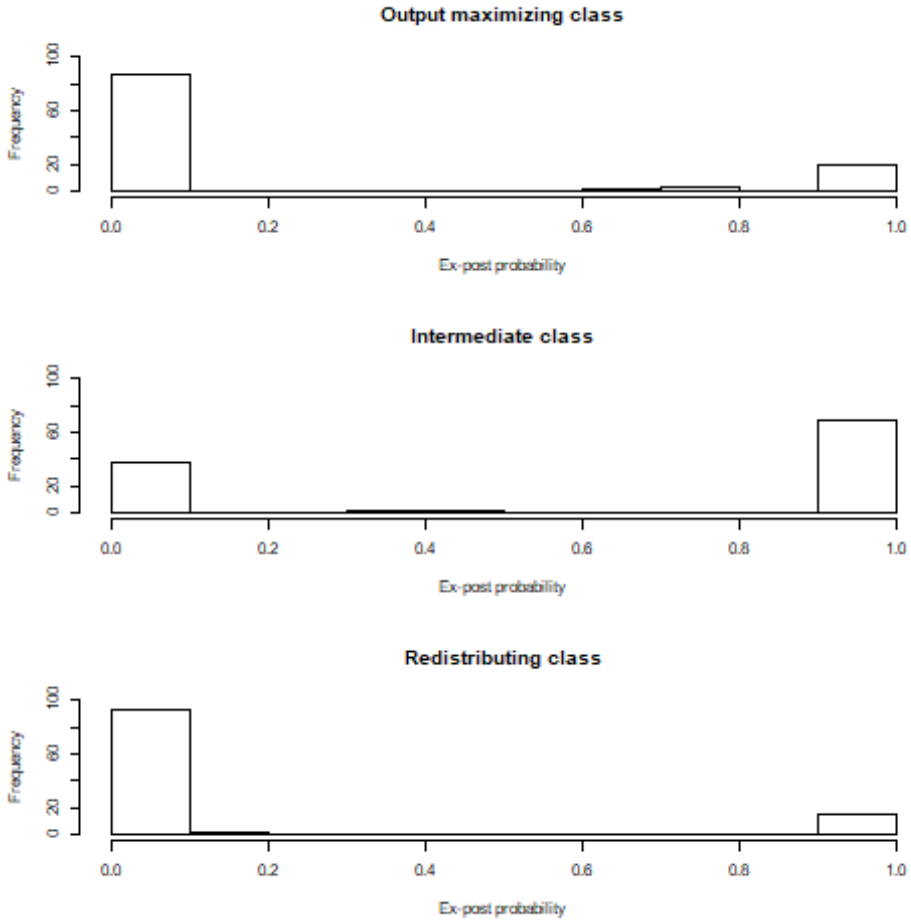
	Theoretical trade-offs		Belief-based trade-offs	
	(1)	(2)	(3)	(4)
Dependent variable: Contract 2 (high inequality) was chosen				
Stakeholder	0.230**	0.230**	0.247***	0.247***
	(0.0909)	(0.0909)	(0.0807)	(0.0813)
$\frac{\Delta(\text{Output 2 and 1})}{10}$	0.0417***	0.0422***	0.0320***	0.0308***
	(0.0127)	(0.0128)	(0.00674)	(0.00661)
$\frac{\Delta(\text{Output 2 and 1})}{10}$ * Stakeholder	0.00721	0.00720	0.0317***	0.0317***
	(0.0170)	(0.0170)	(0.0114)	(0.0113)
$\frac{\Delta(\text{Inequality 2 and 1})}{10}$	-0.0499***	-0.0480***	-0.0234**	-0.0159
	(0.0138)	(0.0147)	(0.0109)	(0.0108)
$\frac{\Delta(\text{Inequality 2 and 1})}{10}$ * Stakeholder	0.00584	0.00572	-0.0111	-0.0110
	(0.0180)	(0.0180)	(0.0150)	(0.0151)
A is equal piece rate	-0.0975**	-0.0888*	-0.0241	0.00373
	(0.0458)	(0.0463)	(0.0435)	(0.0430)
A is equal piece rate * Stakeholder	-0.00454	-0.00507	0.00103	0.000447
	(0.0640)	(0.0641)	(0.0560)	(0.0562)
Choice 1 = 1		0.0506		0.0955**
		(0.0424)		(0.0474)
Choice 13 = 1		-0.00848		0.0704
		(0.0473)		(0.0507)
Constant	0.514***	0.498***	0.429***	0.382***
	(0.0907)	(0.0922)	(0.0835)	(0.0837)
Control variables	Yes	Yes	Yes	Yes
Observations	1532	1532	1620	1620
R^2	0.103	0.104	0.171	0.173

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 about Choice 1 (Choice 1 = 1) or Choice 13 (Choice 13 = 1). 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 3.A1 – Distribution of individual ex post probabilities to be part of a given class



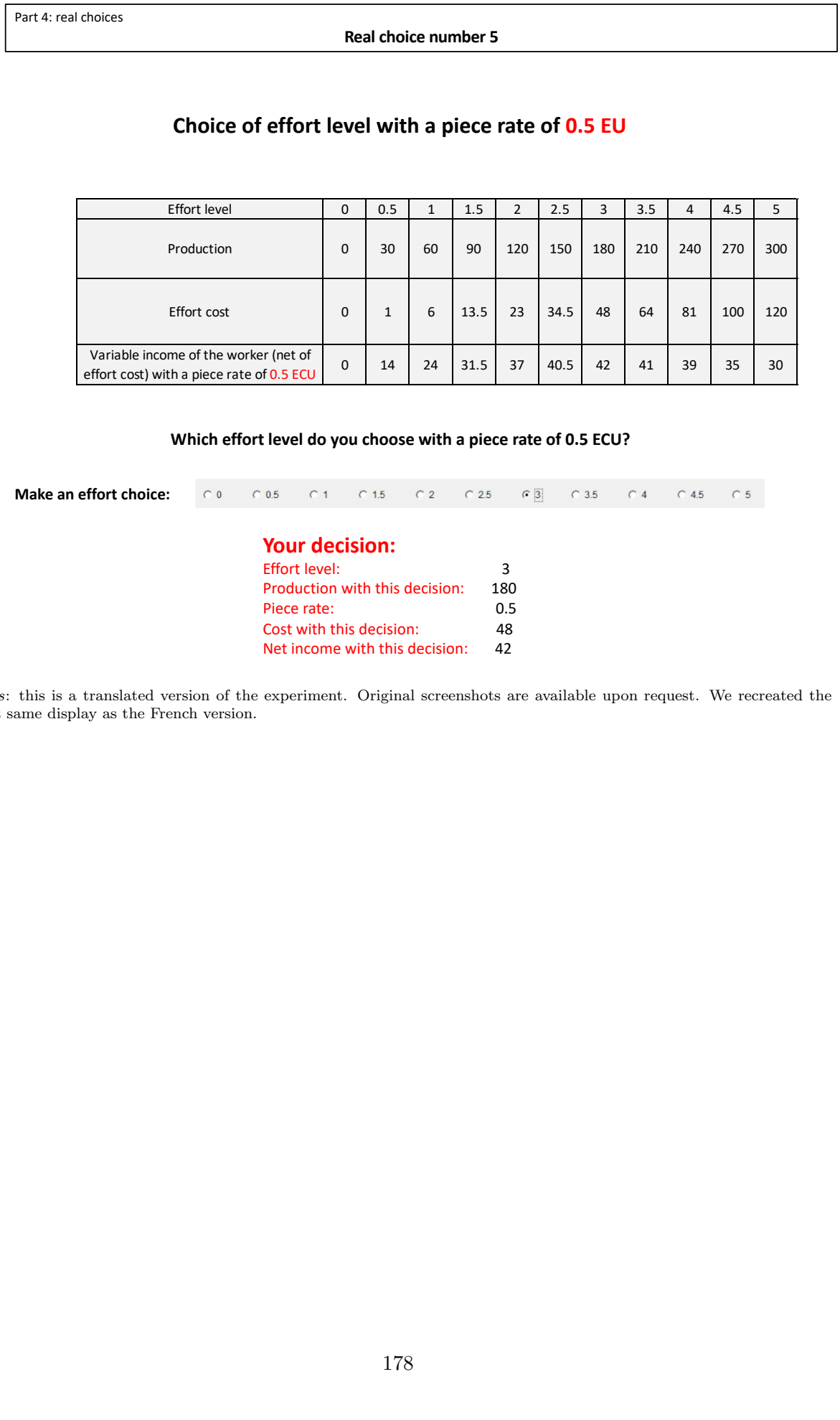
Notes: These histograms characterize the distribution of ex post class probabilities on the individual level. The x-axis characterizes the ex post probability and the y axis reports the number of subjects within a bin. The first histograms shows the distribution of the ex post probability to be in the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

Figure 3.A2 – 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 3.A3 – 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?

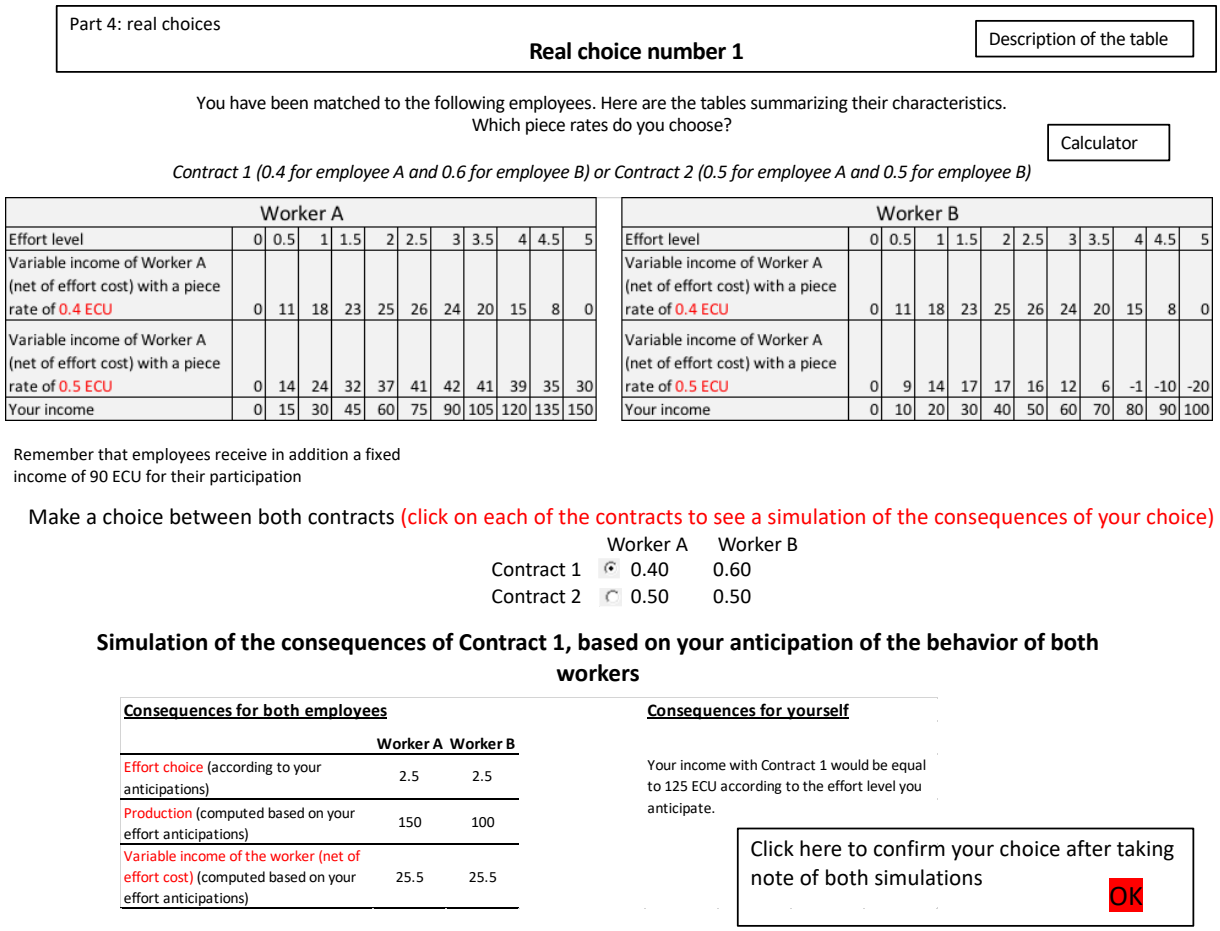
Make an effort choice: ○ 0 ○ 0.5 ○ 1 ○ 1.5 ○ 2 ○ 2.5 ◉ 3 ○ 3.5 ○ 4 ○ 4.5 ○ 5

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 3.A4 – 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 0.4 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	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)

	Worker A	Worker B
Contract 1	0.40	0.60
Contract 2	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

	Worker A	Worker B
Effort choice (according to your anticipations)	2.5	2.5
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

Consequences for yourself

Your income with Contract 1 would be equal to 125 ECU according to the effort level you anticipate.

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.

3.B Robustness Checks

Table 3.B1 – Results from the finite mixture model with three classes including outlier beliefs

	Output maximizers	Intermediate	Strong redistributors
	(1)	(2)	(3)
Parameters			
β	0.00	0.36	10.38
	[0.07]	[0.06]	[2.07]
σ	0.28	0.03	0.003
	[0.07]	[0.012]	[0.04]
Shares			
Full sample	0.22	0.63	0.145
if Stakeholder	0.36	0.44	0.19
	[0.02]	[0.04]	[0.03]
if Spectator	0.00	0.47	0.53
	[0.01]	[0.07]	[0.07]

Standard errors from gradient based estimation in parentheses.
Bootstrapped standard errors (1000 iterations) in squared brackets following (**?**, p.64) .
This table presents results from a finite mixture model outlined in section 3.5.2. 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 3.B1 replicates Table 3.5 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 σ. Furtermore, 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.

3.C Deriving the density of the likelihood on the individual level

The model assumes that utility has a deterministic (u) component characterized by equation (3.1) and a random component (ε) that is stochastic.

$$U^c(X_c, \theta, \sigma) = u(x_c, \theta) + \varepsilon_{X_c} \tag{3.9}$$

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 θ be a vector of parameters e.g. (γ, β, σ) if we fit the model posited in equation (3.1); ε 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:

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

If σ 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

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

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:

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

3.D 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.

3.D.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 3.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 3.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 3.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.

3.D.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 3.D4 – Effort-production 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

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 3.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. You 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 3.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.

3.E 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 3.A3). 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 3.A4). 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.

3.E.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 3.E1 – Effort-production-cost table *[Stakeholder treatment]*

Employé A											
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 revenu	0.0	12.5	25.0	37.5	50.0	62.5	75.0	87.5	100.0	112.5	125.0

Employé B											
Niveau d'effort	0	0.5	1	1.5	2	2.5	3	3.5	4	4.5	5
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	20.0	20.0	31.3	45.0	61.3	80.0	101.3	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 3.E2 – Effort-production-cost table *[Spectator treatment]*

Employé A											
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

Employé B											
Niveau d'effort	0	0.5	1	1.5	2	2.5	3	3.5	4	4.5	5
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	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 3.E3 – Effort-production-cost table *[Stakeholder treatment]*

Employé A

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 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
Votre revenu	0.0	12.5	25.0	37.5	50.0	62.5	75.0	87.5	100.0	112.5	125.0

Employé B

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 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 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	6.3	12.5	18.8	25.0	31.3	37.5	43.8	50.0	56.3	62.5

Figure 3.E4 – Effort-production-cost table *[Spectator treatment]*

Employé A

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 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

Employé B

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 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 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.

3.E.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 3.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 3.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 3.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.

3.E.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.

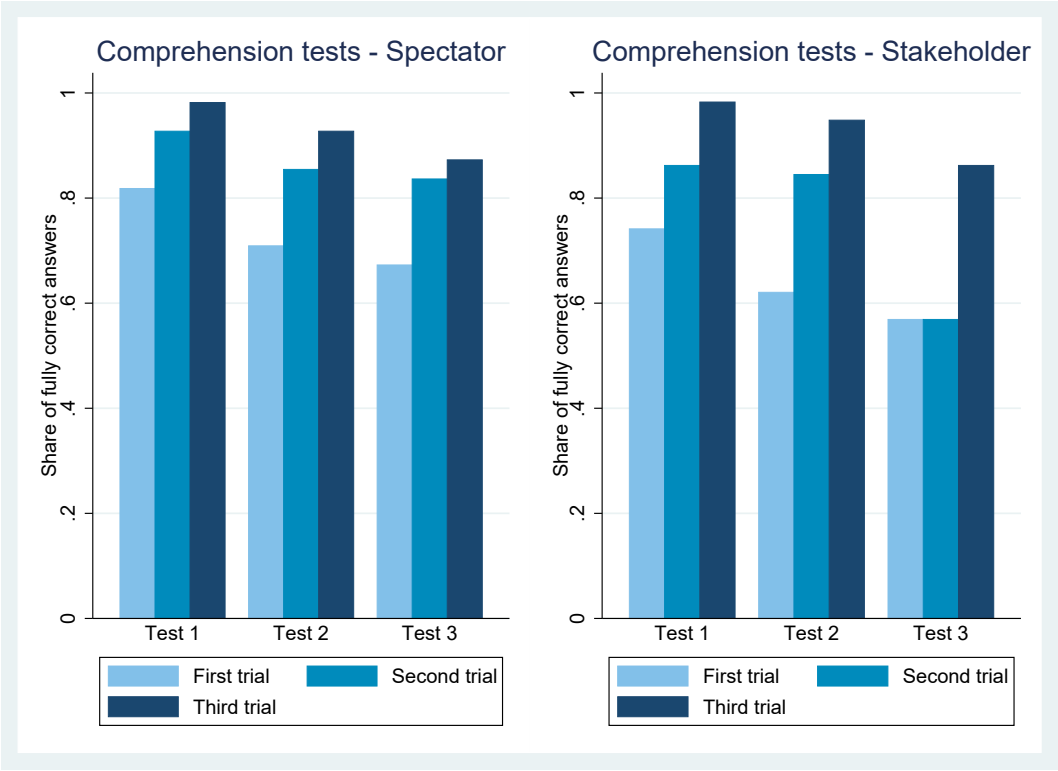


Figure 3.E8 – Principals' comprehension tests

Notes:

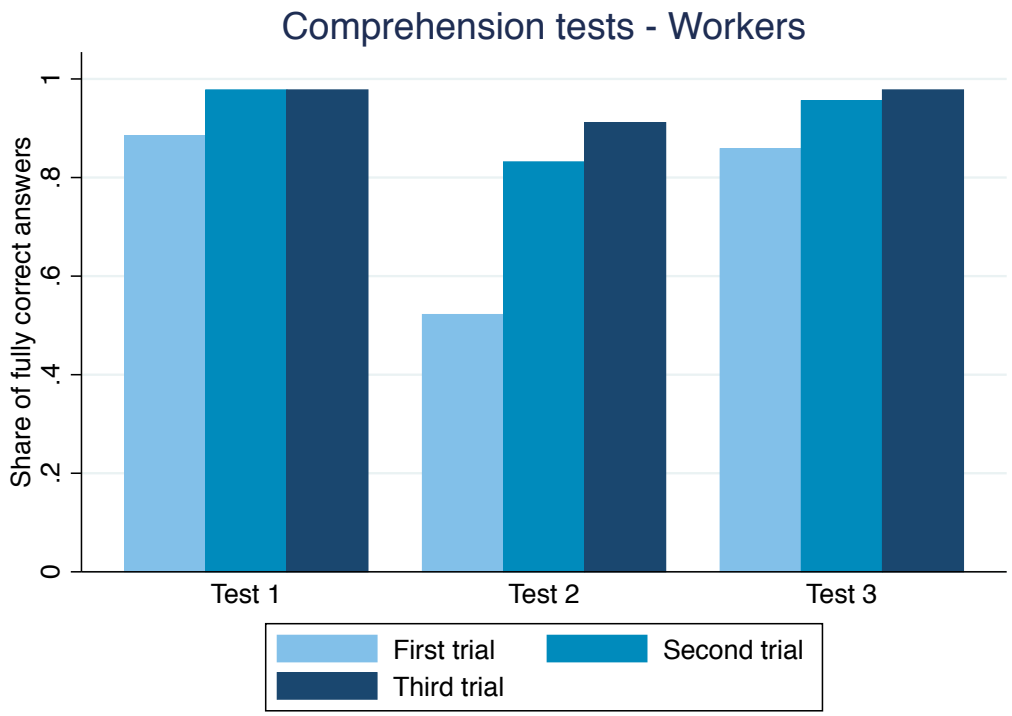


Figure 3.E9 – Workers' comprehension tests

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 3.E1 – True-False average score

Variable	Obs	Mean	Std. Dev.	Min	Man
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.

3.F Aptitude test

Translated from French to English by the Authors.

3.F.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?[37]

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.[38]
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?[39]

1. **Remarkable**
2. Immediate
3. Indiscreet
4. Boaster

[37]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.

[38]Crystal-clear would be the correct translation but then this sentence would be an intruder too

[39]In French immediate can be translated by "imminent" and thus many people are confused about the difference between "éminent" and "imminent"

3.F.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
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

3.F.3 General knowledge

Question 7: Simone Veil[40]

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

[40]Simone Veil was a judge but not a lawyer. She entered the French Panthéon in 2018

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

Chapter 4

Ethnic bias, economic success, and trust: findings from large sample experiments in Germany and the U.S.

This chapter is a joint work with Yann Algan, Gianluca Grimalda, Fabrice Murtin, Louis Putterman, Ulrich Schmidt and Vincent Siegerink.

4.1 Introduction

Populist views are becoming more widespread and anti-establishment parties embracing xenophobic attitudes are receiving increased support in Western countries (Betz, 2018; Algan et al., 2019). Some hypothesize that a cultural shift leading to stronger nationalistic attachment and anti-immigration stances are at the basis of this reversal in public opinion. Such a cultural shift is partly the reaction to the previous spread of progressive ideas in Western culture since the late 1960s (Inglehart and Norris, 2016). This cultural backlash, in the words of Norris and Inglehart (2019), may have been amplified by larger migratory flows for either economic reasons or conflicts in the Middle East and Africa, which increased ethnic diversity. As a result, the demarcation between the "us" and the "them" has become more pronounced in the psychology of many individuals, particularly after widespread feeling of insecurity caused by the 2008 economic crisis in many cohorts of the population (Algan et al., 2017; Guriev, 2018).

Research in experimental social psychology has been studying the foundations of group identity and discrimination for a long time (Allport et al., 1954; Tajfel et al., 1971; Brewer, 1999). One of the main findings is that the tendency to discriminate in favor of those identified as belonging to the same group (the so-called "ingroup") at the expense of others identified as belonging to another group (the so-called "outgroup") is endemic (Balliet et al., 2014; Lane, 2016). Such an ingroup bias has been found in experiments conducted with groups differing for their nationality (Yamagishi et al., 2005; Guillen and Ji, 2011; Akai and Jiro, 2012; Whitt and Wilson, 2007; Romano et al., 2017; Dorrough and Glöckner, 2016), ethnicity (Fershtman and Gneezy, 2001; Fershtman et al., 2005; Bernhard et al., 2006; Simpson et al., 2007; Habyarimana et al., 2007; Ahmed and Ahmed, 2010; Burns, 2012; Tanaka and Camerer, 2016; Felfe et al., 2017; Zhang et al., 2019), religious affiliation (Chuah et al., 2013, 2014), castes (Fehr et al., 2008; Hoff et al., 2011), political groups (Rand et al., 2009; Weisel and Böhm, 2015), associations, communities, or army units within a country (Goette et al., 2006; Ruffle and Sosis, 2006; Degli Antoni and Grimalda, 2016) and also when groups differed for purely arbitrary characteristics induced in the laboratory (Tajfel et al., 1971; Charness et al., 2007; Chen and Li, 2009; Güth et al., 2009; Hargreaves and Zizzo, 2009). Evidence of ingroup bias is widespread even outside the laboratory (Bertrand and Mullainathan, 2004; Tjaden et al., 2018; Adida et al., 2010)[1]

Populism may be interpreted as a trigger making more salient ethnic demarcations and thus amplifying psychological propensity to favor people identified as one's ingroup. The purpose of this article is to use the tools from experimental economics to better understand the extent and the patterns of ethnic-based discrimination and to test whether discrimination may be reduced. We focus on two large Western countries, the US and Germany. While previous research typically focused on cross-national discrimination, or within-country discrimination between two ethnic groups, our use of large samples makes it possible to study discriminatory patterns between the ethnic majority, two specified minority groups, and a residual group, in both countries. In this way we can study whether discrimination is selective or treats other outgroups similarly. We quantify ingroup bias for both the ethnic majority and the ethnic minorities within both countries. We determine whether ingroup favoritism comes from accurate expectations of low trustworthiness or whether it is purely

[1] Criado et al. (2015); Johansson-Stenman et al. (2009) reached more mixed conclusions, while Goerg et al., 2016 find significant miscalibration of beliefs, but rarely in behavior, between three national groups. Some papers do not find any discriminatory patterns (Willinger et al., 2003; Bouckaert and Dhaene, 2004; Georgantzis et al., 2018; Goerg et al., 2016).

taste-based. We also test in a controlled way potential treatments to reduce such bias. Discrimination often hinges upon stereotypical beliefs that ethnic minorities do not share the same work ethic as the ethnic majority. People from ethnic minorities are often depicted as being lazy and taking advantage of welfare benefits (Gilens, 2009; Alesina et al., 2018a). We test whether releasing information that people from ethnic minorities were economically successful alleviates discriminatory attitudes from the ethnic majority.

To do so, we conduct a module on ethnic discrimination within the *Trustlab* platform, a large-scale multi-country incentivized online experiment designed to study social preferences, generalized trust and trust in institutions using experimental games (Murtin et al., 2018; Aassve et al., 2018a,b). The module was implemented in the US and in Germany on about 1000 subjects, representative of the national population of each country. The module consists of several trust games (TGs) involving pairs of players. Both receive an endowment of 10 dollars/euros. The first mover can transfer any fraction of this endowment to a second mover. The transferred amount is multiplied by 3 and the second mover can then return any amount out of this multiplied transfer and her own endowment to the first mover. Our key experimental manipulation is to disclose the second mover's ethnic group to the first mover. In the US, first movers from any ethnic groups are matched in random order with a non-Hispanic White (henceforth "White" for the sake of brevity), an African-American, and a Hispanic second mover. In Germany, first movers are matched in a similar fashion with a rooted German, a subject of Turkish descent and a subject with Eastern European origins. We follow Adida et al. (2014) and define a rooted German as a person who was born in Germany and whose parents were also born in Germany. We measure the prevalence of ingroup favoritism – also referred to as parochial attitudes (Romano et al., 2017) –, by comparing the first mover transfers across the different ethnic groups. In our experiment, discrimination coincides with ingroup favouritism. It is, in other words, the propensity to transfer larger sums to people from one's ingroup than to people from one's outgroup. Since we record the first mover ethnicity, we are able to study how favoritism varies based on the ethnic types, thereby making a distinction between the discriminatory behavior of the ethnic majority compared to ethnic minorities. We also study bias selectivity, i.e. whether first mover transfers depend on the type of outgroup second movers. Are subjects more biased against one outgroup compared to another? Is there an ethnic group that is discriminated against or favored by all groups, or are bias patterns completely ethnicity-specific?

The second part of the experiment tests whether information on second movers' income can alleviate ethnic ingroup bias. We run another round of TGs where first movers are now matched with rich second movers, whose incomes belong to the top 20% of the country's income distribution. We still varied the second mover's ethnicity. Ethnic majority participants are thus confronted with rich ethnic minorities, which contradicts the usual populist narratives picturing immigrants or minorities as idle welfare recipients. We also analyze how ethnic minorities react to being matched with rich people from their own or other ethnic minorities.

Overall, we find that members of all ethnic groups have a significant ingroup bias, except participants of Eastern European descent in Germany. This bias is particularly large for African Americans in the US, and rooted Germans and Turkish descent participants in Germany. We further show that ethnic discrimination is selective in Germany. Rooted Germans discriminate twice as much against Turkish descent participants as against those of Eastern European descent. On the contrary, Eastern European and Turkish descent first movers discriminate against each other, but trust rooted Germans

similarly to how much they trust their ingroup. In the US, the ethnic groups have a more homogeneous non-selective ingroup bias.

We are able to decompose ethnic discrimination into a taste-based and a statistical-based component using first movers' beliefs regarding second movers' transfer. Controlling for expected trustworthiness (expected transfer from second mover to first mover), we infer that 80% of the ingroup bias is driven by taste-based discrimination and 20% by statistical discrimination in both countries. We further show that low trustworthiness stereotypes are mostly inaccurate, except for those concerning Turkish descent second movers, who send back significantly less money to first movers than other groups.

Although participants of all ethnicities reduce transfers when the receiver is known to be rich, matching participants to rich second movers attenuates ethnic discrimination. Ethnic ingroup favoritism almost completely disappears except for African Americans and rooted German first movers who still favor their own ingroup, even if the second mover is rich, but to a much lower extent than when income information is not released. Moreover, we uncover the existence of a "deserving rich ethnic minority effect" in Germany. Rooted Germans discriminate less against rich Turkish second movers than against poor ones, suggesting that narratives of successful ethnic minorities could help changing stereotypes. However, we also show that this treatment can backfire and generate distrust within minority groups. In the US, we also observe a "deserving rich ethnic minority effect" but smaller in magnitude. White Americans had a smaller ingroup bias in the first place, so there was probably less room for intervention. In the US, the treatment does not generate any backlash within minorities. We also find that first movers belonging in the top 20% of income distribution display ingroup loyalty across income lines, as they transfer to fellow top 20% income earners more than first movers from the bottom 80% of the income distribution. This income ingroup bias is however significant only for rooted Germans.

Our contribution to the literature is to apply experimental tools on ingroup favoritism to better understand the populist surge, by designing an experiment with several unique features that are insightful in the current political context. Most of the existing studies using natural groups to analyze ingroup-outgroup relationships have considered interactions between residents of different countries. Only rarely has research looked at within-country relationships. When it did so, it typically involved at most two ethnic groups within a country using small or non-representative samples.[2] Our study is the first, to the best of our knowledge, to study experimentally within-country inter-ethnic relationships in large Western countries using nationally representative samples and including more than two ethnic groups. This enables us to match participants from the ethnic majority and the two largest ethnic minorities living in the country. This is relevant to understanding populist attitudes because most of the populist discourse is targeted to people from ethnic minorities. Ethnic minorities may be citizens of the country but they may be somehow portrayed as 'second class citizens" because they do not descend from the country's "founding fathers". Or they may be immigrants and as such not have citizenship status. Populist discourse also target potential immigrants, but in many cases potential immigrants share the same ethnicity as residents of the country, such as Hispanic Americans in the US and Turkish people in Germany. Moreover, we can study the behavior of both the ethnic majority and the ethnic minorities thanks to our large and representative samples. Studies that did use large and

[2] An exception is Tanaka and Camerer (2016), who studied inter-ethnic relationships among the ethnic majority and two ethnic minorities in Vietnam rural villages.

representative samples focused on between-country biases, without considering within-country ethnic differences (Dorrough and Glöckner, 2016; Romano et al., 2017) or focused on the ethnic majority within a country (Cettolin and Suetens, 2019). Romano et al. (2017) have subjects from 17 countries play a TG where the nationality of the second mover varies. They find strong evidence of an ingroup bias in most countries, especially in Japan and Germany. Their survey was conducted on about 100 subjects per country, which impedes the analysis of within-country ingroup bias. Our paper departs significantly from their work by focusing on within-country discriminatory patterns and the interaction between ethnicity and income. Moreover, while decisions in all countries except two were hypothetical in Romano et al. (2017), all decisions are monetarily incentivized in our experiments. Dorrough and Glöckner (2016) run surveys on representative samples from Germany, India, Israel, Japan, Mexico, and the United States, to analyze how cooperation and expectation patterns depend on the nationalities involved. They show that when the partner's nationality is known, people hold strong and transnationally shared expectations (i.e., stereotypes) concerning the cooperation level of interaction partners from other countries, which are often wrong. Their study does not specifically focus on ingroup favoritism. Cettolin and Suetens (2019) conduct an experiment on a representative sample of the Dutch population and show that the return of trust by Dutch natives is lower for immigrants than for other natives. Their paper does not focus on trust. Buchan et al. (2009) find that propensity to cooperate with foreigners – as opposed to people from one's local area – covaries with country-level and individual-level involvement in global networks.

Our focus on *within*-country ethnic biases is also relevant to understand social cohesion in a country (Dragolov et al., 2014). There is a widespread concern that increased ethnic diversity may threaten social cohesion (Alesina and La Ferrara, 2000; Putnam, 2007; Algan et al., 2016). Indicators of social cohesion show that it is receding in the US (Dragolov et al., 2014). Even if the aggregate index of social cohesion is not decaying in Germany, the sub-component relative to "acceptance of diversity" is indeed falling in Germany (Dragolov et al., 2014).

The second part of our experiment contributes to the literature showing how the ingroup bias is altered by the release of income information, which acts as an exogenous shock on stereotypes. This part of our design is related to the literature studying the relationship between income and trust in TGs. Several papers use endowment manipulations in the lab to investigate how income inequality affects trust (Anderson et al., 2006; Lei and Vesely, 2010; Smith, 2011; Greiner et al., 2012). Lei and Vesely (2010) introduce income inequality by varying the show-up fee of the laboratory subjects. They find that only "rich" subjects display ingroup favoritism (based on income) in a trust game, while "poor" subjects send larger amounts to rich subjects. Similar to our paper, Falk and Zehnder (2013), Trifiletti and Capozza (2011) and Bogliacino et al. (2018) use variation in participants' incomes outside the lab rather than lab-induced variation in order to provide more sizable and realistic income differences. Falk and Zehnder (2013) show in a large experiment in Zurich that first movers in a trust game prefer sending larger amounts to second movers living in high-income neighborhoods compared to participants from poorer locations. Studying the effect of information on real-life traits or group characteristics is an increasingly used method (Cruces et al., 2013; Kuziemko et al., 2015; Alesina et al., 2018a,b; Mai et al., 2019) In their cross-country study, Alesina et al. (2018a) examine a condition in which immigrants are portrayed as being "hard-working", but do not find significant effects on preferences for redistribution compared to baseline. Likewise, Mai et al. (2019) find higher propensities to transfer to immigrants when they are portrayed as performing community work, but

only when controlling for preferences for redistribution. We significantly depart from previous studies, as we focus on the interaction between ethnicity and income achievement.

The paper is organized as follows. Section 4.2 describes the experimental design. Sections 4.3 and 4.4 present and discuss the results. Section 4.5 concludes.

4.2 Experimental design

4.2.1 Trustlab methodology

Trustlab is a cross-country initiative that aims at running large-scale online surveys on social preferences and trust in institutions (see Murtin et al., 2018 for more details). Participants were recruited by the same private sector polling company in every country and are nationally representative by age, gender and socio-economic status. It was not possible to include ethnicity as a criterion for representativeness. The surveys on the US and Germany were conducted in early 2018.

The US sample includes 1090 subjects and the German sample 1108. Both samples differ markedly compared to traditional student lab samples. For examples, in the US (Germany), 55% (47%) of the subjects are over 45, and 55% (62%) are employed. Women represent about half of the sample in both countries and 30.6% (40.8%) of the subjects are in the top two national income quintiles in the US (Germany). The descriptive statistics of the samples can be found in Appendix Tables 4.B1 and 4.B2, where we also report the actual population distribution for various demographic characteristics.

Trustlab includes three sections. The first section consists in a series of monetarily incentivized experimental games, the second is an Implicit Association Test to measure trust in institutions while the third is a questionnaire tapping into demographic and attitudinal characteristics and other measures of trust in institutions. The first game in the experimental section of *Trustlab* is a standard TG (Berg et al., 1995). A first mover (A) and a second mover (B) both receive 10 dollars in the US or 10 Euros in Germany. A can send none, some, or all of her endowment to B, in multiples of 1 dollar or 1 Euro. The latter receives this amount multiplied by 3 and can send back to A any amount out of her endowment (10 dollars) augmented with the amount received from A (multiplied by 3). Participants sequentially play this TG as first mover and as second mover. This is followed by two public goods games, a dictator game, and finally a series of games that differed across countries. In this paper we focus on six TGs that were played in the US and Germany. We refer to three of these TGs as the "Ethnicity block" and to three TGs as the "Ethnicity plus Income" block.[3]

Ethnic block: trust games with only ethnicity information

In the Ethnic block, participants are involved in a TG where we disclose the second mover's ethnicity to first movers. The first mover is truthfully informed that the second mover does not know the first mover's ethnicity. Each subject in the US is matched in random order with an African American, a Hispanic, and a non-Hispanic White second mover. In Germany, subjects are matched randomly with a subject of Eastern European descent, one of Turkish descent, and a rooted German. Therefore, in both countries, all participants play three TGs, each time with a different person of different ethnicity.

We used the three largest ethnic group as prompts in each country. Own ethnic identity is determined by standard survey questions in the third module of *Trustlab*. In the US, the second mover

[3]Our results are robust to controlling for subjects' transfers in the first TG of the experimental part, i.e. the amount transferred from A to B in a TG where no information about the second mover is given (generalized trust).

is described as being either "non-Hispanic White", "African American" or "Hispanic". We elicit the ethnic group of each subject in *Trustlab*'s survey module where we ask participants to select the racial/ethnic group they identify the most with among a list of 12 options.[4] Table 4.B1 shows that these three groups constitute 71.5%, 11.2% and 11.3% of our sample, respectively. Although our sampling strategy did not target ethnic composition, our sample is fairly close to the real US ethnic composition, according to census data, as Table 4.B1 shows. Nevertheless, we slightly over-represent White individuals and under-represent Hispanic ones. In Germany, ethnicity is defined in the following way: *"a person who was born in Eastern Europe/Turkey/Germany or whose parents were born in Eastern Europe/Turkey/Germany"*. We reconstruct ethnic groups in *Trustlab*'s survey module using two questions. We asked subjects in which country they were born (using a drop-down list of countries) and in which country their mother and father were born. We classify as "Eastern Europe descent" ("Turkish descent") a subject who is either born in Eastern Europe (Turkey) *or* whose parents were born in Eastern Europe (Turkey). Table 4.B2 shows that in Germany, ethnic minorities constitute a smaller share of the population compared to Hispanic people and African Americans in the US. For instance, according to *Statistisches Bundesamt*, Turkish descent individuals constitute only 3.4% of the population. We thus decided to oversample people of Turkish origin by conducting an additional wave in June 2018, which targeted them in priority. In the end, our German sample is constituted of 80.1% rooted Germans, 6% Turkish descent subjects and 6.5% Eastern European descent subjects (see Table 4.B2). For brevity of language, in the following we refer to "people of Turkish descent" as "Turkish" and to "people of Eastern European descent" as "Eastern Europeans".

We match first movers and second movers at the end of the experiment, not instantaneously, because we need to obtain ethnic information first. This is made possible by the use of the strategy method to elicit second mover's behavior. First movers receive no feedback regarding second movers' responses between one trust decision and the next (see Section 4.2.1 for more details). Payments have to be made no earlier than 48 hours after the survey is completed, once we are able to match participants with the required characteristics.

The Ethnic block enables us to quantify the extent of the ingroup bias and whether it is uniform across ethnic groups, or whether some outgroups suffer more or less discrimination. Romano et al. (2017) consider three theories that may explain differences in ingroup bias across countries. The first theory argues that ingroup bias thrives when institutions are inefficient or under-performing, because people from an ingroup have larger incentives to cooperate with each other. The second theory argues that ingroup bias should be reduced by the spread of world religions, because of the doctrine of universal brotherhood that these embrace. The third theory claims that the higher the exposure to pathogens, the higher the payoffs from ingroup-based cooperation. Germany and the US are quite similar in all these three accounts[5] and in particular institutions seem to work relatively well for both ethnic majorities. Hence we do not have any reason to believe, a priori, that ingroup bias will be

[4] In the US the question is "*What racial or ethnic group do you belong to?*" and the options are the following: White, African American, White Hispanic, Other Hispanic, American Indian or Alaska native, Asian Indian, Chinese, Filipino, Japanese, Korean, Vietnamese, Native Hawaiian, some other race, more than one. We merged White Hispanic and Other Hispanic to create the Hispanic ethnic group.

[5] According to census data from both countries, Christianity is the main religion, with 70% of US citizens, and 56% of Germans, declaring to be observant. Conversely, 22% of US citizens, and 38% of Germans declare to be unaffiliated. The share of Protestants is higher in the US (47%) than Germany (28%), while that of Roman Catholic is higher in Germany (28%) than in the US (21%). The next largest religious affiliation is Islam in Germany (5%), while less than 1% of the US population is Muslim. All other religious affiliations do not cover more than 2% of the population. Source: https://www.cia.gov/library/publications/the-world-factbook

different between the White Americans and rooted Germans. We therefore posit:

H1a *The ingroup bias is no different in magnitude between White Americans and rooted Germans.*

H1b *The ingroup bias is significant for all the ethnic groups involved in the US.*

H1c *The ingroup bias is significant for all the ethnic groups involved in Germany.*

Conversely, African Americans may perceive that US institutions have been biased against them, and may have therefore, according to the first theory mentioned above, developed stronger ingroup bias than other US ethnic groups. We therefore posit:

H2a *Ingroup bias by African Americans is higher than ingroup bias by other ethnicities in the US.*

Moreover, since White Americans seem to hold more negative views of African Americans than Hispanics (see section 4.2.1), we also posit:

H2b *White Americans discriminate against African Americans more than they do against Hispanics.*

As for Germany, one may posit that discrimination by rooted Germans may be larger against Turkish people than Eastern Europeans, because the former are overwhelmingly Muslims, while the latter are overwhelmingly Christians. Religious differences are undoubtedly a strong reason for outgroup derogation (Brewer, 1999; Putnam, 2007; Bisin et al., 2008). Moreover, we do not have a priori reason to hypothesize that discrimination by ethnic majorities may be bigger for one ethnic minority group over the other.

We therefore posit the following hypotheses:

H3a *Rooted Germans discriminate against Turkish people more than they do against Eastern Europeans.*

H3b *Turkish people and Eastern Europeans have similar levels of ingroup bias.*

***Ethnic plus Income block*: trust games with ethnicity and income information**

In the next series of three TGs, subjects are randomly matched to three different second movers. This time, the first mover knows both the ethnicity (each one of the three types described above) and that the income of the second mover places him or her among the top 20% of the population in the US or Germany. This top income information is real and elicited in the third module of the survey. In this block of TGs, second movers are always top income earners, only their ethnic groups vary.

The Ethnic plus Income block enables us to examine whether discrimination toward outgroups – and ethnic minorities in particular – is reduced when the outgroup has been economically successful. A broad range of literature argues that discrimination is at least in part based on misperceptions or stereotypes about minorities (Gilens, 2009; Alesina and Stantcheva, 2020). Gilens (2009) extensively documented the widespread negative views held by white Americans on minorities and African Americans in particular. US citizens substantially overestimate the percentage of African Americans among the poor. The median US survey respondent believed that half of US poor were African Americans (Gilens, 2009). In reality, according to the US Census Bureau, the percentage of poor African Americans is only around 22% of the total poor population, and there are about 10 million more

white poor than African American poor. Moreover, the stereotype that African Americans are lazy is widespread. In the most recent waves of the US General Social Survey (GSS), 26% of US White respondents thought of African Americans as lazy, while 20% thought of them as hard-working. In contrast, many more US Whites tend to think of Hispanics as hard-working (nearly 50%) and fewer US Whites think of Hispanics as lazy (less than 10%). Attitudes toward minorities were even more negative in the 1990s. US Whites' views about African Americans improved over the laziness dimension (which was 47% in the 1990s), but not in the hard-working dimension (which was 17% in the 1990s). On the contrary, opinions about Hispanics improved in both dimensions, as 37% and 26% of whites thought of Hispanics as lazy and hard-working in the 1990s, respectively. [6]

Gilens (2009) argues that negative beliefs about African Americans are the largest cause of US citizens' lack of support for the welfare state, as they tend to think of welfare recipients as undeserving of the help they receive. Social surveys in Germany do not ask people's opinions about ethnic minorities. We accordingly turn to Alesina et al. (2018a), who focused on attitudes toward immigrants. Since minorities are typically current or past immigrants, their findings are relevant for our experiment. Alesina et al. (2018a) show that negative views over ethnic minorities are not limited to the US but extend to other European countries including Germany. In particular, natives believe that the percentage of unemployed immigrants is more than 20 points larger than what it actually is, and that immigrants are poorer and less educated than than what they actually are. Natives also believe that an immigrant is more likely to receive welfare benefits than a native in the same situation (Alesina et al., 2018a).

Nevertheless, when immigrants do manage to be successful, most natives believe that they deserve their success. 68% of US respondents and 60% of German respondents believe that the reason for immigrants' economic success is their effort, rather than their luck. On the contrary, people tend to believe that luck played a larger role for rich natives than rich immigrants. This is the case for all five countries surveyed in Alesina et al. (2018b). In the US, only 40% of US respondents believe that effort was the cause of success for rich natives.[7] Opinions differ across countries on the reasons for people being poor. In the US people believe that poor immigrants are less responsible for their condition than poor natives, although the difference in perception is small. In some European countries, such as France and Italy, citizens attach a much larger blame to poor immigrants than poor natives, while the blame is similar for natives and immigrants in Sweden and the UK (Alesina et al., 2018b).

We infer from this evidence that natives substantially underestimate the chances of minorities' economic success, but think that rich people from minority groups deserve their success – even more than rich natives. Therefore, the manipulation of our Ethnic plus Income block seems potentially effective to reduce discrimination for economically successful minorities. We also posit, for the same reasons, that ethnic minorities should also reward economically successful people from ethnic minorities

[6] Interestingly, African Americans, too, tend to think of Hispanics as more hard-working than themselves. Within the 2016 and 2018 GSS waves, only 36% of Black American respondents think of African Americans as hard-working, while 60% of African Americans think of Hispanics as hard-working. A 10% gap also exists between African Americans thinking of African Americans as lazy (21%) in comparison with African Americans thinking of Hispanics as lazy (11%). On the contrary, in the 1990s Black Americans tended to have a better opinion of African Americans than Hispanics. It is as if African Americans' self-image got tarnished over time, while the image of Hispanics improved across the spectrum of ethnicities.

[7] Germany was not included in the survey asking about perceptions of the causes of success for rich natives in Alesina et al. (2018b), thus a comparison with perceptions about rich immigrants is not possible in Germany. It seems plausible that Germans' perceptions are not too dissimilar to those of citizens from other European countries included in the survey.

more than successful people from the ethnic majority, with a possible ingroup bias in this respect, too. We therefore posit:

H4a *Ingroup bias will be lower when first movers are matched with rich second movers than with people with unspecified income, both in the US and Germany, and across all ethnic groups.*

For the US, we can formulate the further hypothesis that the success of African Americans should be rewarded even more by white Americans than success by Hispanics, because of the widespread opinion that African Americans are lazier and less hard-working than Hispanics. In other words, since beliefs about the average Black American are more negative than beliefs about the average Hispanic, the observation of a successful African American should lead to a larger correction of initial beliefs than the observation of a successful Hispanic, in comparison with the baseline. We then hypothesize:

H4b *Ingroup bias by White Americans will drop more for African Americans than Hispanics in the "Ehnic plus Income" treatment compared to the "Ethnic" treatment.*

Given the absence of survey evidence on inter-ethnic attitudes in Germany, we cannot formulate precise hypotheses with respect to the drop in ingroup bias by rooted Germans between the two treatments.

Behavior as second mover in the trust game

We use the data from the second decision of the experimental section of *Trustlab* to measure trustworthiness. As mentioned in Section 4.2.1, participants played a standard TG as second movers in the second experiment of this section. No information on ethnicity was given in this decision, so this measure captures the general propensity to reciprocate trust from a member of the general population.[8] The amount that the second mover decides to send back to the first mover in the TG is elicited using the strategy method: subjects have to decide how much they would send back to the first movers for each of the 11 possible decisions of the first movers.

Expectations

For the TGs of the Ethnic block, we also elicited first mover beliefs regarding the second mover trustworthiness, i.e. the amount that the first mover expects to get back from the second mover. We use the following question to elicit expectations in the US: *"Imagine you sent 5 dollars, so Participant B receives 15 dollars, making his or her total budget 25 dollars. Participant B has no information about your identity. What amount would you expect Participant B to return to you? Please enter a number from 0 to 25."*. We ask this question three times, once for each of the ethnicities mentioned in the Ethnic bloc. In Germany the elicitation question substituted euros for dollars. Expectations are elicited between the end of the Ethnic block and the beginning of the Ethnic plus Income block. Because of time restrictions, we could not elicit expectations in the Ethnic plus Income block. Instructions for this module and the other modules in the *Trustlab* survey can be found in Appendix 4.D.

[8] We have only decisions as second mover that were taken for the generic TG, before subjects knew of the TGs to be played as first mover with knowledge of second mover's ethnicity. We truthfully inform first movers' in that module that second movers don't know their counterpart's ethnicity. When making the generic second mover decision, the participant also had no knowledge that a module referencing ethnicity would appear later. For each game, participants were matched randomly with another participant of the survey, living in the same country. They were never matched with the same participant twice.

4.2.2 Duration and payoffs payment

Participants were paid for the result of one randomly selected game among those played in the experimental section of *Trustlab*. In this way no income effect affected decisions. Each block of the experimental section had equal probability of being selected. Participants were paid in their private bank accounts, as is customary for interviewees of the poll agency. The median time needed to complete the whole *Trustlab* survey was 35 minutes.[9] Participants were paid on average 12 euros in Germany and 11.8 dollars in the US. The largest possible payoff in the survey was 40 euros/dollars.

4.3 Results

4.3.1 Overview of the results

Figure 4.1 shows how the ingroup bias varies between countries and ethnic groups and across both relevant experimental blocks. Here we look at the general tendency to trust people from one's ingroup more than people from the outgroup, without distinguishing between differences in trust toward the two outgroups (a disaggregation we discuss subsequently). For each type of first mover, we regress the amount transferred to the second mover on a binary variable indicating whether the second mover belongs to the same ethnic group. The dependent variable is standardized to have a mean equal to 0 and a standard deviation equal to 1 within each sample. We report on the graph the coefficient of the ingroup dummy, by type of first mover and decision block.[10]

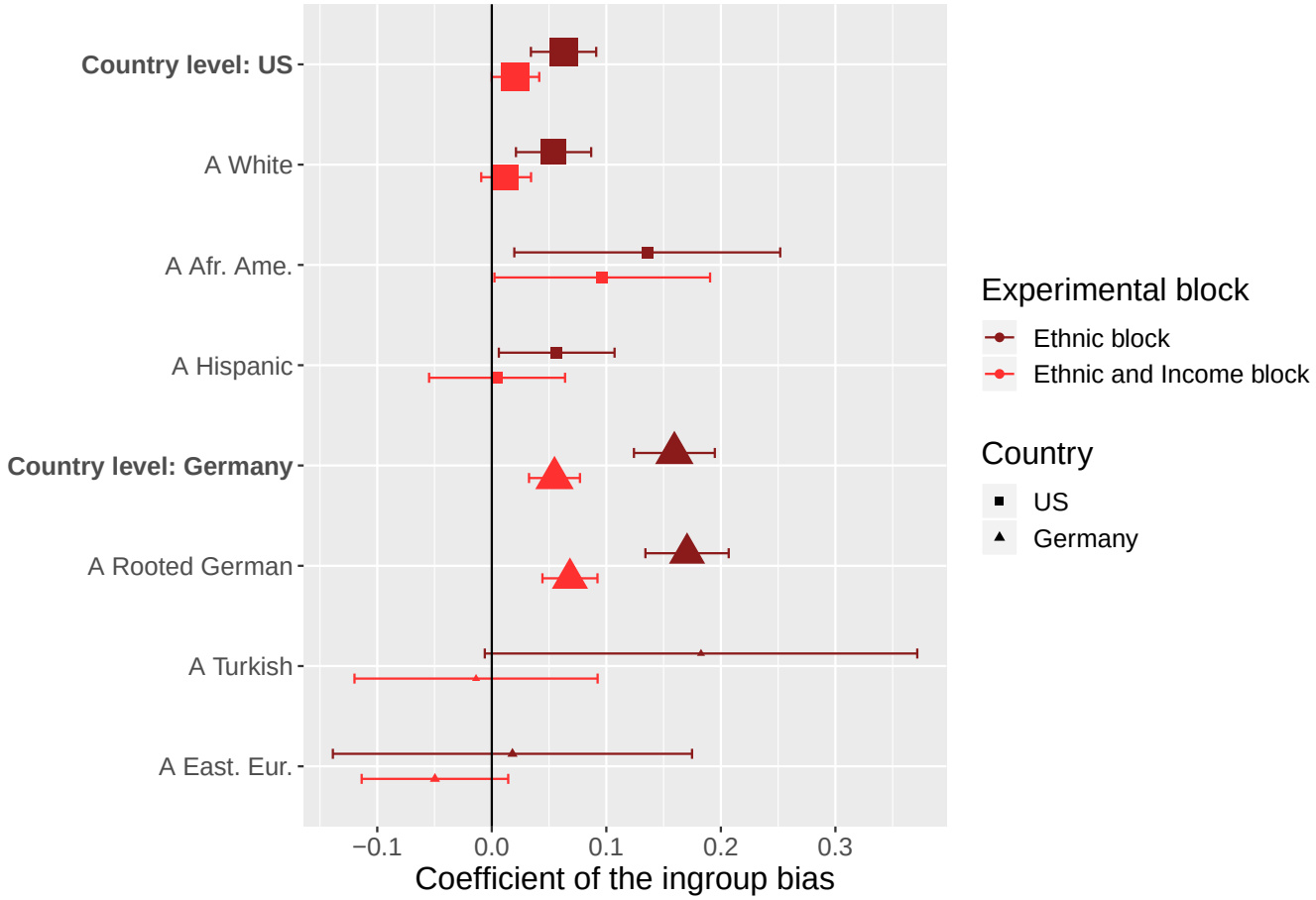
Choices made when only ethnicity is known (Ethnic block) show the existence of significant ingroup biases, especially among rooted Germans and African Americans, followed by White Americans and Hispanics. The ingroup bias is large for Turkish living in Germany, but is imprecisely estimated. There is no ingroup bias for Eastern Europeans living in Germany. Rooted Germans transfer 17% of a standard deviation more money to other rooted Germans compared to outgroup second movers. This corresponds to a transfer gap of about 50 cents (out of the 10 euros endowment) between ingroup and outgroup. White Americans have a statistically significant but smaller ingroup bias: about 5% of a standard deviation, which corresponds to a transfer gap of 15 cents.

Overall, the ethnic ingroup bias is twice as large in Germany as in the US. This result is confirmed in Appendix Table 4.B3, which pools US and Germany data and interacts the ingroup dummy with a binary variable indicating the country of residence of the subject. This result is mostly driven by the difference in attitudes between the majority groups. White Americans engage in ingroup favoritism at a much lower rate than rooted Germans. We conclude:

[9] We judged that the necessary duration to complete the survey carefully enough was about 30 minutes. The average time to complete the survey (71 minutes) was twice as large as the median time, because some participants took a very long time to complete the survey, as is sometimes the case for online surveys. It is likely that some people took breaks from the survey. Nevertheless, such extreme cases are rare. 95% of the participants completed the survey in less than 102 minutes.

[10] See Appendix Figures 4.A1, 4.A2 and 4.A3 for raw statistics of transfers by first mover and second mover types. Appendix Table 4.B4 tests whether the extent of the ingroup bias is statistically different across first mover ethnic groups. More precisely, we can answer for instance the following question: are African Americans favoring significantly more their ingroup compared to White first movers? To make this comparison, we compute the difference between the ingroup transfer and the average of the two outgroup transfers for each type of first mover. Appendix Table 4.B4 shows regressions of this variable on binary variables for each first mover ethnic groups and other socio-demographic characteristics. We find that although African Americans have a larger ingroup bias compared to White and Hispanic first movers, the difference is not statistically significant. However rooted Germans have a significantly larger ingroup bias in both blocks (columns 3 and 4) compared to Eastern European descent first movers.

Figure 4.1 – Size of the ethnic ingroup bias by first mover ethnic group and experimental block



Notes: The Figure plots coefficients of OLS regressions where the dependent variable is the amount transferred from the first mover to the second mover (standardized to have mean 0 and standard deviation 1 within each sample) on a dummy variable for whether the second mover is from the same ethnic group as the first mover currently making the transfer decision. We report the coefficient of this ingroup dummy for different samples, depending on the country of residence and ethnic group of the first mover. The dark dots show regressions on the Ethnic block and lighter dots show the coefficients on the Ethnic plus Income block. The size of the dots is proportional to the sample size. There are 3081 observations in the German sample (2664 rooted Germans, 201 Turkish descent and 216 Eastern European descent first mover observations) and 3072 observations in the US sample (2337 White, 366 African Americans, and 369 Hispanic first mover observations). We control for a variable indicating whether the observation corresponds to the first second mover encountered within the block to capture any first round effects. Subject-level control variables include gender, four age groups, two income levels, two education levels and three employment category binary variables. Standard errors are clustered at the individual level (there are three observations per first mover). We report 95% confidence intervals around the coefficients.

Result 1a *The ingroup bias is higher for rooted Germans than White Americans, thus contradicting H1a.*

Once first movers additionally know that the second mover is from the top 20% (Ethnic plus Income block), the ingroup bias more than halves in both countries and becomes non-significant except for rooted German and African American first movers.

More generally, we find significant differences in trust behavior both between and within countries. Trust is about 7% higher in Germany than the US in the trust decision where ethnicity is not specified (not shown), and about 5% higher in the trust decision where ethnicity is specified (see Figure 4.A2). In the latter case, we use the weighted average of decisions in the Ethnic Block, with weights given by the relative proportion of each ethnic group in real life. Both differences are statistically significant at the 1% level in two-tailed non-parametric Mann-Whitney-Wilcoxon tests. Trust in general others by rooted Germans is also significantly higher than trust by White Americans. In both countries, one group stands out as trusting significantly less than all others. In the US, African Americans trust general others 13% less than White Americans and 5% less than Hispanics. All tests matching pairs of ethnic groups return the result that African Americans trust significantly less than any other group for both general trust and trust in ethnic groups – including the residual group made up of ethnicities not belonging to the three target ones, with only one exception.[11] Differences are significant at the 1% level when African Americans are matched with White Americans, or lower levels of significance in other cases. No difference is instead significant for pairwise tests between any other ethnic groups in the US. In Germany, it is Turkish people who trust significantly less than all other groups. The Turkish trust general others 19% less than rooted Germans and 18% less than Eastern Europeans. Even here, non-parametric pairwise tests reject the null of equality of distributions for any other ethnic group matched with Turkish people, while no other pairwise test between other groups rejects the null. This is the case for both trust in general others and trust in specified ethnicities. The differences are statistically significant at the 1% level in pairwise tests including Turkish and rooted Germans and at lower levels of significance when Turkish are matched with other ethnic groups. Hence, our evidence points to one ethnic group in both countries maintaining significantly different trusting behaviors in comparison with all others.

4.3.2 Ethnic block

In this section, we describe how being confronted with ethnic diversity affects transfer levels in the TG. We first investigate whether first movers make selective transfers depending on the type of outgroup second mover with whom they are matched (Section 4.3.2). We then decompose the ingroup bias into a taste-based and a statistical-based component (Section 4.3.2).

Selective ingroup bias

Table 4.1 displays the results of pooled OLS regressions on the US sample. The dependent variable is the amount transferred from the first mover to the second mover and the main explanatory variables are dummies for the ethnic group of the second mover. Each pair of columns focuses on one particular first mover ethnic group. In the first six columns, the omitted second mover ethnic variable is one's

[11] The only case in which a pairwise test including African Americans fail to reject the null of equality of distributions is that including African Americans and Hispanics in the decision on trust to general others.

ingroup, so that negative coefficients always reflect a positive ingroup bias, that is, that the participant transferred more to the ingroup than the outgroup. We cannot use the same procedure for other US first movers (last two columns) since they are never confronted with an ingroup member. In that case, we make White the omitted category.[12] Looking at columns 1, 3 and 5, we clearly observe that White, African American and Hispanic first movers all display a significant ingroup bias. This ingroup bias is not selective for any group, as shown in the last row of Table 4.1, which reports the p-value of a t-test on the hypothesis that the two second mover ethnic group coefficients are equal to each other. For instance, in columns 1 and 2, the null hypothesis is that the coefficient of B African American and B Hispanic are the same. For any US ethnic group the null of equality of outgroup coefficients is never rejected. We conclude:

Result 1b *Ingroup bias is significant for all ethnic groups in the US, in accordance with H1b.*

Result 2a *Ingroup bias is larger for African Americans than other US ethnic groups, confirming H2a.*

Result 2b *Discrimination by White Americans against African Americans is not different from discrimination against Hispanics, contradicting H2b.*

We also note:

Result 2c *Ingroup bias is not selective in the US for all three ethnic groups.*

[12] There are 40 Asian American first movers and 26 that identified with another race or ethnic category. Since there are 3 rounds per individual this yields 198 observations.

Table 4.1 – Trust game transfers in the US - Only second mover ethnic group is known

	A White		A Afr. Ame.		A Hispanic		A other	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent variable: transfers from 1st mover to 2nd mover in the trust game								
B White			-0.372**	-0.410**	-0.150*	-0.234**		
			(0.173)	(0.186)	(0.0846)	(0.103)		
B Afr. Ame.	-0.185***	-0.143**			-0.163**	-0.140	-0.263	-0.275
	(0.0552)	(0.0557)			(0.0758)	(0.0861)	(0.170)	(0.167)
B Hispanic	-0.126**	-0.0994**	-0.310**	-0.292*			-0.412**	-0.424**
	(0.0490)	(0.0503)	(0.143)	(0.148)			(0.180)	(0.177)
Expected transfer from B to A		0.194***		0.203***		0.264***		0.0791
		(0.0186)		(0.0414)		(0.0301)		(0.0703)
Constant	5.962***	4.277***	5.753***	3.998***	7.213***	4.713***	4.332***	3.870**
	(0.416)	(0.434)	(0.854)	(0.811)	(1.069)	(0.981)	(1.568)	(1.607)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2337	2337	366	366	369	369	198	198
R^2	0.026	0.143	0.100	0.264	0.092	0.350	0.207	0.220
p-value of test H0:								
B ethnic group 1 = B ethnic group 2	0.138	0.280	0.571	0.308	0.869	0.231	0.187	0.178

Standard errors clustered at the individual level in parentheses

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

Notes: Results of pooled OLS regressions using all the three decisions made by each first mover in the Ethnic block of the trust games on the US sample (only ethnicity of second mover is known). The dependent variable is the level of transfers from the first mover to the second mover. The explanatory variables are binary variables for the ethnic group of the second mover. We additionally control for a variable indicating whether the observation corresponds to the first second mover encountered (first transfer made in the Ethnic block) to capture any first round effects. Subject-level control variables (in all columns) include gender, four age groups, two income levels, two education levels and three employment categories binary variables. Columns 1 and 2 restrict the sample to White first movers, 3 and 4 to African Americans first movers, 5 and 6 to Hispanics first movers, 7 and 8 to other US first movers (Asian Americans, American Indian or Alaska Native, more than one ethnic group and other ethnic groups). The last row of the table reports the p-value of a t-test comparing the two second mover ethnic group coefficients. For instance, in columns 1 and 2, the null hypothesis is that the coefficient of B African American and B Hispanic are the same.

Table 4.2 – Trust game transfers in Germany - Only second mover ethnic group is known

	A rooted German		A East Eur.		A Turkish		A other	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent variable: transfers from 1st mover to 2nd mover in the trust game								
B rooted German			0.288	0.191	-0.363	-0.302		
			(0.249)	(0.276)	(0.326)	(0.324)		
B East. Eur.	-0.360***	-0.242***			-0.770**	-0.641**	-0.0409	0.0468
	(0.0582)	(0.0604)			(0.316)	(0.320)	(0.113)	(0.120)
B Turkish	-0.628***	-0.509***	-0.396	-0.375			-0.148	-0.124
	(0.0628)	(0.0622)	(0.297)	(0.336)			(0.184)	(0.159)
Expected transfer from B to A		0.220***		0.214***		0.249***		0.217***
		(0.0167)		(0.0744)		(0.0553)		(0.0499)
Constant	6.165***	3.952***	5.869***	4.384***	5.711**	4.210**	8.200***	5.597***
	(0.548)	(0.530)	(1.581)	(1.308)	(2.155)	(1.843)	(1.539)	(2.106)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2664	2664	216	216	201	201	243	243
R^2	0.035	0.151	0.207	0.296	0.084	0.243	0.209	0.335
p-value of test H0:								
B ethnic group 1 = B ethnic group 2	0.000	0.000	0.009	0.037	0.083	0.154	0.571	0.282

Standard errors clustered at the individual level in parentheses

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

Notes: Results of pooled OLS regressions using all three decisions made by each first mover in the Ethnic block of the Trust games on the German sample. The dependent variable is the level of transfers from the first mover to the second mover. The explanatory variables are binary variables for the ethnic group of the second mover. We additionally control for a variable indicating whether the observation corresponds to the first second mover encountered (first transfer made in the Ethnic block) to capture any first round effects. Subject-level control variables (in all columns) include gender, 4 age, 2 income level, 2 education level and 3 employment category binary variables. Columns 1 and 2 restrict the sample to rooted Germans first movers, 3 and 4 to first movers from Eastern Europe, 5 and 6 to first movers of Turkish descent and 7 and 8 to first movers with another ethnicity. The last row of the table reports the p-value of a t-test comparing the two second mover ethnic group coefficients. For instance, in columns 1 and 2, the null hypothesis is that the coefficient of B Eastern Europe and B Turkish are the same.

Results are markedly different in Germany. First, the ingroup bias is strongly significant at the 1% level for rooted Germans with respect to both outgroups, but is not statistically significantly different from 0 for Eastern Europeans. It is only significant at the 5% level for Turkish against Eastern Europeans, but is not significant for Turkish against rooted Germans (see Table 4.2, column 1, 3 and 5). Hence, the ingroup bias seems to be concentrated in the ethnic majority in Germany. Second, the ingroup bias is selective. Rooted Germans discriminate twice as much against Turkish second movers compared to Eastern Europeans second movers, the difference being significant at the 1% level (see last row of Table 4.2, column 1). Eastern Europeans are also selective. Even if the ingroup bias was statistically insignificant against the two outgroups, it had opposite sign. Our test confirms, at the 1% level, that Eastern Europeans trust rooted Germans significantly more than Turkish (see Table 4.2, column 3). In fact, Eastern Europeans are the only group in the experiment displaying outgroup favoritism, as they transfer more to rooted Germans than to fellow Eastern Europeans. Finally, first movers of Turkish descent are also selective, as they transfer more to rooted Germans than to Eastern Europeans, the difference being in this case weakly significant. Overall, we find that Turkish and Eastern European first movers tend to discriminate against each other and discriminate much less against – or in fact favor – rooted Germans, while the latter discriminate against both of them and particularly against people of Turkish descent.

Results for participants in the German sample who fall into none of the three identified categories are displayed in columns 7 and 8 (labeled "Other") of Table 4.2. These subjects (for whom we make rooted German the omitted category) show no statistically significant selective discrimination, but we may be lacking statistical power to detect a significant effect.

We conclude:

Result 1c *Ingroup bias is significant for rooted Germans, but is not significant for Eastern Europeans and only significant for Turkish against Eastern Europeans but not against Germans, thus partially contradicting H1c.*

Result 3a *Rooted Germans discriminate against Turkish people more than Eastern Europeans, confirming H3a.*

Result 3b *Turkish people and Eastern Europeans have similar levels of ingroup bias, confirming H3b.*

We also note:

Result 3c *Each German ethnic group performs selective discrimination, that is, people from one ethnic group transfer significantly more to one outgroup than to the other outgroup.*

Statistical vs taste-based discrimination

We have documented the variations in the ethnic ingroup bias in the US and Germany across first mover and second mover types. We can decompose this bias into two components: a statistical bias and a taste-based bias. First movers may discriminate against an ethnic group because they expect that they will receive back a lower amount from people from that group in the TG (statistical discrimination), or they can prefer to transfer more to one ethnic group than another independent of any transfer expectations (taste-based discrimination).

To quantify the relative importance of statistical and taste-based discrimination, we control for the amount that the first mover expects to receive from each type of second mover at the end of the TG.[13] Table 4.3 shows regressions for the whole dataset (Germany and the US combined) in columns 1 and 2, US only in columns 3 and 4 and Germany only in the last two columns. For all groups the return expectation is a strongly significant predictor of the transfer to the second mover, thus confirming that first movers' transfers depended on the second movers' expected trustworthiness. The coefficient of the ingroup dummy (A and B belong to the same ethnic group) in column 1 can be directly compared to the coefficient of the same variable in the next column, where we control for the expected transfer from second mover to first mover. If the coefficient of the ingroup variable is driven down to zero, then it would mean that statistical discrimination is fully driving the result. But this is not the case: the ingroup coefficient remains sizable when we control for the expected transfer from B to A. Pooling German and US data, we see that the coefficient drops by about 21%. Separating US and German data, we see that the drop is slightly larger in Germany (25%) than in the US (16%).

We conclude:

Result 5 *Statistical discrimination accounts for about one fifth of ingroup favoritism. Taste-based discrimination is the main driver of the ingroup bias.*

Table 4.3 – Statistical vs taste-based discrimination

	(1) Pooled data	(2) Pooled data	(3) US	(4) US	(5) Germany	(6) Germany
Dependent variable: transfers from 1st mover to 2nd mover in the trust game						
A and B same ethnic group	0.287***	0.226***	0.180***	0.151***	0.395***	0.295***
	(0.0401)	(0.0399)	(0.0534)	(0.0538)	(0.0593)	(0.0585)
Expected transfer from B to A		0.213***		0.201***		0.226***
		(0.0105)		(0.0151)		(0.0146)
Constant	5.804***	3.968***	5.855***	4.160***	5.828***	3.748***
	(0.277)	(0.276)	(0.349)	(0.353)	(0.491)	(0.475)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6594	6594	3270	3270	3324	3324
R^2	0.016	0.145	0.020	0.150	0.031	0.156

Standard errors clustered at the individual level in parentheses

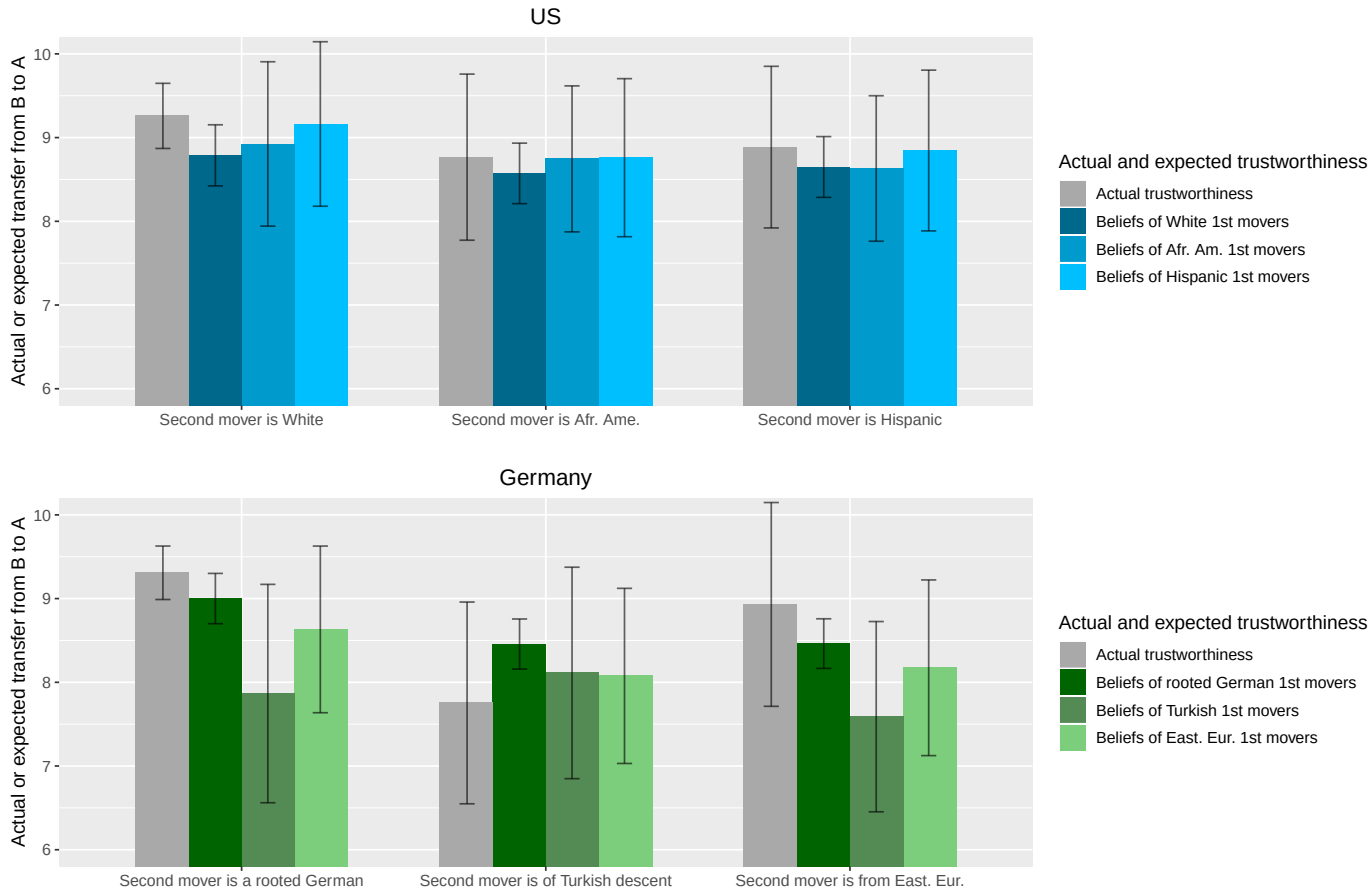
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The table reports the results of pooled OLS regressions using all three decisions made by each first mover in the Ethnic block of the TGs. The dependent variable is the level of transfers from the first mover to the second mover. The explanatory variables are binary variables for whether the first and the second mover have the same ethnicity and the amount that the first mover expects to receive from the second mover in the median transfer scenario (see Section 4.2.1). We additionally control for a variable indicating whether the observation corresponds to the first round of the Ethnic block to capture any first round effects. Subject-level control variables include gender, 4 age, 2 income level, 2 education level and 3 employment category binary variables.

We can then assess which ethnicity (if any) drives this result within countries. The even-numbered columns of Table 4.1 show the impact of controlling for the expected transfer from B to A on the size of the two outgroup binary variables for each first mover ethnicity in the US. White first movers were manifestly driving the overall US results since we observe a similar drop in the size of the ethnic group binary variables (B African American and B Hispanic) when controlling for expected transfers. This indicates that about 23% of the outgroup bias toward African Americans and 21% toward Hispanics comes from statistical discrimination. In other words, lower transfers to minorities

[13] The variable is described in Section 4.2.1.

Figure 4.2 – Expected and actual transfer from second mover to first mover by ethnic group of the second mover



Notes: The extreme-left bar of each panel (grey bar) shows actual trustworthiness by second mover type, i.e. the amount that second movers choose to send back to first movers, in the case where first movers send 5 dollars/euros. The rest of the colored bars show first movers beliefs (expected trustworthiness), i.e. the amount that first movers expect to receive back by second mover type, in the case where they first send 5 dollars/euros. The top panel represents the US sample and the bottom panel for the German sample. 95% confidence intervals are reported.

by White Americans are partly explained by low expectations on minorities' trustworthiness. On the contrary, the coefficients of B White increases in magnitude after controlling for the expected transfers in the regression for both African Americans and Hispanic first movers (columns 4 and 6 of Table 4.1). This means that African Americans and Hispanic first movers hold relatively high expectations on White Americans' trustworthiness, and these expectations attenuate their taste-based discrimination. Hispanic first movers send on average 15 cents less to White second movers compared to Hispanics second movers, but once controlling for expected transfers, they send 23 cents less on average.

In Germany (Table 4.2), there are no clear cut differences among first movers of different ethnicity in terms of the discrimination patterns. All ethnic groups partly discriminate based on statistical consideration since the outgroup binary variables always decrease in magnitude once controlling for expected transfers from B to A.

Even though discrimination is mainly taste-based and only party statistical, we check whether low trustworthiness stereotypes are accurate. Figure 4.2 plots actual and expected trustworthiness side by side, by second mover type. Actual trustworthiness is measured at the median case scenario, i.e. the amount that second movers choose to send back to first movers, in the case where first movers send 5 dollars/euros. The remaining colored bars show first movers beliefs (expected trustworthiness) and correspond to the amount that first movers expect to receive back by second mover type, in the case where they first sent 5 dollars/euros.

In the US, second mover median transfer choices (actual trustworthiness) are not statistically different across second mover types, although they tend to be higher for White people.[14] So the 20% statistical discrimination we observed was based on inaccurate stereotypes.

The picture is quite different in Germany, where both actual behavior and beliefs show larger variations across subject types. Rooted Germans have the highest trustworthiness and Turkish descent ones the lowest.[15] The latter group sends on average 16.4% less money to first movers compared to other subjects in the German sample, this difference being significant at the 5% level.[16] This behavior was only partly anticipated by participants. Non-Turkish first movers expected Turkish descent participants to send back amounts 3.1% lower compared to the other two types of second movers, but this difference in expectations is not significant (p-value = 0.12). Interestingly, Turkish people expect other Turkish to be the most trustworthy of the three groups, whereas they are in fact return the least of all groups, as second movers.

4.3.3 Ethnic plus Income block

Decrease of the ethnic ingroup bias with income information

We now add income into the picture to study its interaction with ethnic favoritism. In the Ethnic plus Income block, the second mover is always in the top 20% of the income distribution. In these

[14] This is formally tested with a Kruskal-Wallis equality-of-populations rank test, which is a generalization of the two-sample Wilcoxon (Mann-Whitney) rank-sum. The Kruskal-Wallis test checks whether two or more samples (here three) come from the same distribution. In our context, this allows to test for the equality of the distributions of trustworthiness across the three ethnic groups. We cannot reject the null hypothesis that trustworthiness behavior is the same across ethnic groups in the US (p-value = 0.27). The same result holds if we take average trustworthiness across all 11 decisions that second movers make instead of the the median case scenario (p-value = 0.37).

[15] A Kruskal-Wallis allows to reject at the 1% level the equality in trustworthiness distributions across the three ethnic groups in Germany.

[16] The difference between Turkish descent second movers and other subjects in the German sample is even larger (20%) and significant at the 1% level, if we take the average trustworthiness across all 11 second mover decisions, instead of the median transfer (i.e., the backtransfer choice when sent the median option of 5 dollars or euro).

three rounds the ethnic group of the second mover varies, while information on the second mover's income is held constant. As argued in Section 4.2.1, the key hypothesis we want to test is whether discrimination may be attenuated by receiving information of the "economic success" of somebody from minority groups.

To answer these questions, we check how transfers vary across different types of second movers, depending on the ethnicity of the first mover, focusing on the Ethnic plus Income block results only. Table 4.4 reports OLS pooled regressions on all the 3 transfer decisions of the Ethnic plus Income block in the US. We regress the amount transferred by each type of first mover on outgroup dummy variables. Focusing on White, Hispanic, and other US first movers, we can observe that the ingroup bias is considerably reduced by information on income. White Americans' ingroup bias disappears with respect to Hispanics and is reduced with respect to African Americans – though it remains weakly significant in the latter case. The t-test on equality of transfers toward outgroups rejects the null for White Americans, albeit at weakly significant levels only (see Table 4.4, column 1, last line). This result goes against our hypothesis H4b (in Section 4.2.1) that African Americans would have benefited more in this treatment compared to Hispanics. Since White Americans tend to view Hispanics' work ethic in a more positive light than that of African Americans, we expected that correcting this belief would led White Americans to increase their transfers to African Americans more than their transfers toward Hispanics. The fact that this is not the case may be possibly due to economically successful Hispanics being seen as more deserving than economically successful African Americans. Or it could mean that deservedness is not the only – or main – determinant of inter-ethnic discrimination by White Americans. The ingroup bias by African Americans seems to be largely unaffected by the information that second movers from outgroups are economically successful. The difference in transfers toward ingroup and outgroup are about the same in this and the previous treatment for African Americans (see Tables 4.1 and 4.4). The ingroup bias disappears for Hispanics and other ethnic groups do not differentiate among the three main US ethnic groups.

Table 4.5 shows that the ingroup ethnic bias for rooted Germans is substantially smaller when information on belonging to the high-income group is revealed. Comparing the coefficients of Table 4.5 with those of Table 4.2 reveals that rooted Germans' discrimination against Turkish second-movers drops by 64% while discrimination against Eastern Europeans drops by 30%. As a result, rooted Germans' discrimination against Turkish and Eastern Europeans is no longer different from each other when second movers are rich (see Table 4.5, column 1, last row). Nonetheless, the ingroup bias does not disappear completely and it remains statistically significant for rooted Germans. Ingroup bias is no longer significant for ethnic minorities in Germany. Interestingly, both Eastern Europeans and Turkish participants transfer more to rich Germans than to a rich ingroup, the effect being weakly significant for Eastern Europeans (see Table 4.5, columns 2 and 3). This tendency suggests distrust by ethnic minorities toward rich people from their ingroup. The raw data of Appendix Figure 4.A1 shows that the level of transfers declines by about 20% once one knows that the second mover is rich. This gap could be motivated by three different factors. First, subjects may have purely taste-based discriminatory attitudes toward the rich. Second, they may believe that rich people are more selfish and feel less concerned about fairness or reciprocity, and hence are less trustworthy. Third, since transfers partly respond to altruistic concerns for the second mover (Cox, 2004), then transfers to the rich should be lower than transfers to the general population, as the rich are obviously less

Table 4.4 – Trust game transfers in the US - Second mover is rich (Ethnic plus Income block)

	(1) A White	(2) A Afr. Ame.	(3) A Hispanic	(4) A Other
Dep. var.: transfers from 1st mover to 2nd mover in the trust game				
B White top 20		-0.296*	-0.0403	
		(0.158)	(0.120)	
B Afr. Ame. top 20	-0.0725*		0.0088	0.0589
	(0.0415)		(0.108)	(0.0909)
B Hispanic top 20	-0.0125	-0.313**		-0.112
	(0.0411)	(0.155)		(0.125)
Constant	5.472***	3.236***	5.885***	4.122**
	(0.494)	(1.119)	(1.337)	(1.570)
Controls	Yes	Yes	Yes	Yes
Observations	2337	366	369	198
R^2	0.038	0.052	0.110	0.414
p-value of test H0:				
B ethnic group 1 = B ethnic group 2	0.075	0.836	0.603	0.111

Standard errors clustered at the individual level in parentheses

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

Notes: Results of pooled OLS regressions using all three decisions made by each first mover in the ethnic and income block of the Trust games in the US sample. The dependent variable is the level of transfers from the first mover to the second mover. The explanatory variables are binary variables for the ethnic group of the second mover. In he ethnic and income block, all second movers are in the top 20% of the income distribution. We additionally control for a variable indicating whether the observation corresponds to the first round of the block to capture any first round effects. Subject-level control variables include gender, 4 age, 2 income level, 2 education level and 3 employment category binary variables. Column 1 restricts the sample to White first movers, 2 to African Americans first movers, 3 to Hispanics first movers, 4 to Asian Americans first movers and 5 to other US first movers. The last row of the table reports the p-value of a t-test comparing the two second mover ethnic group coefficients. For instance, in columns 1 and 2, the null hypothesis is that the coefficient of B African American top 20 and B Hispanic top 20 are the same.

Table 4.5 – Trust game transfers in Germany - Second mover is rich (Ethnic plus Income block)

	(1)	(2)	(3)	(4)
	A rooted German	A East. Eur.	A Turkish	A other
Dep. var.: transfers from 1st mover to 2nd mover in the trust game				
B rooted German top 20		0.265*	0.182	
		(0.149)	(0.238)	
B East Eur. top 20	-0.253***		-0.0869	-0.0372
	(0.0471)		(0.182)	(0.0471)
B Turkish top 20	-0.225***	0.0906		0.0372
	(0.0461)	(0.126)		(0.167)
Constant	4.695***	3.992	1.358	7.469***
	(0.651)	(2.845)	(2.623)	(1.908)
Controls	Yes	Yes	Yes	Yes
Observations	2664	216	201	243
R^2	0.038	0.235	0.106	0.245
p-value of test H0:				
B ethnic group 1 = B ethnic group 2	0.442	0.240	0.177	0.635

Standard errors clustered at the individual level in parentheses

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

Notes: Results of pooled OLS regressions using all three decisions made by each first mover in the Ethnic plus Income block of the TGs on the German sample. The dependent variable is the level of transfers from the first mover to the second mover. The explanatory variables are binary variables for the ethnic group of the second mover. In the ethnic and income block, all second movers are in the top 20% of the income distribution. We additionally control for a variable indicating whether the observation corresponds to the first round of the block to capture any first round effects. Subject-level control variables include gender, 4 age, 2 income level, 2 education level and 3 employment category binary variables. Column 1 restricts the sample to rooted German first movers, 2 to first movers from Eastern Europe, 3 to first movers of Turkish descent and 4 to other ethnic groups. The last row of the table reports the p-value of a t-test comparing the two second mover ethnic group coefficients. For instance, in columns 1 and 2, the null hypothesis is that the coefficient of B Eastern Europe top 20 and B Turkish top 20 are the same.

needy than others.[17] Time limits with the overall survey prevented us from introducing additional experimental choices or treatments to disentangle these various motivations. Nevertheless, in the next section we develop a strategy that controls for the above confounds and enables us to perform meaningful comparisons on ingroup discrimination when income information is revealed.

Normalized High-Income Bias

In the previous section, we performed comparisons of transfers within the Ethnic Plus Income block and found that ethnic discrimination was largely reduced. But this did not take into account baseline levels of ethnic discrimination. To measure the *net* treatment effect of our exogenous shock on the wealth of the second mover, we need to compare the level of discrimination in the Ethnic Plus Income block with the one of the Ethnic block.[18]

To fix ideas, we define the High-Income Bias ($HIB_{i,ki}$) as the difference between the transfers that a first mover of ethnic group i sends to a rich second-mover of outgroup k and a rich second-mover of ingroup i.

$$HIB_{i,ki} = T_{i,Rich_k} - T_{i,Rich_i} \tag{4.1}$$

When i is the ethnic majority and k is an ethnic minority, $HIB_{i,ki}$ measures the difference in transfers that the ethnic majority sends to a rich outgroup k relative to a rich ingroup i. Our hypothesis H4a concerns the *reduction* of the ingroup bias when the treatment is in place compared to when it is not in place, the treatment being the revelation that the outgroup second mover has high income. For this reason, we need to subtract the ingroup bias between ingroup i and outgroup k observed in the absence of the treatment from $HIB_{i,ki}$. This leads to what we call the Normalised $HIB_{i,ki}$.

$$NHIB_{i,ki} = (T_{i,Rich_k} - T_{i,Rich_i}) - (T_{i,k} - T_{i,i}) \tag{4.2}$$

Where $(T_{i,k} - T_{i,i})$ is the difference between the transfers from a first mover of ethnic group i to a second mover of outgroup k and a second mover of ingroup i in the Ethnic block. $NHIB_{i,ki}$ would be *positive* if our "rich second mover" treatment effectively reduced the ethnic discrimination of the ethnic majority against the ethnic minority. This would suggest the existence of a "deserving rich ethnic minority effect".

Finally, we also measure how i treats the ethnic outgroup 1 relative to the ethnic outgroup 2. In that case, we compute the Normalized High-Income Bias, $NHIB_{i,12}$ as follows:

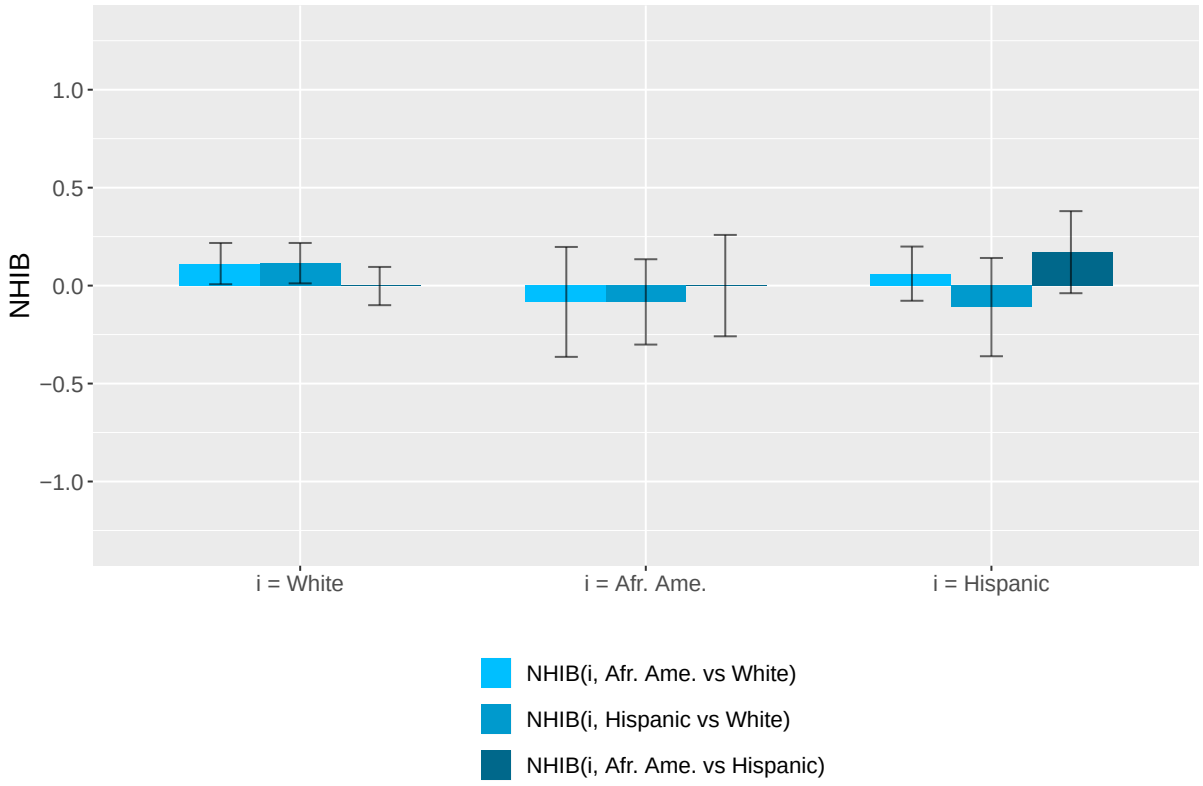
$$NHIB_{i,12} = (T_{i,Rich_1} - T_{i,Rich_2}) - (T_{i,1} - T_{i,2}) \tag{4.3}$$

Figures 4.3 and 4.4 plot the Normalized High-Income Biases for each ethnic group, in the US and in Germany respectively. Both NHIB(White, Afr. Ame. vs White) and NHIB(White, Hispanic vs

[17] We also cannot rule out that the drop in transfers toward the rich could be partially driven by order effects, since the Ethnic plus Income block always comes after the Ethnic block. We preferred not to opt for the alternative design where the order with which the Ethnic block and the Ethnic plus Income block are presented is randomized. In those cases when the Ethnic plus Income block had came before the Ethnic block, it would have been likely, in our view, that the participant expectation over the second mover's income were still affected, in some non-obvious ways, by the information received in the previous block. This would have prevented comparability between the two blocks. In other words, it seemed more natural to move from one setting with less information to one with more information rather than the other way round.

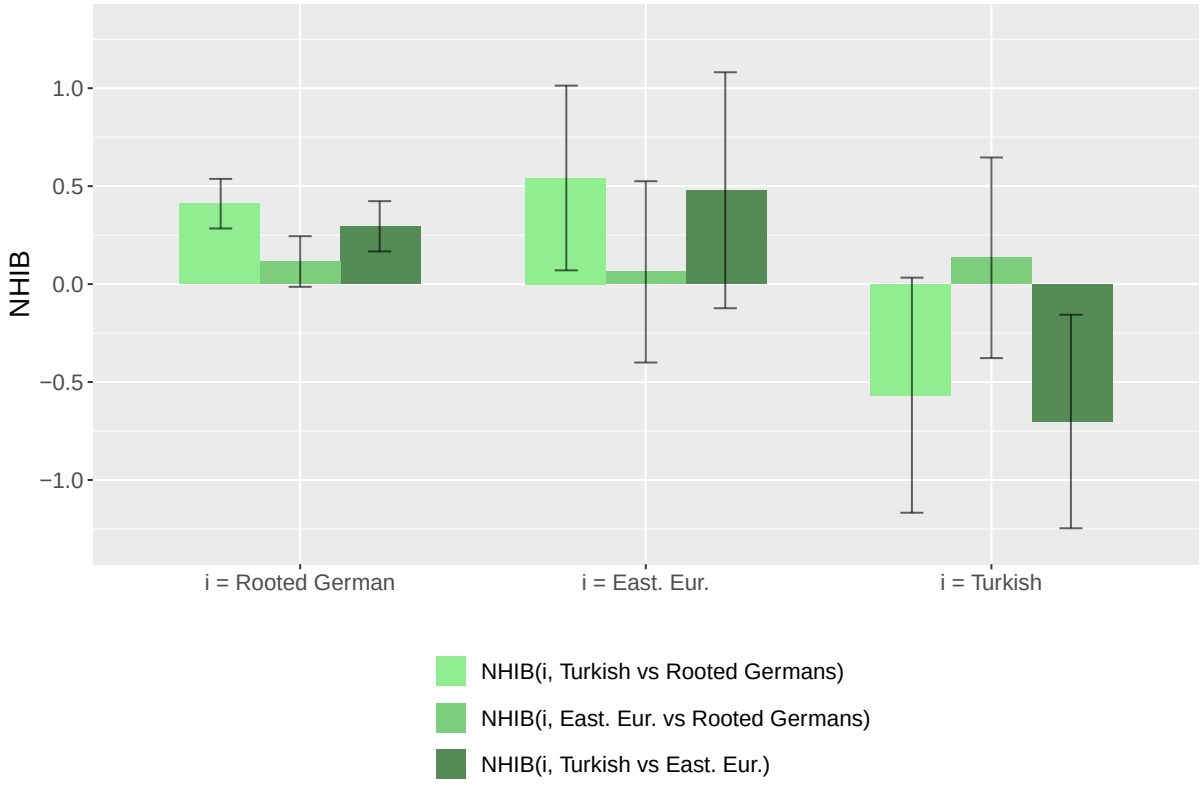
[18] The intuition is the same as the one underlying the diff-in-diff strategy.

Figure 4.3 – Normalized High-Income Bias by first mover ethnicity i – US



Notes: The figure plots the Normalized High-Income Biases computed using Equations (4.2) and (4.3) by type of first mover i in the US sample. We report 95% confidence intervals.

Figure 4.4 – Normalized High-Income Bias by first mover ethnicity i – Germany



Notes: The figure plots the Normalized High-Income Biases computed using Equations (4.2) and (4.3) by type of first mover i in the German sample. We report 95% confidence intervals.

White) are positive, revealing the existence of a "deserving rich ethnic minority effect" – that is, a rise in relative trust or in appraisal of deservingness of minority group members by the majority when the latter learn that the minority members in question have high incomes. White Americans favor equally the African American rich and the Hispanic rich (positive light and medium blue bars of the left panel). NHIB(White, Afr. Ame. vs Hispanic) is thus close to zero.

Figure 4.4 shows the Normalized High-Income Biases in Germany. The size of the NHIBs appears overall larger than in the US, meaning income information causes a greater reduction in discrimination against the outgroups.[19] NHIB(Rooted Germans, Turkish vs Rooted Germans) is larger than NHIB(Rooted Germans, East. Eur. vs Rooted Germans), only the first bias being statistically significantly different from zero. Hence, unlike the US, the NHIB of the ethnic majority is selective in Germany. Turning to NHIB's of the ethnic minorities with respect to their own outgroups, these appear sizable, but in an unexpected direction. The rightmost panel of Figure 4.4 shows that NHIB(Turkish, Turkish vs Rooted Germans) and NHIB(Turkish, Turkish vs East. Eur.) are both negative. In other words, Turkish first movers transfer *less* to a rich Turkish than they transfer to a rich rooted German or a rich Eastern European, in comparison to baseline. There is in fact no significant difference in the NHIB toward the other two ethnic groups by Turkish first movers (middle bar of the rightmost panel). The existence of an NHIB of Turkish participants toward both ethnic majority and Eastern Europeans suggests a general *mistrust* by Turkish people. Eastern Europeans have a positive, and statistically significant, NHIB toward Turkish people relative to rooted Germans, while they treat their own rich in the same manner as they treat rooted Germans. This also means that Eastern Europeans have a more positive NHIB toward Turkish relative to that of Eastern Europeans themselves, though this effect is not statistically different from zero at conventional levels.

We conclude:

Result 4a *Releasing information that second movers are economically successful in real life significantly reduces the ingroup bias, confirming H4a.*

Result 4b *The "deserving rich ethnic minority effect" of White Americans toward African Americans is no larger than the one toward Hispanics, contradicting H4b.*

Moreover:

Result 4c *The "deserving rich ethnic minority effect" of rooted Germans is larger toward Turkish people than toward Eastern Europeans.*

Result 4d *Both Turkish and Eastern Europeans transfer more to rich outgroups than to rich ingroups.*

Finally, it must be kept in mind that members of each ethnic group send less, not more, on average to a second-mover who is in the top 20% by income than to one for whom income information is not given. Thus NHIB or "deserving rich effects" play out against a backdrop of generally lower sending to high earners, and they take the form of a lowering of discrimination against the rich of ethnic minorities relative to the general discrimination against high income second movers, rather than of absolutely larger sending to rich ethnic minorities than to others (see the end of Section 4.3.3).

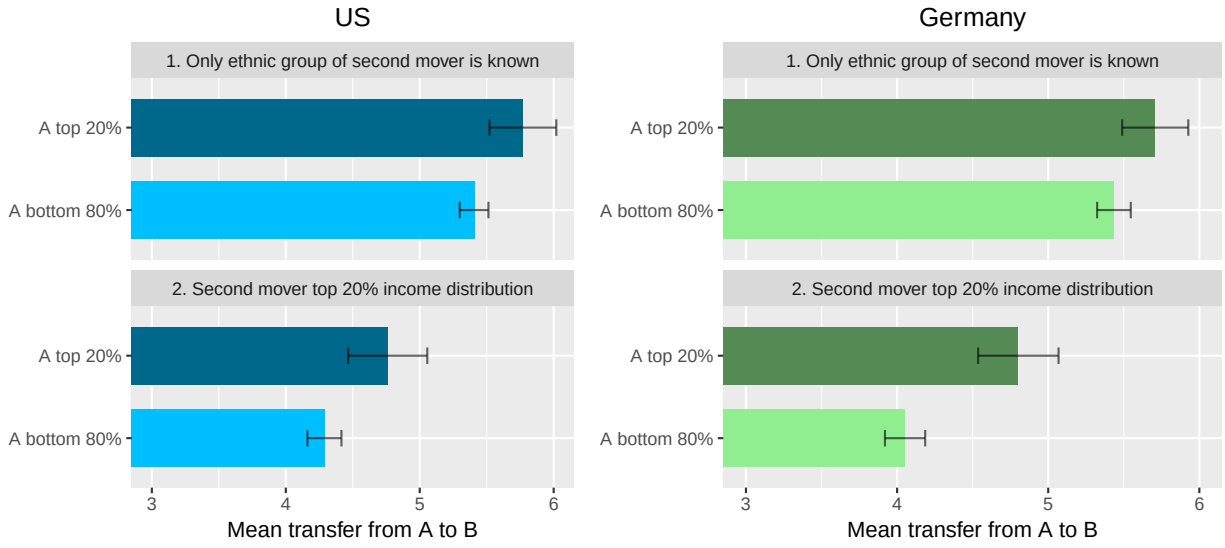
[19] The larger reduction of discrimination in Germany can be partly driven by the fact that the discrimination had been larger in Germany in the first place, so there was more room for intervention.

Income ingroup bias

So far we have considered ethnicity as the relevant source of identification to tell apart ingroups and outgroups. However, people may identify with different groups at the same time and many people experience multiple identities (Brewer, 1999). The Ethnic plus Income block made income salient in addition to ethnicity, by portraying a rich second mover. We conjecture that income may have been a source of identification for first movers who are rich in real life. If this was the case, then we should observe favoritism by rich first movers toward rich second movers, especially if they come from the same ethnic group. We test for this hypothesis in this section.

Figure 4.5 investigates the existence of an ingroup bias based on income rather than ethnicity. It shows how transfers depend on the first mover income level (either top 20% or bottom 80%) and the information delivered regarding the second mover income (no information in the top panel and the second mover is in the top 20% in the bottom panel). Rich first movers transfer significantly larger amounts than bottom 80% subjects. All subjects decrease their transfers when the second mover is known to be rich, but rich first movers do so to a lesser extent than bottom 80% first movers.

Figure 4.5 – Transfer decisions from first mover to second mover in the US and Germany by first mover income



Notes: The bars display the average level of transfers by first mover income (either top 20% of the income distribution or bottom 80%) pooling over the whole US sample for the left panel and the whole German sample for the right panel. The upper part displays *Ethnic block* transfers (only ethnicity is known) and the lower part shows *Ethnic plus Income block* transfers (the second mover is additionally in the top 20% of the national income distribution). 95% confidence intervals are reported.

To investigate this question more systematically, Table 4.6 shows the respective effects of ethnic ingroup ("A and B belong to the same ethnic group") and income ingroup (with the variable "A top 20%"), using Ethnic plus Income block data only. This latter variable enables us to check how rich first movers treat other rich second movers, compared to what bottom 80% first movers do. The last line of the table shows the p-value of a t-test for the equality of both coefficients. It tells us whether the income ingroup bias has the same size as the ethnic ingroup bias. We show the results for all ethnic groups for completeness but the reader should keep in mind that most top 20% first movers in our sample are from the ethnic majority.[20]

[20]In fact, we have only 10 African Americans, 14 Hispanic, 6 Turkish descent and 14 Eastern European descent subjects

Our analysis shows that the hypothesis of a significant income ingroup bias is rejected for all ethnic groups except for rooted Germans. In Germany, the top 20% subjects from the ethnic majority transfer almost one euro more to rich people compared to the amount that bottom 80% first movers choose to transfers to top 20% subjects. Furthermore, for rooted Germans, the size of the income ingroup bias is almost four time larger than the magnitude of the ethnic ingroup bias, and this difference is significant at the 5% level (last line of Column 4 of Table 4.6). We do not find a significant income ingroup bias for White Americans. Although both the ethnic ingroup variable and the income ingroup variables show positive coefficients (suggesting an ingroup bias in both cases), they are not significant. These results should be interpreted cautiously since the ethnic ingroup bias is much smaller in the Ethnic plus Income block than when only ethnicity is disclosed.

that are within the top 20% of their respective national income distribution. So the coefficient of the variable "A top 20%" should be interpreted cautiously for columns 2, 3, 4 and 5 of Table 4.6.

Table 4.6 – Comparison of the ingroup ethnic bias and the ingroup income bias when 2nd mover is in top 20% by income

	US			Germany		
	A White	A Afr. Ame.	A Hispanic	A Rooted German	A East. Eur.	A Turkish
	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Var.: transfers from 1st mover to 2nd mover in the trust game						
A Top 20%	0.191	-0.0704	0.0305	0.992***	-2.552*	0.154
	(0.338)	(1.048)	(0.929)	(0.291)	(1.372)	(1.485)
A and B belong to the same ethnic group	0.0426	0.305**	0.0158	0.239***	-0.177	-0.0475
	(0.0377)	(0.151)	(0.104)	(0.0429)	(0.116)	(0.187)
Constant	5.064***	3.127***	5.508***	3.702***	6.383**	1.141
	(0.473)	(1.116)	(1.316)	(0.625)	(2.786)	(2.346)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2337	366	369	2664	216	201
R^2	0.033	0.051	0.072	0.044	0.252	0.104
p-value of test H0: Income ingroup bias = Ethnic ingroup bias	0.662	0.726	0.988	0.011	0.087	0.894

Standard errors clustered at the individual level in parentheses

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

Notes: Results of pooled OLS regressions using all three decisions made by each first mover in the ethnic and income block of the Trust games in the US sample (columns 1 to 3) and the German sample (columns 4 to 6). The dependent variable is the level of transfers from the first mover to the second mover. The explanatory variables are binary variables for whether the first mover has an income within the top 20% of the distribution and whether A and B belong to the same ethnic groups. In the Ethnic plus Income block, all second movers are in the top 20% of the income distribution. We additionally control for a variable indicating whether the observation corresponds to the first round of the block to capture any first round effects. Subject-level control variables include gender, 4 age, 2 education level and 3 employment category binary variables. The last row of the table reports the p-value of a t-test comparing the size of the income ingroup bias (A top 20% coefficient) with the size of the ethnic ingroup bias (A and B belong to the same ethnic group coefficient).

Table 4.7 – In group bias based on income

	(1) All	(2) US	(3) Germany
Dep. Var.: transfers from 1st mover to 2nd mover in the trust game			
A Top 20%	0.223	0.125	0.368*
	(0.160)	(0.239)	(0.214)
B top 20%	-1.183***	-1.095***	-1.282***
	(0.0590)	(0.0784)	(0.0886)
B top 20% X A Top 20%	0.297**	0.108	0.476***
	(0.127)	(0.193)	(0.171)
A and B belong to the same ethnic group	0.191***	0.130***	0.252***
	(0.0346)	(0.0476)	(0.0497)
Constant	5.575***	5.720***	5.314***
	(0.276)	(0.346)	(0.502)
Controls	Yes	Yes	Yes
Observations	13188	6540	6648
R^2	0.051	0.049	0.067
p-value of test H0:			
Income ingroup bias = Ethnic ingroup bias	0.419	0.911	0.207

Standard errors clustered at the individual level in parentheses

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

Notes: Results of pooled OLS regressions using all six decisions made by each first mover in the Ethnic block and the ethnic and income block of the Trust games. The dependent variable is the level of transfers from the first mover to the second mover. All regressions include the following binary variables. A belongs to Top 20% is equal to 1 if the first mover is from the top 20% of the income distribution. B top 20% is equal to 1 if the decision was made in the ethnic and income block (second mover is rich). The third variable is an interaction of the two. The next variable is an ethnic ingroup variable equal to 1 if the first mover and the second mover belong to the same ethnic group. We additionally control for two binary variables indicating whether the observation corresponds to the first round of the Ethnic block and the Ethnic plus Income blocks, to capture any first round effects. Subject-level control variables include gender, 4 age, 2 income level, 2 education level and 3 employment category binary variables. Column 1 considers the US and Germany jointly, columns 2 and 3 focus on the US and German samples respectively. The last row of the table reports the p-value of a t-test comparing the two ingroup bias coefficients (A and B belong to the same ethnic group and the income-group bias, i.e. the interaction between B top 20% and A top 20%.)

To deal with this issue, we can compare the ethnic bias with the income bias by pooling the results of the 6 TGs from the Ethnic block and the Ethnic plus Income block. Do rich first movers transfer higher amounts in the Ethnic plus Income block compared to the Ethnic block, relative to the amount transferred by bottom 80% first movers? Table 4.7 shows pooled OLS regressions using all six decisions made by each first movers in the TGs. The large and negative coefficient of the variable "B top 20%" indicates that participants decrease on average their transfers between both blocks. In the US, the null effect of the interaction term "B top 20% X A top 20%" indicates that first movers with different income levels do not behave too differently. However, in Germany, the behavior of rich vs non-rich first movers is more polarized. They all tend to decrease their transfers when the second mover is among the top 20% of the income distribution, but rich participants decrease their transfers much less. On average, in Germany, top 20% first movers decrease their transfers by about 0.8 euros between the two blocks, while first movers within the bottom 80% of the national income distribution

decrease their transfers by 1.29 euros. The difference is significant at the 1% level.[21]

We run a test for the equality of the income inequality bias coefficient (the interaction between A bottom 80% and B Top 20%) and the ethnic outgroup bias coefficient. Although the income bias is almost twice as large as the ethnic bias in Germany, the test for equality of coefficients does not find the difference significant at conventional levels.

We conclude:

Result 6a *Rich rooted Germans transfer significantly more to fellow rich rooted Germans than do non-rich rooted Germans. This "income-based" ingroup bias is economically but not statistically significantly larger in magnitude than is the ethnic outgroup bias, and does not hold for other ethnic groups.*

Result 6b *There is no statistically significant ingroup bias based on income in the US.*

Gender Effects

Romano et al. (2017), Fershtman and Gneezy (2001) and Ahmed and Ahmed (2010) all find significant gender effects, in that men display a larger ingroup bias than women. The meta-analysis by Balliet et al. (2014) demonstrates that studies containing more (if not all) men than women yield larger intergroup discrimination. We also find evidence of a gender effect in our experiments. As shown in Table A4, overall men have a significantly larger ingroup bias than women. However, this result is only significant in the US and in the Ethnic block.

4.4 Discussion

Our analysis has shown substantial variation in ingroup bias across countries and across conditions. The first obvious question to a study of this kind concerns the external validity of our results. To what extent can we be sure that our results reflect attitudes and preferences that hold outside of the research medium that we used? This is of course a very general question. Some have expressed concerns about whether experiments that are not run in natural conditions can be generalized (Levitt and List, 2007). Others have argued that experiments permit incentive-compatible elicitation of human preferences (Falk and Heckman, 2009), and that experiments' results generalize to field settings (Camerer, 2015; Herbst and Mas, 2015). Ultimately, we share the view of Falk and Heckman (2009) that even if non-natural experiments may suffer from so-called experimenter demand effects or social desirability biases, they are nonetheless valid instruments of research that should complement evidence coming from other methods.[22] Moreover, our findings are likely to be of greater general validity for the issues we study than are average lab experiments since they are obtained using samples that are more representative with respect to age, occupation and income than is often the case.

A striking result of our study is the higher rate of discrimination observed in Germany compared to the US in our experiment. Can we infer that this pattern of preferences also holds in reality? Or could it be the case that US participants are more sensitive to social desirability than the Germans,

[21] This income ingroup bias in Germany is driven by first movers from the ethnic majority and is stronger if the second mover is from a different ethnic group (not shown). This is consistent with the positive Normalized High-Income Bias (deserving rich ethnic minority effect) we documented above.

[22] Experimenter demand effects and social desirability bias can be defined as the tendency of participants in surveys or experiments to manifest patterns behavior that they perceive as being expected by the experimenter, or socially desirable (Zizzo, 2010; Quidt et al., 2018).

possibly because the social stigma – or internal guilt – associated with manifesting discriminatory preferences toward the groups in question may be higher in the US than Germany?

First of all, we have to bear in mind the big differences in the history of inter-ethnic relationship in the US vis-a-vis Germany. The Black/White racial divide has a complex history since the arrival of African Americans' as slaves in the 17th century and the mixed record of emancipation and integration into the society as equal members. Tensions over large scale Hispanic immigration and American Whites' discomfort over widespread use of Spanish in Latino communications are more recent and more complex to deconstruct. The immigration of Turkish and Eastern European began much later, in the second half of the 20th century, and the reasons were partially different. Turkish people were mainly "guest workers" migrating for economic reasons, while some – though not all – Eastern Europeans escaped war zones after the break-up of Yugoslavia. We may conjecture that in the US survey respondents may feel more restrained in revealing their real attitudes than people in Germany, given possibly different ideas of what is politically correct. Or Germans may be less concerned with their self-image than Americans. In what follows, we address these concerns further analyzing our data, although we hasten to say that we can only offer speculative answers.

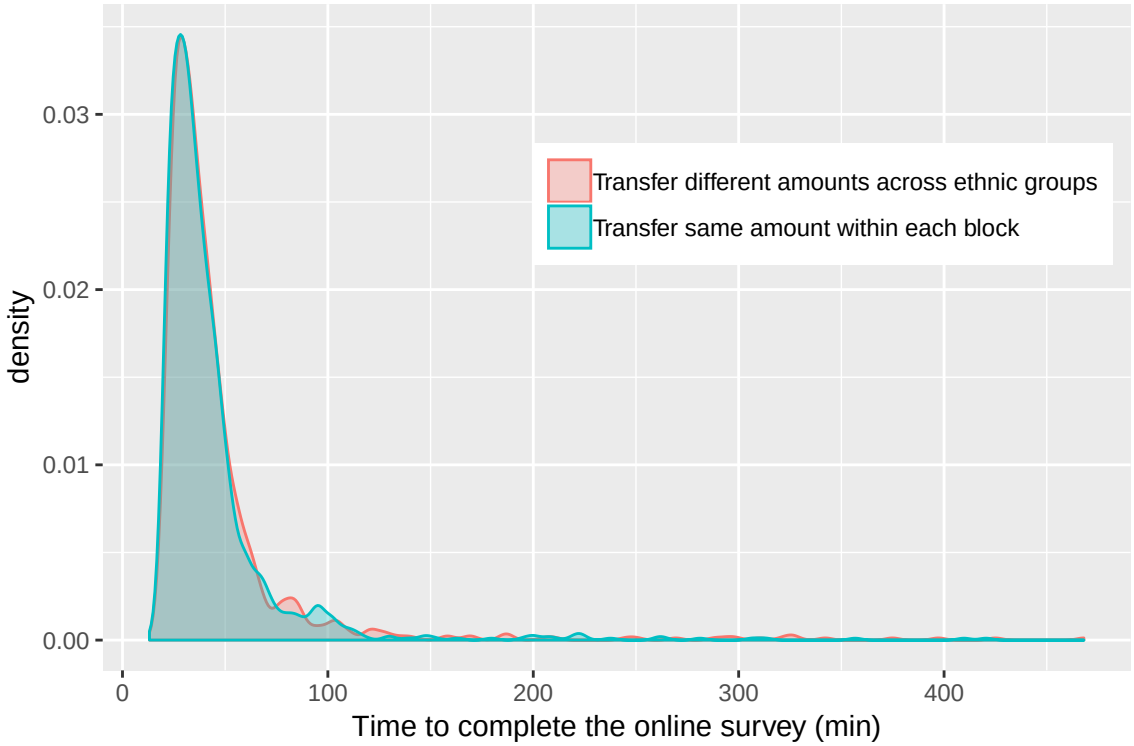
By construction, non-discriminating individuals are those who transfer the same amount of money to second movers, regardless of their ethnicity. In fact, we find a significantly larger percentage of US participants (76%) who transfer exactly the same amount to second movers, regardless of their ethnicity, in comparison to German ones (56%), as shown in the last lines of Appendix Tables 4.B1 and 4.B2.

A preliminary conjecture is that this result is caused by differing degrees of attention. One could think that some first movers were less attentive and selected the same transfer to finish the survey early. We would then underestimate the extent of discrimination. We can discard this first explanation. Figure 4.6 shows the distribution of the total time used to complete the whole *Trustlab* survey divided by people choosing (at least once) different amounts across ethnic groups and subjects sticking to the same transfers within each block. The distributions are nearly identical, suggesting that both types of subject do not differ in the total time they took to answer the survey. This is also the case if we plot the distributions separately for both countries (not shown). A more formal test confirms this visual interpretation. We cannot reject the null hypothesis of no difference between both distributions using a Kolmogorov-Smirnov test on the whole sample (p-value = 0.183). And this holds true when running the Kolmogorov-Smirnov test separately on the US sample and on the German sample.

Another potential explanation is that subjects always choosing the same transfers may still be racially biased but hide their true preferences from the experimenter, out of image concerns and social desirability bias. One could argue, however, that the social distance with the experimenter is higher in an online experiment than a laboratory one, which should diminish social desirability bias. We show in Appendix 4.C that differences in TG transfer decisions correlate in an intuitive way with answers to traditional questions of *Trustlab*'s survey module. For instance, we show that people declaring a low racial bias in survey questions are also more likely to send the same transfers to all second movers, regardless of their ethnicity. Of course this piece of evidence cannot prove that the extent of people hiding their discriminatory attitudes was limited. At best, it shows some "consistency" in the way "hiding" discriminators behaved throughout the survey.

The fact that rooted Germans are selective in their bias, discriminating more against Turkish than Eastern Europeans, is after all not too surprising. The growing support of political parties and social

Figure 4.6 – Time to complete the survey by type of transfer decisions



Notes: The figures show the distribution of the time to complete the whole *Trustlab* survey, in minutes, keeping only duration smaller than 500 minutes and pooling over the US and German sample. The graph plots two distributions: the blue one corresponds to the 1202 subjects that always transferred the exact same amount within each block. The red one is for the 996 subjects that sent at least once different amounts across two second movers (in both blocks).

movements that put at the core of their agenda the protection of the country against the perceived risk of 'islamization", in Germany as well as in other European countries, mirrors the ingroup biases that we observe in our experiment. It suggests that social desirabilty bias may be limited in Germany. Moreover, a higher ingroup bias in Germany than the US also emerges in Romano et al. (2017). It could be the case that the US data only show the lower bound of discrimination, thus enlarging the difference between the US and Germany. Even so, the size of the difference is such that it is implausible that it is exclusively caused by differences in social desirability biases or in views of "political correctness" in the two countries.

One may also question the lack of selective ingroup bias in the US. We expected higher White Americans' discrimination against African Americans than Hispanics. Even in this case, because of historical reasons, social desirability bias might demand treating African Americans on a par with Hispanics. We cannot oppose any firm counterargument to this objection, except for noting that we do find some form of selective ingroup bias in the Ethnic plus Income block even among White Americans. Ultimately, we believe that our research has uncovered previously unexplored patterns of ingroup bias, and that further research should refine our understanding of these topics.

Our design in the Ethnic plus Income block was inspired by the idea that discrimination is often associated with the belief that ethnic minorities are relatively unsuccessful in economic activities. We showed that discrimination is significantly reduced when people from the ethnic majorities are faced with successful ethnic minorities. This finding is potentially relevant for policy. It suggests that spreading stories of economic success by ethnic minorities may be an effective way to reduce

discrimination, because it would contribute to reduce the stereotype that ethnic minorities inevitably fall into the categories of the "undeserving poor". Stories of individual success may be spread at various levels – by the press, the government, at schools, and through the entertainment industry – e.g. in TV serials or films. Shaping more desirable individual behavior through the portrayal of "success stories" or "role models" has been proven to be an effective way to, for instance, improve healthy behavior, increase financial literacy, and knowledge about a public works program (Banerjee et al., 2017, 2019; Ravallion et al., 2015; Berg and Zia, 2017). We posit that the same could be the case with respect to inter-ethnic discrimination. This manipulation could be a way to break the "vicious circle" conjectured by Adida et al. (2010) to explain patterns of discrimination in Christian-heritage societies. In their analysis, rooted ethnic majorities discriminate against ethnic minorities – especially Muslims – because of perceived lack of willingness to integrate, and ethnic minorities do not integrate because of perceived discrimination.

There are two important caveats in our policy recommendation. First, even if people may accept that some individuals from ethnic minorities have been successful, they may still refrain from updating their beliefs about the group as a whole. As shown in psychology research (Kunda and Oleson, 1995; Yzerbyt et al., 1997), when individuals are confronted with new evidence purportedly contradicting previously held stereotypes about a group, they might further typify the group into "virtuous" members and "non-virtuous" members, so that the additional evidence does not lead to significant belief updates. For instance, the group of successful people from ethnic minorities may be typified as the group of "business-people in the ethnic minority". The positive treatment reserved to this group may fail to extend to the rest of the ethnic minority. This resistance to removing stereotyping may reduce the size of the effect on discrimination reduction. However, it offers a basis to update beliefs in a direction consistent with reduction of discrimination. The second caveat, which was unexpected, concerns the observation that ethnic minorities in Germany seem not to trust successful people of their own ethnic group. This may point to a breakdown of internal trust and internal cohesion if "role model" or "success stories" from ethnic minorities are diffused through the media. Our experiments could not analyze the underlying reasons for this behavior. We do not know if it was due to envy, or to genuine mistrust toward co-ethnic rich, or to some other factor. It also has to be added that, while Turkish identity was arguably a meaningful source of attachment for participants in our study, it is probably less the case for Eastern European descent subjects, given that this group encompasses many different national groups for whom a shared sense of identity may be less developed or lacking altogether. In this case, too, more research is needed to quantify and understand the nature of this effect and whether it may disrupt social capital within ethnic minorities.

Another result that stands out is the significantly lower trust and trustworthiness by the only recognizable Muslim group in our sample – Turkish in Germany. Even if this result is only partially anticipated by others and only plays a small part in discrimination by the ethnic majority, it signals the persistence of significant cultural differences between Turkish people and the rest of the population. This confirms results from other studies showing sluggishness in cultural convergence between immigrant groups and natives (Guiso et al., 2006; Bigoni et al., 2019), or between different cultures within a country (Guiso et al., 2016). Cultural convergence by Muslims seems to be particularly slow (Bisin et al., 2008; Adida et al., 2014). A conjecture to explain such sluggishness is that immigrants' social networks, instead of including people from the ethnic majority, remain mainly confined to other immigrants. Since both trust and trustworthiness are essential components of social capital (Putnam,

2007), increasing trustworthiness by immigrant groups would seem a goal worth pursuing. However, it is not clear how this can be achieved, except for stimulating interactions across different ethnic groups.

4.5 Conclusions

The goal of this paper has been to investigate levels of inter-ethnic trust in the US and Germany, focusing on the three main ethnic groups within each country. The general picture that emerges is one of substantial differences in inter-ethnic relations, both cross-country and within-country. Between countries, we document greater ethnic discrimination in Germany than in the US.

Within countries, while all ethnic groups, except Eastern Europeans living in Germany, show an ingroup bias, there are large differences in the extent and direction of the bias. In Germany, the ethnic majority holds the strongest ingroup bias and discriminates against Turkish descent subjects about 1.75 times more than against Eastern Europeans. In turn, Eastern Europeans and Turkish people favor their ingroup over the other ethnic minority but do not discriminate against rooted Germans. In the US, African Americans have the largest ingroup bias, but White and Hispanic first movers' ingroup biases are significant, albeit smaller. Unlike Germany, American Whites do not treat members of the two minority groups differently from one another. We show that in both countries, about 80% of the discrimination is taste-based rather than based on differing expectations of trustworthiness.

Our design further tests how interacting with successful ethnic minorities may reduce the bias of the ethnic majority. In the Ethnic plus Income block, we match first movers to second movers within the top 20% of the income distribution and let the ethnic group of the second mover vary. Our treatment acts like an exogenous shock on the negative stereotypes from which ethnic minorities often suffer. The treatment successfully changes behavior of rooted Germans, as they reward rich Turkish substantially more than rich Eastern Europeans. This treatment also reduces White Americans' ingroup bias, but the effect is smaller and seems to favor more Hispanics than African Americans. However, the treatment leads to the unexpected result that ethnic minorities in Germany distrust rich people from their own ethnic group. Finally, we document significant and large favoritism by rich rooted Germans toward fellow rich rooted Germans in comparison with low-income rooted Germans. This effect does not extend to other ethnic groups.

Overall, the variety of inter-ethnic relations presented in this study and the differences in the response to our intervention suggest that the policy-maker's task, if aiming to reduce to reduce discrimination, increase social cohesion, and tame the populist surge, entails complex challenges. First of all, discrimination and ingroup bias is ubiquitous and common not just to ethnic majorities, but also ethnic minorities – with the exception of Eastern Europeans in our sample. Second, discrimination can be selective, either in its basic form or in the response to an intervention. Hence, the policy-maker should have a clear map of the actual state of inter-ethnic relationships and might want to differentiate the type of intervention depending on the ethnicities involved. Third, the behavior of Turkish people and rooted Germans is reminiscent of the vicious circle proposed by Adida et al. (2014) to explain inter-ethnic relations in France. Many rooted Germans mistrust Turkish people, mainly out of taste-based rather than statistical discrimination. In turn, Turkish people are at the same time less trusting and trustworthy compared to other ethnic groups. A similar situation occurs in the US with respect to African Americans, albeit less markedly. African Americans trust others much less than

other ethnic groups, although their level of trustworthiness is on a par with others. The persistence of these behavioral differences, which arguably can be construed in terms of cultural differences, is worrying and can provide an easy ground for populist messages to spread. There is nonetheless some hope. Our intervention of showing examples of success from the ethnic minority does reduce discrimination substantially and, in Germany, benefits especially Turkish people. As argued in the previous section, more testing is needed to ascertain whether this intervention leads to the generalized removal of stereotypical views that ethnic minorities lack work ethic, or remains limited to successful people. Moreover, even if this intervention is successful for the ethnic majority, it may disrupt social cohesion within the ethnic minority.

As argued by Putnam (2007), inter-group relationships can move at a glacial pace over time, but they can undoubtedly improve. This is however not a foregone conclusion. The populist surge represents a setback in this progress by making ethnic demarcations more salient. Our paper has contributed to this debate by bringing novel empirical evidence on the mapping of inter-ethnic relationships in two large Western countries, and discussing possible avenues for interventions.

Bibliography

Aassve, Arnstein, Pierluigi Conzo, and Francesco Mattioli. Was Banfield right ? New insights from a nationwide laboratory experiment. 2018a.

Aassve, Arnstein, Gianmarco Daniele, and Marco Le Moglie. Never Forget the First Time : The Persistent Effects of Corruption and the Rise of Populism in Italy. 2018b.

Adida, Claire L., David D. Laitin, and Marie Anne Valfort. Identifying barriers to muslim integration in France. *Proceedings of the National Academy of Sciences of the United States of America*, 107 (52):22384–22390, 2010. ISSN 00278424. doi: 10.1073/pnas.1015550107.

Adida, Claire L., David D. Laitin, and Marie Anne Valfort. Muslims in France: Identifying a discriminatory equilibrium. *Journal of Population Economics*, 27(4):1039–1086, 2014.

Ahmed, Ali M and Ali M Ahmed. What is in a surname ? The role of ethnicity in economic decision making. *Applied Economics*, 42:2715–2723, 2010.

Akai, Kenju and Robert Jiro. Trust and reciprocity among international groups : Experimental evidence from Austria and Japan. *Journal of Socio-Economics*, 41(2):266–276, 2012.

Alesina, Alberto and Eliana La Ferrara. Participation in Heterogeneous Communities. *The Quarterly Journal of Economics*, 2000.

Alesina, Alberto and Stefanie Stantcheva. Diversity, Immigration, and Redistribution. *National Bureau of Economic Research*, 2020.

Alesina, Alberto, Armando Miano, and Stefanie Stantcheva. Immigration and Redistribution. *NBER Working Paper 24733*, 2018a.

Alesina, Alberto, Stefanie Stantcheva, and Edoardo Teso. Intergenerational Mobility and Preferences for Redistribution. *American Economic Review*, 108(2):521–554, 2018b.

Algan, Yann, Camille Hémet, and David D. Laitin. The Social Effects of Ethnic Diversity at the Local Level: A Natural Experiment with Exogenous Residential Allocation. *Journal of Political Economy*, 124(3):696–733, 2016.

Algan, Yann, Sergei Guriev, Elias Papaioannoun, and Evgenia Passari. The European Trust Crisis and the Rise of Populism. *Brookings Papers on Economic Activity*, (Fall):309–382, 2017.

Algan, Yann, Elizabeth Beasley, Daniel Cohen, and Martial Foucault. *Les Origines du populisme*. Le Seuil, 2019.

Allport, Gordon Willard, Kenneth Clark, and Thomas Pettigrew. *The nature of prejudice*. Addison-wesley Reading, MA, 1954.

Anderson, Lisa R., Jennifer M. Mellor, and Jeffrey Milyo. Induced heterogeneity in trust experiments. *Experimental Economics*, 9(3):223–235, 2006.

Balliet, Daniel, Junhui Wu, and Carsten K.W. De Dreu. Ingroup favoritism in cooperation: a meta-analysis. *Psychological bulletin*, 140(6):1556–1581, 2014.

Banerjee, Abhijit, Sharon Barnhardt, and Esther Duflo. Movies, Margins, and Marketing Encouraging the Adoption of Iron-Fortified Salt. *NBER*, (March):285–306, 2017.

Banerjee, Abhijit, Eliana La Ferrara, and V. H. Orozco-Olvera. The entertaining way to behavioral change: Fighting HIV with MTV. *The World Bank*, 2019.

Berg, Gunhild and Bilal Zia. Harnessing emotional connections to improve financial decisions: Evaluating the impact of financial education in mainstream media. *Journal of the European Economic Association*, 15(5):1025–1055, 2017.

Berg, Joyce, John Dickhaut, and Kevin McCabe. Trust, Reciprocity, and Social History, 1995. ISSN 08998256.

Bernhard, Helen, Urs Fischbacher, and Ernst Fehr. Parochial altruism in humans. *Nature*, 442(7105): 912–915, 2006.

Bertrand, Marianne and Sendhil Mullainathan. Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination. *American Economic Review*, 94 (4):991–1013, 2004.

Betz, Hans-georg. The Radical Right and Populism. *Oxford Hanbook of the Radical Right*, (April): 1–22, 2018.

Bigoni, Maria, Stefania Bortolotti, Marco Casari, and Diego Gambetta. At the root of the North–South cooperation gap in Italy: Preferences or beliefs? *The Economic Journal*, 129(619):1139—-1152, 2019.

Bisin, Alberto, Eleonora Patacchini, Thierry Verdier, and Yves Zenou. Are Muslim immigrants different in terms of cultural integration? *Journal of the European Economic Association*, 6(2-3): 445–456, 2008.

Bogliacino, Francesco, Laura Jimenez Lozano, and Daniel Reyes. Socioeconomic stratification and stereotyping: lab-in-the-field evidence from Colombia. *International Review of Economics*, 65(1): 77—-118, 2018.

Bouckaert, Jan and Geert Dhaene. Inter-ethnic trust and reciprocity: Results of an experiment with small businessmen. *European Journal of Political Economy*, 20(4):869–886, 2004.

Brewer, Marilynn B. The Psychology of Prejudice : Ingroup Love or Outgroup Hate ? *Journal of Social*, 55(3):429–444, 1999.

Buchan, Nancy R, Gianluca Grimalda, Rick Wilson, Marilynn Brewer, Enrique Fatas, and Margaret Foddy. Globalization and human cooperation. pages 1–5, 2009.

Burns, Justine. Race, diversity and pro-social behavior in a segmented society. *Journal of Economic Behavior and Organization*, 81(2):366–378, 2012.

Camerer, Colin F. The Promise and Success of Lab-Field Generalizability in Experimental Economics: A Critical Reply to Levitt and List. *Handbook of experimental economic*, 2015.

Cettolin, Elena and Sigrid Suetens. Return on Trust is Lower for Immigrants. *The Economic Journal*, 2019.

Charness, By Gary, Luca Rigotti, and Aldo Rustichini. Individual Behavior and Group Membership. *American Economic Review*, 97(4):1340–1352, 2007.

Chen, By Yan and Sherry Xin Li. Group Identity and Social Preferences. *American Economic Review*, 2005:431–457, 2009.

Chuah, Swee Hoon, Reema Fahoum, and Robert Hoffmann. Fractionalization and trust in India: A field-experiment. *Economics Letters*, 119(2):191–194, 2013.

Chuah, Swee-hoon, Robert Hoffmann, Bala Ramasamy, and Jonathan H W Tan. Religion , ethnicity and cooperation : An experimental study. *Journal of Economic Psychology*, 45:33–43, 2014.

Cox, James C. How to identify trust and reciprocity. *Games and Economic Behavior*, 46(2):260–281, 2004.

Criado, Henar, Francisco Herreros, Luis Miller, and Paloma Ubeda. Ethnicity and Trust : A Multifactorial Experiment. *Political Studies*, 63:131–152, 2015.

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.

Degli Antoni, Giacomo and Gianluca Grimalda. Groups and trust: Experimental evidence on the Olson and Putnam hypotheses. *Journal of Behavioral and Experimental Economics*, 61:38–54, 2016.

Dorrough, Angela Rachael and Andreas Glöckner. Multinational investigation of cross-societal cooperation. *Proceedings of the National Academy of Sciences*, 113(39):10836–10841, 2016.

Dragolov, Georgi, Zsofia Ignacz, Jan Lorenz, Jan Delhey, and Klaus Boehnke. Social Cohesion Radar, Measuring Common Ground: An International Comparison of Social Cohesion. Methods Report. pages 1–55, 2014.

Falk, A and J J Heckman. Lab Experiments Are a Major Source of Knowledge in the Social Sciences. *Science*, 326(5952):535–538, 2009.

Falk, Armin and Christian Zehnder. A city-wide experiment on trust discrimination. *Journal of Public Economics*, 100:15–27, 2013.

Felfe, Christina, Martin Kocher, Helmut Rainer, Judith Saurer, and Thomas Siedler. Immigration, Naturalization, and Discrimination: Combining a Natural Experiment with a Large-Scale Trust Experiment in Schools. 2017.

Fershtman, Chaim and Uri Gneezy. Discrimination in a segmented society: An experimental approach. *Quarterly Journal of Economics*, 116(1):351–377, 2001.

Fershtman, Chaim, Uri Gneezy, and Frank Verboven. Discrimination and Nepotism: The Efficiency of the Anonymity Rule. *The Journal of Legal Studies*, 34(2):371–396, 2005.

Georgantzis, N., J. A. Lacomba, F. Lagos, and J. Milgram. Trust and reciprocity between neighbour countries: Morocco, France and Spain. *Revista Internacional de Sociologia*, 76(1), 2018.

Gilens, Martin. *Why Americans hate welfare: Race, media, and the politics of antipoverty policy*. University of Chicago Press, 2009.

Goerg, Sebastian J., Heike Hennig-Schmidt, Gari Walkowitz, and Eyal Winter. In Wrong anticipation – Miscalibrated beliefs between germans, israelis, and palestinians. *PLoS ONE*, 11(6):1–16, 2016.

Goette, Lorenz, David Huffman, Stephan Meier, By Lorenz Goette, and David Huffman. The impact of Group Membership on Cooperation and Norm: Evidence Using Random Assignment to Real Enforcement : Social Groups. *American Economic Review*, 96(2):212–216, 2006.

Greiner, Ben, Axel Ockenfels, and Peter Werner. The dynamic interplay of inequality and trust-An experimental study. *Journal of Economic Behavior and Organization*, 81(2):355–365, 2012.

Guillen, Pablo and Daniel Ji. Trust, discrimination and acculturation. Experimental evidence on Asian international and Australian domestic university students. *Journal of Socio-Economics*, 40 (5):594–608, 2011.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales. Does culture affect economic outcomes? *Journal of Economic Perspectives*, 20(2):23–48, 2006. ISSN 08953309. doi: 10.1257/jep.20.2.23.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales. Long-Term Persistence. *Journal of the European Economic Association*, 14(6):1401–1436, 2016.

Guriev, Sergei. Economic Drivers of Populism. *AEA Papers and Proceedings*, pages 200–203, 2018.

Güth, Werner, Matteo Ploner, and Tobias Regner. Determinants of in-group bias : Is group affiliation mediated by guilt-aversion ? *Journal of Economic Psychology*, 30(5):814–827, 2009.

Habyarimana, James, Macartan Humphreys, Daniel N Posner, and Jeremy M Weinstein. Why Does Ethnic Diversity Undermine Public Goods Provision? *The American Political Science Review*, 101 (4):709–725, 2007.

Hargreaves, Shaun P and Daniel John Zizzo. The Value of Groups. *The American Economic Review*, 99(1):295–323, 2009.

Herbst, D. and A. Mas. Peer effects on worker output in the laboratory generalize to the field. *Science*, 350(6260):545—-549, 2015.

Inglehart, Ronald F and Pippa Norris. Trump, Brexit, and the Rise of Populism: Economic Have-Nots and Cultural Backlash. 2016.

Johansson-Stenman, Olof, Minhaj Mahmud, and Peter Martinsson. Trust and religion: Experimental evidence from Rural Bangladesh. *Economica*, 76(303):462–485, 2009.

Kunda, Ziva and Kathryn C Oleson. Maintaining Stereotypes in the Face of Disconfirmation. *Journal of Personality and Social Psychology*, 68(4), 1995.

Kuziemko, Ilyana, Michael I. Norton, Emmanuel Saez, and Stefanie Stantcheva. How elastic are preferences for redistribution? Evidence from randomized survey experiments. *American Economic Review*, 105(4):1478–1508, 2015.

Lane, Tom. Discrimination in the laboratory: A meta-analysis of economics experiments. *European Economic Review*, 90:375–402, 2016.

Lei, Vivian and Filip Vesely. In-Group versus Out-Group Trust: The Impact of Income Inequality. *Southern Economic Journal*, 76(4):1049–1063, 2010.

Levitt, Steven D and John A List. What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World? *Journal of Economic Perspectives*, 21(2):153–174, 2007.

Mai, Robert, Tobias Stöhr, and Philipp Wichardt. Signs of Integration Affect Helping Behavior Towards Immigrants. 2019.

Murtin, Fabrice, Lara Fleischer, Vincent Siegerink, Arnstein Aassve, Yann Algan, Santiago Gonzalez, Gianluca Grimalda, Rafael Hortala, Vallve Soonhee, Kim David, Lee Zsuzsanna, Lonti Louis, Putterman Ulrich, and Schmidt Conal. Trust and its Determinants : Evidence from the Trustlab Experiment. 2018.

Norris, Pippa and Ronald Inglehart. *Cultural backlash: Trump, Brexit, and authoritarian populism*. Cambridge University Press, 2019.

Putnam, Robert D. E pluribus unum: Diversity and community in the twenty-first century the 2006 johan skytte prize lecture. *Scandinavian Political Studies*, 30(2):137–174, 2007.

Quidt, By Jonathan De, Johannes Haushofer, and Christopher Roth. Measuring and Bounding Experimenter Demand. 108(11):3266–3302, 2018.

Rand, David G, Thomas Pfeiffer, Anna Dreber, Rachel W Sheketoff, Nils C Wernerfelt, and Yochai Benkler. Dynamic remodeling of in-group bias during the 2008 presidential election. *Proceedings of the National Academy of Sciences*, 106(15):6187–6191, 2009.

Ravallion, Martin, Dominique Van De Walle, Puja Dutta, and Rinku Murgai. Empowering poor people through public information ? Lessons from a movie in rural India. *Journal of Public Economics*, 132:13–22, 2015.

Romano, Angelo, Daniel Balliet, Toshio Yamagishi, and James H. Liu. Parochial trust and cooperation across 17 societies. *Proceedings of the National Academy of Sciences*, 114(48):12702–12707, 2017.

Ruffle, Bradley J. and Richard Sosis. Cooperation and the in-group-out-group bias: A field test on Israeli kibbutz members and city residents. *Journal of Economic Behavior and Organization*, 60(2): 147–163, 2006.

Simpson, Brent, Tucker McGrimmon, and Kyle Irwin. Are blacks really less trusting than whites? Revisiting the race and trust question. *Social Forces*, 86(2):525–552, 2007.

Smith, Alexander. Income inequality in the trust game. *Economics Letters*, 111(1):54–56, 2011.

Tajfel, Henri, M.G. Billig, R.P. Bundy, and Claude Flament. Social categorization and intergroup behaviour. *European Journal of Social Psychology*, 1(2):149–178, 1971.

Tanaka, Tomomi and Colin F. Camerer. Trait perceptions influence economic out-group bias: lab and field evidence from Vietnam. *Experimental Economics*, 19(3):513–534, 2016.

Tjaden, Jasper Dag, Carsten Schwemmer, and Menusch Khadjavi. Ride with Me — Ethnic Discrimination , Social Markets , and the Sharing Economy. *European Sociological Review*, 34(4):418–432, 2018.

Trifiletti, Elena and Dora Capozza. Examining Group-Based Trust with the Investment Game. *Social Behavior and Personality: an international journal*, 39(3):405–409, 2011.

Weisel, Ori and Robert Böhm. "Ingroup love" and "outgroup hate" in intergroup conflict between natural groups. *Journal of Experimental Social Psychology*, 60:110–120, 2015.

Whitt, Sam and Rick K. Wilson. The Dictator Game, Fairness and Ethnicity in Postwar Bosnia. *American Journal of Political Science*, 51(3):655–668, 2007.

Willinger, Marc, Claudia Keser, Christopher Lohmann, and Jean Claude Usunier. A comparison of trust and reciprocity between France and Germany: Experimental investigation based on the investment game. *Journal of Economic Psychology*, 24(4):447–466, 2003.

Yamagishi, Toshio, Yosuke Makimura, Margaret Foddy, Masafumi Matsuda, Toko Kiyonari, and Michael J Platow. Comparisons of Australians and Japanese on group-based cooperation. *Asian Journal of Social Psychology*, pages 173–190, 2005.

Yzerbyt, Vincent, Steve Rocher, and Georges Schadron. Stereotypes as explanations: A subjective essentialistic view of group perception. 1997.

Zhang, Zhe, Xu Zhang, and Louis Putterman. Trust and Cooperation at a Confluence of Worlds: An Experiment in Xinjiang, China. *Journal of Economic Behavior and Organization*, 161:128–144, 2019.

Zizzo, Daniel John. Experimenter demand effects in economic experiments. *Experimental Economics*, pages 75–98, 2010.

Appendices

4.A Figures

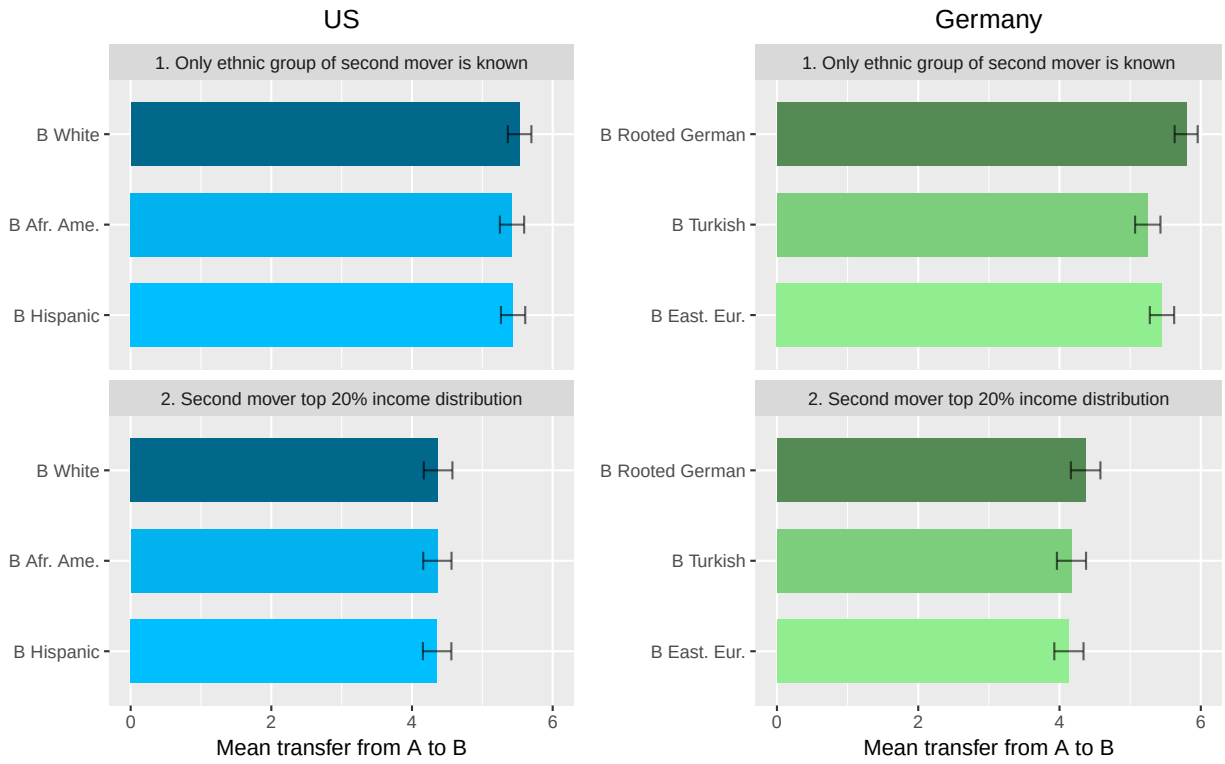


Figure 4.A1 – Transfer decisions from first mover to second mover in the US and Germany by second mover ethnic group

Notes: The bars display the average level of transfers to each type of second mover pooling over the whole US sample for the left panel and the whole German sample for the right panel. The upper part displays *Ethnic block* transfers (only ethnicity is known) and the lower part shows *Ethnic plus Income block* transfers (the second mover is additionally in the top 20% of the national income distribution). 95% confidence intervals are reported.

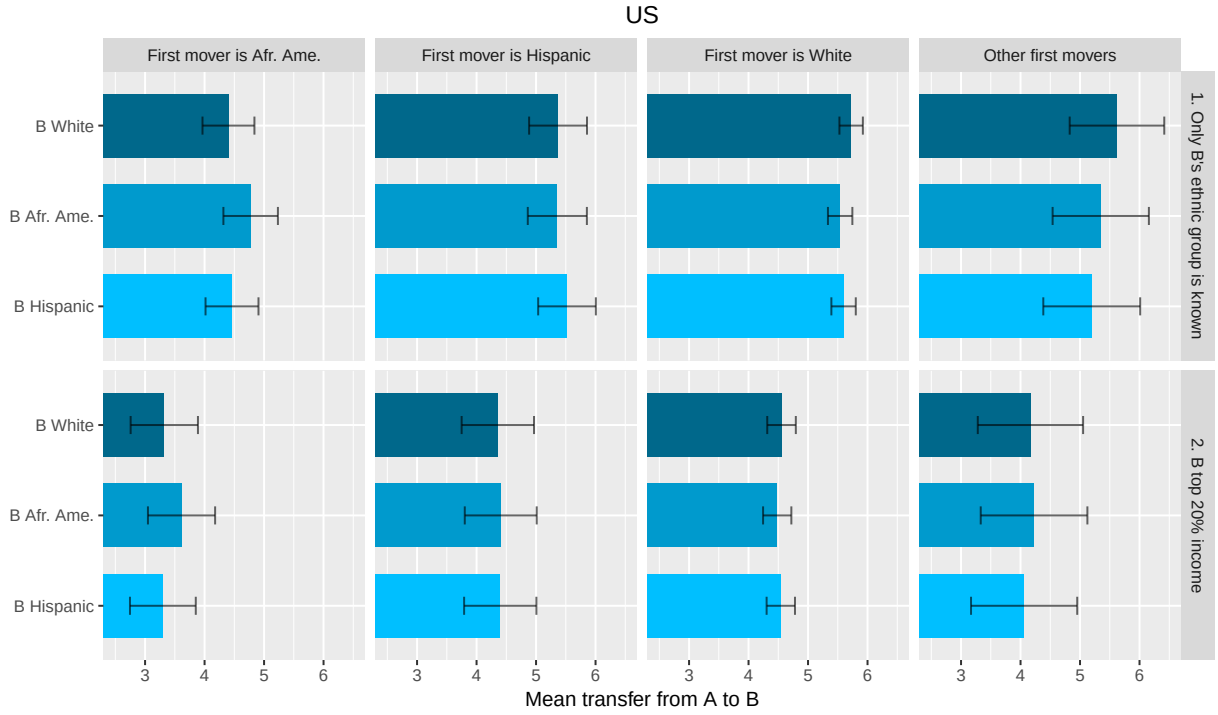


Figure 4.A2 – Transfer decisions from first mover to second mover in the US by first mover ethnicity

Notes: The bars display the average level of transfers to each type of second mover by first mover ethnic group. The upper part displays *Ethnic block* transfers (only ethnicity is known) and the lower part shows *Ethnic plus Income block* transfers (the second mover is additionally in the top 20% of the national income distribution). 95% confidence intervals are reported. The computation are made on the whole US sample. Other first movers correspond to Asian Americans, American Indian or Alaska Native, people with more than one ethnic group and other ethnic groups.

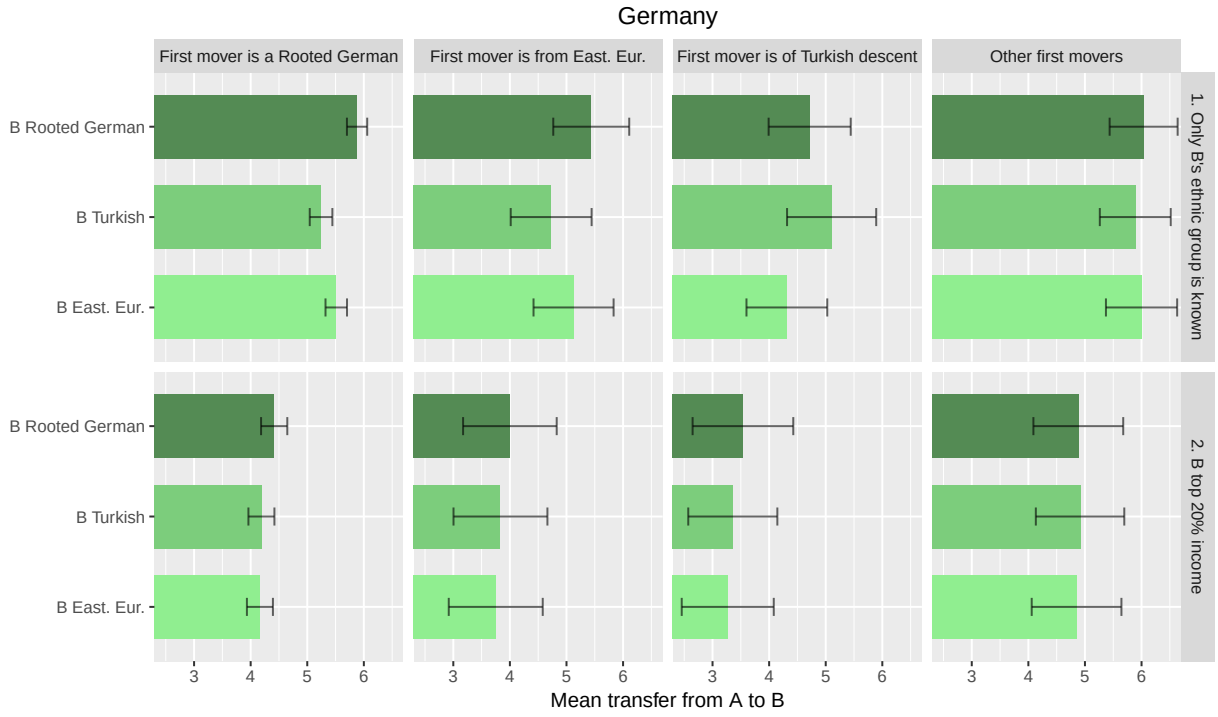


Figure 4.A3 – Transfer decisions from first mover to second mover in Germany by first mover ethnicity

*Notes:*The bars display the average level of transfers to each type of second mover by first mover ethnic group. The upper part displays *Ethnic block* transfers (only ethnicity is known) and the lower part shows *Ethnic plus Income block* transfers (the second mover is additionally in the top 20% of the national income distribution).95% confidence intervals are reported. The computation are made on the whole German sample.

4.B Tables

Table 4.B1 – Summary statistics - US

Variable	Obs	Mean	Std. Dev.	**Population mean**
Socio-demographic characteristics				
A white	1090	.715	.452	.604
A African American	1090	.112	.315	.134
A Hispanic	1090	.113	.317	.161
A Asian American	1090	.037	.188	.059
A other non-white race	1090	.024	.153	.042
Female	1090	.512	.5	.508
Age: 0-14	1090	0	0	.185
Age: 15-24	1090	.1	.3	.129
Age: 25-54	1090	.569	.495	.389
Age: 55-64	1090	.326	.469	.129
Age: 65+	1090	.006	.074	.166
A poor	1090	.498	.5	.4
A medium income cat.	1090	.195	.397	.2
A rich	1090	.306	.461	.4
High school or less	1090	.2	.4	.4
Some college	1090	.381	.486	.28
Tertiary diploma	1090	.419	.494	.32
Employed	1090	.55	.498	.57
Self-employed	1090	.076	.265	.038
Unemployed	1090	.12	.325	.024
Out of the labor force	1090	.253	.435	.37
Right wing	968	.449	.498	
Trust games behavior				
Transfer same amount in Ethnic block	1090	.806	.395	
Transfer same amount in Ethnic plus Income block	1090	.843	.364	
Always transfer the same amount within each block	1090	.763	.425	

Notes: All variables are binary. The population mean column refers to the real socio-demographic characteristics of the US. Ethnic composition and education data comes from the 2018 US Census: https://www.census.gov/quickfacts/fact/table/US/PST045218 and https://www.census.gov/data/tables/2018/demo/education-attainment/cps-detailed-tables.html. Age structure data comes from the CIA World Fact book https://www.cia.gov/library/publications/the-world-factbook/. Employment data comes from the US Bureau of Labor Statistics;: https://www.bls.gov/cps/cpsaat01.htm. All our subjects are at least 17 years old. *A poor* means that the subject's household income falls within the bottom two income quintiles of the distribution of her country, third income quintile for *A medium income cat.* and the top two income quintiles for *A rich*. Household income comes from all salaries, wages, profit from self-employment, interest, rent, pension, social insurance payments and other benefits. Income is calculated before tax but after transfers. *Right-wing* results from the dichotomization of answers to a left-right political attitude question on a 0 (far-left) to 10 (far-right) scale. A subject is considered right-wing if her answer at the political attitude question is greater than 5. *Transfer same amount in Ethnic block (Ethnic plus Income block)* is equal to 1 if the subject chooses to transfer the same amount to all Ethnic block (Ethnic plus Income block) second movers. *Always transfer the same amount within each block* is equal to 1 if the two latter variables are both equal to 1 and 0 otherwise (the level of transfers can differ between blocks as long as second movers are treated the same way within each block).

Table 4.B2 – Summary statistics - Germany

Variable	Obs	Mean	Std. Dev.	**Population mean**
Socio-demographic characteristics				
A rooted German	1108	.801	.399	.745
A East. European	1108	.065	.247	.065
A Turkish	1108	.06	.238	.034
A other non-native race	1108	.073	.26	.156
Female	1108	.503	.5	.507
Age: 0-14	1108	0	0	.129
Age: 15-24	1108	.106	.307	.098
Age: 25-54	1108	.665	.472	.389
Age: 55-64	1108	.213	.41	.157
Age: 65+	1108	.016	.126	.23
A poor	1108	.361	.481	0.4
A medium income cat.	1108	.231	.422	0.2
A rich	1108	.408	.492	0.4
High school or less	1108	.293	.455	.26
Some college	1108	.369	.483	.56
Tertiary diploma	1108	.338	.473	.18
Employed	1108	.623	.485	.65
Self-employed	1108	.07	.256	.068
Unemployed	1108	.046	.21	.024
Out of the labor force	1108	.261	.439	0.26
Right wing	1017	.296	.457	
Trust games behavior				
Transfer same amount in Ethnic block	1108	.606	.489	
Transfer same amount in Ethnic plus Income block	1108	.75	.433	
Always transfer the same amount within each block	1108	.557	.497	

Notes: All variables are binary. The population mean column refers to the real socio-demographic characteristics of Germany. Ethnic composition, education and employment data comes from the Statistisches Bundesamt: https://www.destatis.de/. Age structure data comes from the CIA World Fact book https://www.cia.gov/library/publications/the-world-factbook/. All our subjects are at least 17 years old. *A poor* means that the subject's household income falls within the bottom two income quintiles in her country, third income quintile for *A medium income cat.* and the top two income quintiles for *A rich*. Household income was elicited using the following question: *In the last 12 months, what was the total income of your household before taxes have been deducted? (Income can come salaries and wages, profit from self-employment, interest, rent, pension, social insurance payments and other benefits, among others).* *Right-wing* results from the dichotomization of answers to a left-right political attitude question on a 0 (far-left) to 10 (far-right) scale. A subject is considered right-wing if her answer at the political attitude question is greater than 5. *Transfer same amount in Ethnic block (Ethnic plus Income block)* is equal to 1 if the subject chooses to transfer the same amount to all Ethnic block (Ethnic plus Income block) second movers. *Always transfer the same amount within each block* is equal to 1 if the two latter variables are both equal to 1 and 0 otherwise (the level of transfers can differ between blocks as long as second movers are treated the same way within each block).

Table 4.B3 – Ingroup bias at the country level in the Ethnic block

	(1)	(2)	(3)
Dep. Var.: transfers from 1st mover to 2nd mover in the trust game			
A and B belong to the same ethnic group	0.290***	0.180***	0.176***
	(0.0403)	(0.0533)	(0.0532)
Germany		-0.0328	-0.0641
		(0.123)	(0.127)
Germany * A and B belong to the same ethnic group		0.219***	0.222***
		(0.0804)	(0.0801)
Constant	5.319***	5.336***	5.838***
	(0.0628)	(0.0884)	(0.279)
Controls	No	No	Yes
Observations	6594	6594	6594
R^2	0.003	0.004	0.017

Notes: Columns 1 to 3 report the results of pooled OLS regressions on the US and German samples jointly using all the three decisions made by each first mover in the *ethnic bloc* Trust games (only ethnicity of second mover is known). The dependent variable is the level of transfer from the first mover to the second mover. The explanatory variables are binary variables for whether the first and the second mover have different ethnicity, whether the first mover lives in Germany and an interaction of these two variables. We additionally control for a variable indicating whether the observation corresponds to the first second mover encountered (first transfer made in the *Ethnic block*) to capture any first round effects. Subject-level control variables (in column 3) include gender, 4 age, 2 income level, 2 education level and 3 employment category binary variables.

Table 4.B4 – Individual determinants of ingroup ethnic bias

	US		Germany	
	Ethnic block	Ethnic plus Income block	Ethnic block	Ethnic plus Income block
	(1)	(2)	(3)	(4)
Dependent variable: ingroup transfer - average outgroup transfer				
A White	-0.192	-0.242		
	(0.152)	(0.155)		
A Hispanic	-0.205	-0.281		
	(0.163)	(0.183)		
A rooted German			0.430*	0.387***
			(0.252)	(0.118)
A Turkish			0.554	0.101
			(0.370)	(0.206)
Female	-0.163**	-0.0782	-0.0630	-0.0734
	(0.0814)	(0.0740)	(0.107)	(0.0790)
Age: 18-24	-0.109	0.0402	0.0372	-0.0165
	(0.128)	(0.0663)	(0.202)	(0.137)
Age: 25-34	0.0455	0.114	-0.00914	0.000979
	(0.132)	(0.108)	(0.174)	(0.132)
Age: 35-44	0.0697	-0.00000797	-0.404***	-0.118
	(0.112)	(0.112)	(0.148)	(0.111)
Age: 45-54	0.0387	0.0171	-0.125	0.0738
	(0.115)	(0.0985)	(0.162)	(0.124)
A poor	-0.0971	0.0175	-0.0377	0.0516
	(0.105)	(0.0920)	(0.132)	(0.107)
A medium income cat.	-0.163	-0.0793	-0.107	0.0640
	(0.135)	(0.112)	(0.136)	(0.112)
High school or less	0.218*	0.0120	-0.0222	-0.0152
	(0.121)	(0.109)	(0.152)	(0.122)
Some college	0.100	0.0435	-0.108	-0.0179
	(0.0959)	(0.0869)	(0.133)	(0.0965)
Employed	0.156	0.0596	0.374	0.161
	(0.118)	(0.106)	(0.241)	(0.237)
Self-employed	0.105	0.0276	-0.0516	0.0118
	(0.176)	(0.190)	(0.317)	(0.266)
Out of the labor force	0.200*	0.0946	0.135	0.180
	(0.118)	(0.104)	(0.259)	(0.244)
Constant	0.266	0.227	0.0378	-0.275
	(0.192)	(0.176)	(0.440)	(0.287)
Observations	1024	1024	1027	1027
R^2	0.013	0.009	0.021	0.014

Standard errors clustered at the individual level in parentheses

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

Notes: The table reports the results of OLS regressions on the US sample (only White, African American and Hispanic first movers) in columns 1 and 2 and on the German sample (only rooted Germans, Eastern European descent and Turkish descent first movers) in columns 3 and 4. Odd-numbered columns use the data of the Ethnic block, when only the ethnic group of the second mover is known and even-numbered columns use the data of the Ethnic plus Income block when income information is additionally disclosed. The dependent variable is the difference in the ingroup transfer and the average outgroup transfer. For instance, for African Americans this corresponds to the difference between transfers to other African Americans and the average of the transfers to White and Hispanic second movers. The explanatory variables are binary variables for the ethnic group of the first mover. Subject-level control variables include gender, 4 age, 2 income level, 2 education level and 3 employment category binary variables. See the notes of Tables 4.B1 and 4.B2 for more information about these variables.

4.C Behavior consistency

We describe here the difference in attitudes between people sticking to the same transfer decisions and those that deviate across ethnic groups using *Trustlab*'s survey module (see Appendix Table 4.C1 for a description of the variables). Is it true that subjects always transferring the same amount also report being less racially biased in more traditional survey questions? It seems to be the case. Table 4.C2 and 4.C3 show the difference in means between the two types of behavior to several questions on trust and opinion toward immigrants in the US and Germany respectively. In both countries, subjects changing transfers and those always transferring the same amount (within each block) do not differ much in their answers to trust questions that are not racially oriented. They both report similar levels of generalized trust, trust in their family, neighborhood and people they know personally. However, switchers trust less people of another religion, of another nationality or who immigrated. This result is more pronounced in Germany than in the US.

Table 4.C1 – Description of the variables in *Trustlab*'s survey module

Variable name	Description	Scale
Generalized trust (OECD)	*In general, how much do you trust most people?*	0 = not at all; 10 = completely
Generalized trust (Rosenberg)	*Generally speaking, would you say that most people can be trusted, or that you can't be too careful in dealing with people?*	0 = You can't be too careful; 10 = Most people can be trusted
Trust in ...	*Could you tell us for each of these groups how much do you trust them?* • Family • People in your neighbourhood • People you know personally • People you meet for the first time • People of another religion • People who immigrated • People of another nationality • People who seek refuge	0 = No trust at all; 10 = I fully trust them
Immigrants exert less effort	**US:** *On average, Blacks/African Americans have worse jobs, income, housing than white people. Do you think the differences are mainly due to discrimination and disadvantages of educational opportunity, mainly due to differences in in-born ability, motivation and effort, or some combination?* **Germany** *The average immigrants have worse jobs, income, housing than native Germans. Do you think the differences are mainly due to discrimination and lack of opportunity, mainly due to differences in in-born ability, motivation and effort, or some combination?*	**US** 0 = Mainly discrimination and educational disadvantage; 10 = Mainly lesser ability, motivation and effort **Germany** 0 = Mainly discrimination and lack of opportunity; 10 = Mainly lesser ability, motivation and effort
Immigrants are well integrated	*To what extent do you agree with the following statement? Immigrants are well integrated in our society*	0 = Immigrants are not integrated in our society; 10 = Immigrants are well integrated in our society
Our culture is enriched by immigrants	*To what extent do you agree with the following statement? Our culture is enriched by immigrants*	0 = Our culture is undermined by immigrants; 10 = Our culture is enriched by immigrants
Right-wing	*In political matters, people often talk of "the left" and "the right". How would you place your views on this scale, generally speaking?*	0 = Left, 10 = right. The variable is dichotomized so that it is equal to 1 for answers greater than 5

Table 4.C2 – Attitudes by type of transfer decision - US

Variable	(1) Change transfer	(2) Same transfers	(3) Diff.	(4) Obs.
Generalized trust (OECD)	6.844	6.423	-0.422	1,083
	(2.468)	(2.200)	(0.162)***	
Generalized trust (Rosenberg)	5.808	5.850	0.042	1,075
	(2.720)	(2.343)	(0.175)	
Do you think a stranger would return your wallet?	0.414	0.458	0.043	751
	(0.494)	(0.499)	(0.041)	
Trust in your family	8.585	8.423	-0.163	1,086
	(1.911)	(1.867)	(0.134)	
Trust in people in your neighbourhood	6.063	6.161	0.097	1,061
	(2.332)	(2.200)	(0.161)	
Trust in people you know personally	7.629	7.747	0.118	1,085
	(1.893)	(1.668)	(0.123)	
Trust in people you meet for the first time	5.087	4.989	-0.098	1,061
	(2.380)	(2.079)	(0.155)	
Trust in people of another religion	5.962	6.359	0.396	1,019
	(2.297)	(1.943)	(0.150)***	
Trust in people who immigrated	6.220	6.388	0.168	1,031
	(2.419)	(1.997)	(0.154)	
Trust in people of another nationality	6.102	6.502	0.400	1,040
	(2.275)	(1.867)	(0.144)***	
Trust in people who seek refuge	6.000	5.961	-0.039	1,021
	(2.704)	(2.197)	(0.171)	
Immigrants exert less effort	6.297	5.359	-0.938	970
	(2.910)	(2.878)	(0.217)***	
Immigrants are well integrated	5.967	5.985	0.017	1,023
	(2.546)	(2.356)	(0.176)	
Our culture is enriched by immigrants	6.704	6.655	-0.049	1,039
	(2.796)	(2.671)	(0.196)	
Right wing	0.647	0.469	-0.179	1,090
	(0.479)	(0.499)	(0.035)***	
Observations	258	832		1,090

Notes: Column 1: sample of US participants sending different amounts across ethnic groups at least once either in the Ethnic block or in the Ethnic plus Income block. Column 2: sample of US participants sticking to the same transfer throughout the Ethnic block and also throughout the Ethnic plus Income block (the level of transfer can differ across blocks as long as second movers are treated the same way within a block). See appendix table 4.C1 for a description of the variables.

Table 4.C3 – Attitudes by type of transfer decision - Germany

Variable	(1) Change transfer	(2) Same transfers	(3) Diff.	(4) Obs.
Generalized trust (OECD)	6.544	6.663	0.119	1,096
	(2.157)	(2.053)	(0.128)	
Generalized trust (Rosenberg)	5.696	6.171	0.475	1,093
	(2.005)	(2.009)	(0.122)***	
Do you think a stranger would return your wallet?	0.406	0.489	0.083	764
	(0.492)	(0.500)	(0.036)**	
Trust in your family	8.921	8.882	-0.039	1,096
	(1.484)	(1.789)	(0.101)	
Trust in people in your neighbourhood	6.230	6.406	0.175	1,097
	(2.004)	(2.052)	(0.123)	
Trust in people you know personally	7.467	7.600	0.132	1,100
	(1.511)	(1.583)	(0.094)	
Trust in people you meet for the first time	4.397	4.764	0.367	1,087
	(1.992)	(1.993)	(0.122)***	
Trust in people of another religion	5.076	5.710	0.635	1,066
	(2.120)	(1.945)	(0.125)***	
Trust in people who immigrated	4.996	5.686	0.691	1,069
	(2.139)	(2.041)	(0.128)***	
Trust in people of another nationality	5.331	5.920	0.590	1,069
	(2.080)	(1.914)	(0.122)***	
Trust in people who seek refuge	4.881	5.588	0.707	1,062
	(2.259)	(2.172)	(0.137)***	
Immigrants exert less effort	5.908	5.337	-0.571	973
	(2.273)	(2.408)	(0.152)***	
Immigrants are well integrated	4.695	5.082	0.386	1,079
	(2.274)	(2.052)	(0.132)***	
Our culture is enriched by immigrants	5.324	6.228	0.905	1,083
	(2.746)	(2.673)	(0.166)***	
Right wing	0.403	0.314	-0.089	1,108
	(0.491)	(0.465)	(0.029)***	
Observations	491	617		1,108

Notes: Column 1: sample of German participants sending different amounts across ethnic groups at least once either in the Ethnic block or in the Ethnic plus Income block. Column 2: sample of German participants sticking to the same transfer throughout the Ethnic block and also throughout the Ethnic plus Income block (the level of transfer can differ across blocks as long as second movers are treated the same way within a block). See appendix table 4.C1 for a description of the variables.

4.D Experimental instructions

Before turning to the detailed instructions, we first summarize the sequence of experimental games. They are organized in 5 tasks and several decisions have to be made within a task. The sequence of tasks is the same for everyone. One task and one decision within the selected task are randomly chosen to become payoff relevant. The 5 tasks are the following:

1. A (generalized) trust game as first mover and a (generalized) trust game as second mover
2. An unconditional public good game and a conditional public good game
3. A dictator game
4. The ethnic discrimination module (Ethnic block, Ethnic plus Income block and expectations)
5. A risk aversion module.

For each game where participants have to make decisions within a pair or a group, they are matched randomly with other participants of the survey, living in the same country. In each game, new participants are randomly drawn to be matched with the subject. Note that participants obtain no feedback regarding others' decisions or their earnings until after all games have been played and *Trustlab*'s survey module has been completed. Subjects received their money via Paypal up to 48 hours after the end of the survey. The currency was the dollar in the US and the euro in Germany. We report everything in dollars here for simplicity.

Screen 1

Welcome! Our research team[23] invites you to participate in a quick online study on decision-making. The aim of this study is to learn more about how we as human beings behave – how do we make decisions? How do we interact with one another when faced with different choices? How do we feel about the people and institutions around us? To find this out, you will be participating in different tasks: In the first part, you will participate in five simple tasks, in anonymous interaction with one or more other people. In the second part, you are going to sort different sets of words. In the third part, we ask you to answer a few questions about yourself and your opinions. The whole study should take you about 30 minutes. Note that you should complete this study in one sitting, without any extensive period of inactivity. For best results, minimize distractions and close other programs. You can participate in the study via your laptop computer or tablet (we support recent iPads). If you are having trouble accessing the platform, we advise you to switch to Google Chrome. If problems persist, please contact GMI, specifying your device model and browser.

By participating in the study's tasks, you can earn an amount of up to [currency amount: United States: "40 dollars"]. This amount will depend on the decisions you make together with the other participants during the study's tasks. At the end of the study, one of the several tasks you have completed will be randomly selected. The amount of money you will receive will correspond to your earnings in this selected task. Your decisions will also affect the earnings that other people will receive! You will receive your money at the end of the study via Paypal. Your payment will be processed after

[23] including researchers from Sciences Po Paris, Brown University and Kiel University

your decisions and those of other participants are collected. Because other participants may not be online at the same time as you, the calculation of your earnings may take up to 48 hours.

The data gathered in this study is subject to national privacy protocols. We will use it for research purposes only.

Screen 2 – SECTION 1: TASKS

We will start by giving you five tasks. Note that each task may include several different decisions. This is the part of the study that will allow you to earn additional money. Each of these decisions may determine your final payments.

At the beginning of each task, you may be grouped with other study participants. All participants in this study are from [country name (United States: "the United States"] like you. In each task, the other participants you are grouped with will be different: the same person will never be in your group more than once. How will your earnings be calculated? Your earnings in each task will depend on your and the other participants' decisions. At the end of the study, one of the five tasks you have completed will be randomly selected. The amount of money you will receive will correspond to your earnings in one of the decisions in this selected task. In short, each task may determine your final payoff!

Screen 3 – TASK 1: INTRODUCTION

In the first task, two people participate: Participant A and Participant B. As mentioned before, this other participant also lives in [country name (United States: "the United States"]. These are the rules of the task:

- At the beginning of the task, both participants receive [currency amount, United States: "10 dollars"].
- Participant A has the option to transfer none, part or all of his or her [currency amount, United States: "10 dollars"] to Participant B.
- Whatever amount Participant A sends is multiplied by 3.
- Participant B, after receiving the transfer of Participant A, has to decide how much money, if any, he or she wants to send back to Participant A.

You are asked to make decisions in both role A and B. Which role you will be assigned to for payment will be determined randomly. In either case, your interaction will be with a person who gets randomly assigned to the other role.

Screen 4 – TASK 1: TEST SIMULATION

This is not the real task yet, but a simulation to help you understand the rules better. You can use the test screen below to experiment with the different choices of the two participants. Between each test, click the "reset to zero" button below to reset the calculator.

Screen 5 – TASK 1: REAL TASK (PARTICIPANT A)

Now the real task 1 starts. Once you have made your decision and clicked the "Next" button, you cannot return to this screen.

Suppose you are selected to be in the role of Participant A. You have [currency amount, United States: "10 dollars"] in your possession. How much (if any) do you want to send to Participant B? Please enter a number from 0 to 10:

Screen 6 – TASK 1: REAL TASK (PARTICIPANT B)

Now, suppose you are selected to be in the role of Participant B. On this screen you will make the decisions that will count if you are selected for that role. Once you have made your decision and clicked the "Next" button, you cannot return to this screen. As always, your initial endowment is [currency amount, United States: "10 dollars"]. Remember that Participant A also starts with an endowment of [currency amount, United States: "10 dollars"]. If Participant A sends you any of the amounts listed in the table below, how much money (if any) do you want to send back to Participant A? All of your choices below can impact how much money you and the other participant will receive at the end of the study.

- If Participant A sends you [currency amount, United States: "0 dollars"] (your total endowment is now [currency amount, United States: "10 dollars"]. How much will you send back to Participant A:
- If Participant A sends you [currency amount, United States: "1 dollar"] (your total endowment is now [currency amount, United States: "13 dollars"]. How much will you send back to Participant A:
- If Participant A sends you [currency amount, United States: "2 dollars"] (your total endowment is now [currency amount, United States: "16 dollars"]. How much will you send back to Participant A:
- If Participant A sends you [currency amount, United States: "3 dollars"] (your total endowment is now [currency amount, United States: "19 dollars"]. How much will you send back to Participant A:
- If Participant A sends you [currency amount, United States: "4 dollars"] (your total endowment is now [currency amount, United States: "22 dollars"]. How much will you send back to Participant A:
- If Participant A sends you [currency amount, United States: "5 dollars"] (your total endowment is now [currency amount, United States: "25 dollars"]. How much will you send back to Participant A:
- If Participant A sends you [currency amount, United States: "6 dollars"] (your total endowment is now [currency amount, United States: "28 dollars"]. How much will you send back to Participant A:

- If Participant A sends you [currency amount, United States: "7 dollars"] (your total endowment is now [currency amount, United States: "31 dollars"]. How much will you send back to Participant A:
- If Participant A sends you [currency amount, United States: "8 dollars"] (your total endowment is now [currency amount, United States: "34 dollars"]. How much will you send back to Participant A:
- If Participant A sends you [currency amount, United States: "9 dollars"] (your total endowment is now [currency amount, United States: "37 dollars"]. How much will you send back to Participant A:
- If Participant A sends you [currency amount, United States: "10 dollars"] (your total endowment is now [currency amount, United States: "40 dollars"]. How much will you send back to Participant A:

Screen 7 – TASK 1: REAL TASK (HYPOTHETICAL SCENARIO)

You have just had made decisions as Participant A and Participant B. The following question is about your expectations of other people's decisions. You are not actually deciding as Participant A or Participant B, and this decision will not affect your earnings. We want you to imagine the following scenario: Imagine you sent [currency amount, United States: "5 dollars"], so Participant B receives [currency amount, United States: "15 dollars"], making his or her total budget [currency amount, United States: "25 dollars"]. Participant B has no information about your identity. What amount would you expect Participant B to return to you? Please enter a number from 0 to 25.

Screen 8 – THANK YOU

Thank you very much for entering your choice. We have recorded your decision. Now, please proceed to the second task.

Screen 9 – TASK TWO: INTRODUCTION

In the second task, groups of 4 participants (yourself and 3 other people) are formed. Remember, the participants in this group are different from the person you interacted with in the previous task. However, they all live in [country name (United States: "the United States"]. These are the rules:

- At the beginning, each group member has [currency amount, United States: "10 dollars"].
- Every group member has to choose how much of these [currency amount, United States: "10 dollars"] he or she wants to keep and how much he or she wants to transfer into a joint project.
- The total amount transferred to the joint project is multiplied by 1.6.
- At the end, the money in the joint project will be re-divided and split equally between all 4 group members (including yourself).

Screen 10 – TASK TWO: TEST SIMULATION

This is not the real task yet, but a simulation to help you understand the rules better. You can use the test screen below to experiment with the different choices of the four participants. Whenever you are ready to proceed to the real task, click next.

Screen 11 – TASK TWO: REAL TASK

Now the real task starts. Once you have made your decision and clicked the "Next" button, you cannot return to this screen. You have [currency amount, United States: "10 dollars"] in your possession. You may choose to keep this money, or choose to invest some (or all) of it in the joint project. How much (if any) do you want to transfer to the project?

Screen 12 – TASK TWO: REAL TASK CONTINUED

Now imagine that this time, you find out how much money the other three members of your group are investing in the joint project. All of your choices below can impact how much money you will receive at the end of the study.

Please indicate how much (if any) you would like to transfer to the joint project:

- if on average, each of the other group members contributes [currency amount, United States: "0 dollars"]:
- if on average, each of the other group members contributes [currency amount, United States: "1 dollar"]:
- if on average, each of the other group members contributes [currency amount, United States: "2 dollars"]:
- if on average, each of the other group members contributes [currency amount, United States: "3 dollars"]:
- if on average, each of the other group members contributes [currency amount, United States: "4 dollars"]:
- if on average, each of the other group members contributes [currency amount, United States: "5 dollars"]:
- if on average, each of the other group members contributes [currency amount, United States: "6 dollars"]:
- if on average, each of the other group members contributes [currency amount, United States: "7 dollars"]:
- if on average, each of the other group members contributes [currency amount, United States: "8 dollars"]:
- if on average, each of the other group members contributes [currency amount, United States: "9 dollars"]:
- if on average, each of the other group members contributes [currency amount, United States: "10 dollars"]:

Screen 13 – THANK YOU

Thank you very much for entering your choice. We have recorded your decision. Again, your payoff will depend on the actions of the other participants. Now, please proceed to the third task.

Screen 14 – TASK THREE: INTRODUCTION

The third task involves two participants – Participant A and Participant B. Remember, the other participant is different from the ones you interacted with in the previous two tasks. However, he or she also lives in [country name (United States: "the United States"]. These are the rules:

- At the beginning, Participant A receives [currency amount, United States: "10 dollars"].
- Participant B does not receive any money – he or she has [currency amount, United States: "0 dollars"].
- Participant A must now decide if he or she wants to transfer any of his or her [currency amount, United States: "10 dollars"] to Participant B.

This transfer is not multiplied by any number and Participant B cannot transfer any amount back to Participant A. Your role (Participant A or Participant B) will be determined later. We ask you to make a choice as A in case this is your role. B has no decision to make. Remember that someone will be assigned to role B and that person's payment will be affected by your decision as A. Because this task is simple, there will be no simulator to test out different choices.

Screen 15 – TASK THREE: REAL DECISION

This is the real third task. Once you have made your decision and clicked the "Next" button, you cannot return to this screen. Suppose that you are selected to be in the role of Participant A. You have [currency amount, United States: "10 dollars"] in your possession. How much (if any) do you want to transfer to Participant B?

Screen 16 - THANK YOU

Thank you very much for entering your choice. We have recorded your decision. Now, please proceed to the fourth task, which will be similar to Task 1.

Screen 17 - TASK FOUR: INTRODUCTION

Task 4 follows the same rules as Task 1 in which you participated earlier. In Task 4 you will be assigned to the role of Participant A. In this task, you will again be asked to make choices about how much money you want to transfer to a Participant B, but now you will have more information about the background of the Participant B that you are interacting with. Participant B will have no information about your own identity, except that you also live in [country]. If this Task is selected as the one determining your payments, you will be paid for one of the choices you make during this task, which will be randomly selected. In short, as always, every decision you make can determine your final payoff!

Screen 18 - TASK FOUR: RULES

Remember, in this task, two people participate: Participant A and Participant B. As mentioned before, the other participants you are interacting with all live in [country name (United States: "the United States"]. These are the rules of the task:

- At the beginning of the task, both participants receive [currency amount, United States: "10 dollars"].
- Participant A has the option to transfer none, part or all of his or her [currency amount, United States: "10 dollars"] to Participant B.
- Whatever amount Participant A sends is multiplied by 3.
- Participant B, after receiving the transfer of Participant A, has to decide how much, if any, money he or she want to send back to Participant A.

You are asked to make decisions as Participant A only. Your payoff will be based on the decision of one Participant B who you will be randomly matched with.

Screen 19 – TASK FOUR: REAL TASK (INTRODUCTION)

Now the real task 4 starts. Once you have made your decision and clicked the "Next" button, you cannot return to this screen. [The following three questions follow a randomized order]

Screen 20 - TASK FOUR: REAL TASK 1

Suppose that Participant B is [*US:* non-Hispanic White] [*Germany:* a person who was born in Germany and whose parents were also born in Germany]. You have [currency amount, United States: "10 dollars"] in your possession. How much (if any) do you want to transfer to this Participant B? Please enter an amount from 0 to 10.

Screen 21 – TASK FOUR: REAL TASK 2

Suppose that Participant B is [*US*: African American] [*Germany*: a person who was born in Eastern Europe or whose parents were born in Eastern Europe]. You have [currency amount, United States: "10 dollars"] in your possession. How much (if any) do you want to transfer to this Participant B? Please enter an amount from 0 to 10.

Screen 22 – TASK FOUR: REAL TASK 3

Suppose that Participant B is [*US:* Hispanic] [*Germany*: a person who was born in Turkey or whose parents were born in Turkey]. You have [currency amount, United States: "10 dollars"] in your possession. How much (if any) do you want to transfer to this Participant B? Please enter an amount from 0 to 10.

Screen 23 – TASK FOUR: REAL TASK CONTINUED

There will be three more interactions, with yet again more information about the Participant B that you are interacting with. The rules remain the same. Your payoff depends on how much you decide to transfer to Participant B and how much this participant will decide to transfer back to you. Again, each of these choices may determine your final payoff. [The following three questions follow a randomized order]

Screen 24 – TASK FOUR: REAL TASK 4

Suppose that Participant B is [*US:* non-Hispanic White and the income he or she receives places him or her among the 20% richest people in the US] [*Germany:* a person who was born in Germany and whose parents were also born in Germany. Moreover, the income he or she receives places him or her among the 20% richest people in Germany]. You have [currency amount, United States: "10 dollars"] in your possession. How much (if any) do you want to transfer to this Participant B? Please enter an amount from 0 to 10.

Screen 25 – TASK FOUR: REAL TASK 5

Suppose that Participant B is [*US:* African American and the income he or she receives places him or her among the 20% richest people in the US] [*Germany:* a person who was born in Eastern Europe or whose parents were born in Eastern Europe. Moreover, the income he or she receives places him or her among the 20% richest people in Germany]. You have [currency amount, United States: "10 dollars"] in your possession. How much (if any) do you want to transfer to this Participant B? Please enter an amount from 0 to 10.

Screen 26 – TASK FOUR: REAL TASK 6

Suppose that Participant B is [*US:* Hispanic and the income he or she receives places him or her among the 20% richest people in the US] [*Germany:* a person who was born in Turkey or whose parents were born in Turkey. Moreover, the income he or she receives places him or her among the 20% richest people in Germany]. You have [currency amount, United States: "10 dollars"] in your possession. How much (if any) do you want to transfer to this Participant B? Please enter an amount from 0 to 10.

Screen 27 – TASK FOUR: EXPECTATIONS

As before, we will ask you about your expectations of other people's decisions. Remember, all participants live in [country] like you. You are not making a decision as Participant A or Participant B, and what you enter will not affect your earnings. In the following screens, we want you to imagine different scenarios: [The following three questions follow a randomized order]

Screen 28 – TASK FOUR: EXPECTATIONS 1

Suppose that Participant B is [*US:* non-Hispanic White] [*Germany:* a person who was born in Germany and whose parents were also born in Germany]. You have no information regarding B's income, in this case. Imagine you sent [currency amount, United States: "5 dollars"], so Participant B receives

[currency amount, United States: "15 dollars"], making his or her total budget [currency amount, United States: "25 dollars"]. Participant B has no information about your identity. What amount would you expect Participant B to return to you? Please enter a number from 0 to 25.

Screen 29 – TASK FOUR: EXPECTATIONS 2

Suppose that Participant B is [*US*: African American] [*Germany*: a person who was born in Eastern Europe or whose parents were born in Eastern Europe]. You have no information regarding B's income, in this case. Imagine you sent [currency amount, United States: "5 dollars"], so Participant B receives [currency amount, United States: "15 dollars"], making his or her total budget [currency amount, United States: "25 dollars"]. Participant B has no information about your identity. What amount would you expect Participant B to return to you? Please enter a number from 0 to 25.

Screen 30 – TASK FOUR: EXPECTATIONS 3

Suppose that Participant B is Suppose that Participant B is [*US:* Hispanic] [*Germany*: a person who was born in Turkey or whose parents were born in Turkey]. You have no information regarding B's income, in this case. Imagine you sent [currency amount, United States: "5 dollars"], so Participant B receives [currency amount, United States: "15 dollars"], making his or her total budget [currency amount, United States: "25 dollars"]. Participant B has no information about your identity. What amount would you expect Participant B to return to you? Please enter a number from 0 to 25.

Screen 31 - THANK YOU

Thank you very much for entering your choices. We have recorded your decision. Again, your payoff will depend on the actions of the other participants. Now, you will proceed to the last task of this section of the study.

Screen 32 – TASK FIVE: INTRODUCTION

In this task you have the option to choose from six different gambles. In each gamble, you can win one out of two amounts. You must select one and only one of these gambles. Each gamble has two possible outcomes: outcome A and outcome B. Only one of these outcomes will occur. The gamble works as a random draw, comparable to a coin toss. Like in a coin toss, each possible outcome has a 50% chance of occurring.

Your compensation for this part of the study will be determined by:

1. Which of the six gambles you select. This is your choice.
2. Which of the two possible outcomes occur. This is determined by chance. The random draw is conducted by our computer. Either outcome has the same probability of occurring. The gamble selection table below shows your possible options. You will be asked to choose one of these gambles.

Examples:

- For instance, if you choose Gamble 2, you will earn 7 dollars if outcome A occurs, or 10 dollars if outcome B occurs.

- If you choose Gamble 5, you will earn 4 dollars if outcome A occurs, or 16 dollars if outcome B occurs.
- If you choose Gamble 1, you will earn 8 dollars, regardless of which outcome occurs.

Gamble	Outcome	Payoff	Probabilities	Choice
1	A	8	50%	
	B	8	50%	
2	A	7	50%	
	B	10	50%	
3	A	6	50%	
	B	12	50%	
4	A	5	50%	
	B	14	50%	
5	A	4	50%	
	B	16	50%	
6	A	1	50%	
	B	19	50%	

Screen 33 – TASK FIVE: REAL TASK

Now the real task 5 starts. Once you have made your decision and clicked the "Next" button, you cannot return to this screen. These are the six gambles from which you can choose. If this task is chosen for payment, then your earnings will depend on the gamble you choose and the outcome of the gamble. Please select the gamble of your choice.

Gamble	Outcome	Payoff	Probabilities	Choice
1	A	8	50%	
	B	8	50%	
2	A	7	50%	
	B	10	50%	
3	A	6	50%	
	B	12	50%	
4	A	5	50%	
	B	14	50%	
5	A	4	50%	
	B	16	50%	
6	A	1	50%	
	B	19	50%	

Screen 34 – TASK FIVE: THANK YOU

Thank you once again! You have completed all tasks in this section and we have recorded all your choices. Let's go to the second section of the study, where you will be asked to sort a number of words.

Résumé en français

Etudes des déterminants des inégalités salariales

Les inégalités au cœur de la profession économique

Mon principal argument tout au long de cette thèse est que les inégalités de revenus sont avant tout un objet micro-fondé. Les individus font des choix, ont des préférences, une perception subjective de l'équité, se conforment aux normes sociales, les rejettent ou les renouvellent. Toutes ces forces microéconomiques façonnent la vie et le statut économique et social des individus par leurs propres choix – contraints ou non – et aussi par les décisions prises par d'autres individus. L'agrégation et les interactions de ces forces microéconomiques ont des implications macroéconomiques. Collectivement, les individus façonnent les idéologies, changent les institutions économiques et les régimes politiques. Ceci est particulièrement pertinent pour le thème des inégalités de revenus. Avant de commencer à mesurer des indices agrégés tels que le coefficient de Gini[24], ou la part du revenu qui revient aux 1% les plus riches, il est utile de réfléchir aux processus microéconomiques sous-jacents. Les gens choisissent ou non d'aller à l'université, expriment leurs préférences en votant, font des choix sur le marché du travail, etc.

Le choix de cette approche a été, comme toute entreprise de recherche, profondément influencé par le contexte économique, social, politique et universitaire dans lequel j'ai commencé mon travail de doctorat. Ces dernières années ont vu la montée des populismes incarnés par les partis radicaux [25], un fort mécontentement civil tel que le mouvement des Gilets jaunes en France, et la baisse de la confiance dans les institutions politiques dans plusieurs pays (OECD, 2017; Citrin and Stoker, 2018). Ce contexte troublé a suscité l'intérêt de chercheurs actifs dans un domaine universitaire particulièrement riche, reconfirmant sa pertinence. Ma thèse est largement issue du débat intellectuel qui a suivi les travaux pionniers d'Anthony B. Atkinson, Branko Milanovic, Thomas Piketty, Emmanuel Saez, Gabriel Zucman et d'autres. Le *Capital au XXIème siècle* de Piketty a été publié pour la première fois en Français trois ans avant le début de ma thèse et en anglais deux ans auparavant. De nouvelles données sur la richesse et l'inégalité des revenus ont été publiées et présentées au monde entier – posant ainsi des chiffres sur des intuitions – ce qui a déclenché des débats nationaux sur l'élaboration de politiques de redistribution (Atkinson, 2014; Ostry et al., 2019; Saez and Zucman, 2020). Cette recherche révolutionnaire a eu juste le temps d'imprégner la profession, de telle sorte que, pendant la dernière année de mon master et la première année de mon doctorat, les séminaires de recherche étaient remplis d'économistes qui s'étaient appropriés le sujet à leur manière, l'enrichissant de leurs méthodes et données, produisant de nouvelles perspectives et questions de recherche. Certains auteurs qui s'étaient déjà attaqués à la question de la répartition des revenus plusieurs décennies auparavant – comme Kuznets (1955) – ont suscité un regain d'attention.

[24] Le coefficient de Gini est une mesure synthétique de la répartition des revenus dans la population d'un pays. Il va de 0 (égalité totale, tout le monde a le même revenu) à 1 (inégalité extrême, une personne gagne tout).

[25] Beaucoup de chercheurs ont tenté de déterminer les origines de cette dynamique, en soulignant l'importance des facteurs économiques (Algan et al., 2017; Guriev, 2018). La montée des inégalités de revenus (Piketty, 2013), les effets polarisants de la mondialisation sur les salaires créant des gagnants (travailleurs hautement qualifiés) et des perdants (travailleurs peu qualifiés) (Autor et al., 2014, 2016) et la crise financière de 2008 avec ses conséquences dramatiques pour les plus pauvres, sont des déclencheurs de ressentiment, de mécontentement et de méfiance envers les élites et les institutions politiques (Algan et al., 2017, 2019). Cette vision dichotomique du monde opposant le peuple à une élite est souvent considérée comme une définition minimale consensuelle du populisme (Bonikowski and Gidron, 2016; Mudde, 2017). Une autre cause hypothétique du populisme est un changement culturel avec la renaissance d'un animus racial qui pourrait avoir été en sommeil pendant plusieurs décennies, masqué par la vitalité apparente des idées progressistes (Stephens-Davidowitz, 2014; Inglehart and Norris, 2016). Ce contrecoup culturel, pour reprendre la terminologie de Norris and Inglehart (2019), a pu être amplifié par l'accroissement de la diversité ethnique et la crise des migrants, entraînant une redéfinition de l'identité "nous contre eux".

Questions de recherche

Cette thèse analyse le déroulement de plusieurs événements qui peuvent conduire des individus à recevoir des salaires différents. La première question que je me pose est la suivante : *Pourquoi* des personnes différentes sont-elles rémunérées de manière inégale ? J'examine ensuite les points de vue des individus en ce qui concerne les inégalités de revenus : est-ce que des personnes différentes *devraient* être rémunérées inégalement ? La distinction entre "différent" et "inégal" est importante. Les salaires sont inégaux dans le sens où ils varient selon une métrique monétaire unique, ce qui implique que nous pouvons ordonner les salaires selon une hiérarchie verticale. Mais si les salaires sont unidimensionnels, les individus varient de façon multidimensionnelle, et ne peuvent être ordonnés de manière verticale. Les individus peuvent varier en fonction de leurs talents, capacités, motivation pour travailler dur, leur mérite, leur personnalité, etc. En tant qu'économiste, mon objectif est d'analyser la manière dont les marchés, les institutions et les individus convertissent ces facteurs multidimensionnels en un résultat unidimensionnel : le salaire. Répondre à la première question est une entreprise ambitieuse, et je ne prétends pas y apporter une réponse exhaustive. J'ai axé mes recherches sur trois déterminants des inégalités salariales : les choix éducatifs (Chapitre 1), les préférences normatives des managers et leurs préoccupations en matière d'incitation (Chapitre 3) et la discrimination ethnique (Chapitre 4).

La seconde question est également vaste, et je tiens à souligner que même si elle ressemble à une question normative, elle n'en est pas une. Dans les Chapitres 2 et 3, je montre comment des *individus* répondent à cette question, avec leurs propres points de vue et perceptions de la justice et de l'équité – concepts que je décrirai plus longuement dans la Section . En répondant à cette question, mon objectif n'est pas seulement de clarifier les préférences sociales et politiques de nos sociétés, je veux aussi décrire comment ces préférences varient d'un individu à l'autre. Tenir compte de l'hétérogénéité nous permet de prendre du recul par rapport aux récits et idéologies dominants, et de mieux comprendre le comportement des individus les moins favorisés (Chapitre 2).

Un récit des inégalités

Les chapitres de ma thèse n'ont pas pour but de répondre à ces deux questions de manière séquentielle, ils suivent plutôt un récit chronologique. Comme ma thèse porte sur les choix et les préférences des individus, elle pourrait être entièrement résumée sous la forme d'un récit. Le Chapitre 1 commence par une étudiante qui hésite entre différents domaines d'études. Ses choix influenceront fortement sa carrière professionnelle et son parcours salarial. Au lycée, elle s'intéresse beaucoup à l'économie et lit régulièrement la presse sur ce sujet. Une fois à l'université, elle se rend compte que, même si ses notes en économie sont très respectables, elle est meilleure en sciences politiques et décide de suivre plutôt cette voie. Son ami, au contraire, s'en tient à sa vocation initiale, malgré des notes assez faibles dans son domaine d'étude préféré. Elle se demande pourquoi il ne tient pas compte de ses notes, contrairement à elle.

Dans le Chapitre 2, elle poursuit ses études et peut voter pour la première fois à l'élection présidentielle. Elle doit réfléchir au candidat qu'elle aimerait voir en fonction. En lisant les différents programmes électoraux, elle identifie les différentes idéologies sous-jacentes. L'une d'elle considère que les gens devraient être payés en fonction de leurs mérites et de leurs talents, ce qui rendrait équitables les inégalités économiques qui en résultent. Une autre considère que même dans un contexte d'inégalité fondée sur le mérite, les différences de revenus peuvent encore être indécentes, ce qui

justifie l'intervention de l'État pour redistribuer les revenus. Pour se faire une opinion, elle s'inspire de la *Theory of Justice* du philosophe John Rawls, qu'elle a lu pour ses cours de sciences politiques. Elle s'interroge sur le type de société dans laquelle elle aimerait grandir si elle n'avait aucune idée de ses futures circonstances de vie, de ses talents et de son environnement familial. L'exercice est concluant, elle sait pour qui elle va voter !

Au Chapitre 3, elle obtient son premier emploi et rencontre son premier patron. Tous les membres de son équipe reçoivent presque le même salaire, tandis que dans l'entreprise de son ami, les salaires varient considérablement et sont indexés sur les performances. En discutant avec lui, elle se rend compte que leurs patrons respectifs ont des points de vue très différents sur ce qui constitue une répartition équitable des salaires. Son propre patron considère que les personnes qui font le même travail devraient être payées de la même façon, tandis que le patron de son ami pense que les employés ont besoin d'incitations financières. Elle se rend compte que ces différents points de vue sur l'équité ont des implications majeures pour l'ensemble des structures salariales et des styles de gestion des entreprises.

Enfin, au Chapitre 4, elle a de plus en plus de mal à travailler avec certains collègues, qui semblent se méfier d'elle pour une raison qu'elle essaye d'identifier. Elle a été la cible de blagues racistes, et craint que la couleur de sa peau n'explique le manque de volonté de ses collègues blancs à coopérer avec elle, ce qui affecte ses performances au travail. Elle en parle à son patron et ils élaborent ensemble une stratégie : il reconnaîtra publiquement son travail et son mérite afin de mettre un terme à la discrimination à laquelle elle est confrontée. Cela fonctionne – ses collègues blancs la respectent davantage – mais la stratégie se retourne contre elle lorsque des collègues de la même origine qu'elle deviennent antipathiques et manifestement jaloux de son succès.

Quelles conclusions pouvons-nous tirer de ce récit ? Les inégalités de revenus découlent de l'agrégation de millions de choix économiques, sociaux et politiques individuels, faits à des moments différents. En outre, ces choix sont façonnés par les préférences, les perceptions de l'équité et les croyances des gens. Enfin, les individus ne font pas seulement des choix qui influencent leur propre trajectoire salariale, ils prennent également des mesures qui peuvent être bénéfiques ou préjudiciables aux personnes avec lesquelles ils interagissent.

Résumé des chapitres

Chapitre 1 : *Updating of academic tastes and ability signals*

Dans ce chapitre co-écrit avec Ghazala Azmat, Anne Boring et Roberto Galbiati, nous nous concentrons sur la première question décrite dans la section 4.5, dans le contexte des choix dans l'enseignement supérieur. Nous documentons le processus d'apprentissage qui conduit les étudiants à choisir un master. Nous partons du principe que les étudiants connaissent plus ou moins bien leurs vraies préférences académiques. Certains étudiants peuvent avoir des convictions très fortes et être déjà passionnés par un sujet, tandis que d'autres sont plus incertains. Leur première année à l'université est l'occasion de découvrir de nouvelles matières, d'obtenir des notes qui fonctionnent comme des signaux sur leurs capacités, et peuvent influencer leur orientation dans l'enseignement supérieur. Les goûts académiques des étudiants évoluent donc au fil du temps, sont actualisés, en fonction des signaux que l'étudiant reçoit. Ce processus de "mise à jour" est essentiel car il conduit à des choix qui ont des conséquences importantes sur les opportunités qu'auront les étudiants une fois sur le marché du travail. Par exem-

ple, les salaires de départ à la fin des études varient considérablement selon les masters et les filières (Black et al., 2003; Carnevale et al., 2013).

De plus, savoir comment et pourquoi les étudiants changent leurs goûts académiques peut alimenter le débat sur la flexibilité des systèmes d'enseignement supérieur. Il existe de grandes différences institutionnelles dans la manière dont les étudiants choisissent leurs filières d'études. Dans des pays comme les États-Unis, les étudiants peuvent prendre des cours dans des matières très différentes avant de se spécialiser, tandis que dans d'autres pays, comme le Royaume-Uni et la France, les filières sont choisies avant même d'entrer à l'université. Un système peu flexible présente l'inconvénient d'entraver les changements d'orientation, mais a l'avantage d'accélérer l'accumulation de compétences (Malamud, 2011).

Dans ce chapitre, nous documentons ce processus d'actualisation en nous demandant comment et pourquoi les convictions des étudiants concernant leurs préférences académiques évoluent tout au long de leurs études de premier cycle. Pour mener notre analyse, nous utilisons un ensemble de données unique en son genre provenant d'une université française spécialisée dans les sciences sociales – Sciences Po. Ces données présentent plusieurs avantages. Tout d'abord, elles sont longitudinales, car nous pouvons suivre les étudiants tout au long de leurs études, de leur dernière année de lycée jusqu'à leurs études supérieures. Nous mesurons les goûts académiques à deux moments précis en utilisant le contenu des deux lettres de motivation que tous les étudiants doivent rédiger. La première est une pièce à inclure dans leur dossier de candidature pour Sciences Po, alors qu'ils sont encore au lycée, et la seconde est rédigée deux ans plus tard lorsqu'ils justifient leurs choix d'universités pour le programme obligatoire d'études à l'étranger de troisième année. De plus, ce système universitaire est flexible dans le sens où les étudiants se spécialisent tard dans leurs études, mais ils n'ont aucune latitude dans leurs choix de cours en première année. Le programme d'études est identique pour tous les étudiants de première année ; ils doivent se familiariser toutes les grandes disciplines des sciences sociales (économie, histoire, droit, sciences politiques et sociologie) avant de faire des choix académiques importants. Cette caractéristique est souvent absente dans d'autres universités où les étudiants choisissent eux-mêmes leurs cours et ne vont pas pouvoir savoir s'ils aiment certaines matières ou pas. L'auto-sélection n'est donc pas une préoccupation dans notre contexte. Enfin, il n'y a pas de contrainte du côté de l'offre pour les choix de master, ce qui atténue les préoccupations stratégiques liées à la compétition qui se poseraient si les étudiants devaient anticiper les choix des autres étudiants, dans l'optique de se soustraire à la concurrence ou, au contraire, de s'y confronter.

Nous montrons que les goûts académiques au lycée et en deuxième année sont positivement corrélés avec les choix révélés (les choix de master), cette relation étant plus forte avec les goûts plus récents qu'avec les plus anciens. Cela indique que les étudiants changent d'avis en ce qui concerne leurs préférences après avoir été exposés au programme de l'université. Nous étudions ensuite les raisons qui motivent ce processus d'actualisation des préférences académiques. Nous nous concentrons sur trois types d'information mis à la disposition des étudiants. Premièrement, nous utilisons des données sur les goûts initiaux des étudiants. Ils n'entament pas leurs études supérieures avec le même degré de certitude concernant leurs préférences académiques. Nous évaluons la force de ces croyances initiales quand l'étudiant est en dernière année de lycée. Ensuite, nous étudions l'effet de plusieurs signaux que les étudiants reçoivent concernant leurs capacités : (i) ils découvrent leur aptitude dans le domaine dans lequel ils voulaient initialement se spécialiser ; (ii) ils découvrent leurs capacités dans d'autres domaines ; (iii) ils peuvent comparer leurs performances à celles de leurs camarades de classe. Nous

étudions également si la précision des informations sur les performances des pairs importe dans le processus d'actualisation des goûts académiques.

Nous constatons qu'environ 75% des étudiants changent de matière préférée au cours des deux premières années de leurs études. Les signaux qu'ils reçoivent par rapport à leurs capacités jouent un rôle important dans ce processus : ils incitent à persévérer lorsqu'ils indiquent une bonne adéquation avec les goûts initiaux, et favorisent le changement lorsqu'ils indiquent que l'étudiant est meilleur dans une autre matière. Néanmoins, pour les étudiants qui ont la conviction qu'ils finiront par choisir un certain domaine d'études, nous constatons une plus grande inertie au changement, même lorsqu'ils reçoivent des informations négatives sur leurs aptitudes scolaires dans leur matière favorite. Nous montrons également que les performances relatives, par comparaison avec les camarades de classe, ne jouent pas un rôle significatif. Un degré de précision plus élevé concernant les performances des pairs ne change pas ce résultat. Nous ne constatons pas non plus de différences dans le processus d'actualisation des goûts en fonction du sexe ou du milieu socio-économique des étudiants.

Chapitre 2 : *Preferences over income distribution, evidence from a choice experiment*

Dans ce chapitre co-écrit avec Max Lobeck, Claudia Senik et Thierry Verdier, nous nous concentrons sur la deuxième grande question de recherche de ma thèse en analysant la façon dont les individus répondent à la question suivante : *devrait-on payer inégalement des personnes différentes ?* Nous évaluons 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.

Chapitre 3 : *Principals distributive preferences and the incentivization of agents*

Dans ce chapitre co-écrit avec Max Lobeck, nous abordons la question des croyances concernant l'arbitrage entre l'égalité et l'efficacité, les idéaux en matière d'équité, et comment tout cela se traduit *in fine* par des choix de compensation salariale. Nous nous concentrons sur les choix et préférences des employeurs et managers, qui sont les entités pertinentes en termes de distribution des salaires.

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.

Chapitre 4: *Ethnic bias, economic success, and trust*

Le dernier chapitre est co-écrit avec l'équipe de chercheurs du *Trustlab* : Yann Algan, Gianluca Grimalda, Fabrice Murtin, Louis Putterman, Ulrich Schmidt et Vincent Siegerink.

Nous nous intéressons à la question des origines des inégalités en nous concentrant sur le rôle joué par la discrimination ethnique. Dans un cadre expérimental, nous nous concentrons sur des interactions économiques minimales entre des individus appartenant à des groupes ethniques différents. Nos objectifs sont de mieux comprendre l'ampleur et les caractéristiques de la discrimination fondée sur l'origine ethnique et de tester si elle peut être réduite. Nous étudions deux pays occidentaux – les États-Unis et l'Allemagne. Alors que la recherche sur ce sujet se concentre en général sur la discrimination transnationale, ou sur la discrimination à l'intérieur d'un pays entre deux groupes ethniques, nous utilisons des échantillons larges permettant d'étudier les schémas discriminatoires entre la majorité ethnique, deux minorités et un groupe résiduel, dans chacun des pays. Nous pouvons ainsi étudier si la discrimination est sélective ou si le groupe ethnique de l'outsider importe. Par exemple, nous pouvons non seulement déterminer si une personne blanche aux Etats-Unis tend à favoriser les Blancs, mais aussi savoir si elle traite de façon similaire une personne noire et une personne hispanique. Nous déterminons si le favoritisme ethnique découle d'une défiance basée sur l'anticipation d'une fiabilité moindre des personnes hors de son propre groupe (discrimination dite statistique), ou si elle est une pure question de préférence ethnique (discrimination dite de goût).

Nous testons également l'efficacité d'une intervention destinée à réduire la discrimination. Un biais ethnique repose souvent sur des stéréotypes selon lesquels les minorités ne partagent pas la même éthique de travail que la majorité ethnique. Les personnes issues de minorités ethniques sont souvent dépeintes, dans l'imaginaire de nombreuses personnes, comme paresseuses et profitant des prestations sociales (Gilens, 2009; Alesina et al., 2018). Nous vérifions si la publication d'informations indiquant que les personnes issues de minorités ethniques réussissent sur le plan économique atténue les attitudes discriminatoires de la majorité ethnique à leur égard.

Notre expérience est issue d'un module sur la discrimination ethnique sur la plateforme *Trustlab*, une enquête en ligne à grande échelle et multi-pays conçue pour étudier les préférences sociales, la confiance dans les institutions et la confiance interpersonnelle à l'aide de jeux expérimentaux rémunérés (Murtin et al., 2018; Aassve et al., 2018a,b). Le module a été utilisé aux États-Unis et en Allemagne sur environ 1000 sujets, représentatifs de la population nationale de chaque pays. Le module consiste en plusieurs jeux de confiance impliquant des paires de joueurs. Chaque joueur de la paire reçoit une dotation de 10 dollars/euros. Le premier joueur peut transférer n'importe quelle fraction de cette dotation à un second joueur. Le montant transféré est multiplié par 3 et la deuxième personne peut ensuite reverser une partie de la somme transférée. Notre principale manipulation expérimentale consiste à révéler le groupe ethnique du deuxième joueur au premier. Aux États-Unis, chaque premier joueur est mis en contact dans un ordre aléatoire avec un Blanc, un Afro-Américain et un Hispanique, qui jouent le rôle de second joueur. En Allemagne, le premier joueur est mis en contact, toujours de façon aléatoire, avec des Allemands d'origine, une personne d'origine turque et une personne originaire d'Europe de l'Est. Nous mesurons la prévalence du favoritisme pour son propre groupe ethnique en comparant les transferts choisis par les premiers joueurs en fonction de l'ethnicité du second joueur. Dans notre expérience, une attitude discriminatoire consiste à favoriser les personnes appartenant à son propre groupe ethnique, au détriment des autres. En d'autres termes, il s'agit de la propension à transférer des sommes plus importantes aux personnes de son groupe qu'à celles des autres groupes. Comme nous connaissons également l'ethnicité du premier joueur, nous pouvons étudier comment le favoritisme varie selon les types ethniques. Cela permet d'étudier les différences de comportement entre la majorité ethnique et les minorités. Nous étudions également la sélectivité des préjugés, c'est-à-dire si les transferts des premiers joueurs dépendent du type d'outsider. Les sujets ont-ils plus de préjugés à l'égard d'un groupe ethnique qu'à l'égard d'un autre ? Existe-t-il un groupe ethnique qui est discriminé ou favorisé par tous les groupes, ou les biais sont-ils totalement spécifiques à chaque ethnicité ?

La deuxième partie de l'expérience vise à déterminer si donner des informations sur le revenu des minorités peut atténuer les préjugés ethniques, surtout ceux venant de la majorité. Nous effectuons une autre série de jeux de la confiance où les premiers joueurs sont maintenant associés à des seconds joueurs dont les revenus appartiennent au top 20% de la distribution des revenus du pays. Nous continuons à faire varier l'origine ethnique du second joueur. Les sujets issus de la majorité ethnique sont donc confrontés à des personnes riches, même si issues d'une minorité, ce qui contredit les habituels récits populistes qui dépeignent les immigrants ou les minorités comme des assistés et des personnes oisives. Nous analysons également comment les minorités réagissent au fait de jouer avec des riches issus de leur propre minorité ethnique ou d'autres minorités.

Dans l'ensemble, nous constatons que les membres de tous les groupes ethniques favorisent de façon significative les personnes du même groupe ethnique que le leur, à l'exception des participants originaires d'Europe de l'Est. Ce préjugé est particulièrement important pour les Afro-Américains aux États-Unis, les Allemands d'origine et les participants d'origine turque en Allemagne. Nous montrons en outre que la discrimination ethnique est sélective en Allemagne. Les Allemands d'origine discriminent deux fois plus les participants d'origine turque que ceux originaire d'Europe de l'Est. Au contraire, les premiers joueurs d'Europe de l'Est et ceux originaires de Turquie se discriminent mutuellement, mais sont plus neutres dans leur confiance quand ils sont associés aux Allemands d'origine. Aux États-Unis, les groupes ethniques ont un préjugé non sélectif et donc plus homogène.

Nous pouvons décomposer la discrimination ethnique en une composante basée sur la préférence pure et une composante statistique en utilisant l'information sur ce que les premiers joueurs pensent que les seconds joueurs vont leur transférer en retour de leur propre transfert. Nous appelons cette anticipation la fiabilité attendue. Nous en déduisons que 80% des préjugés sont dus à une discrimination purement fondée sur un goût (ou plutôt dégoût) ethnique et 20% à la discrimination statistique. Ces proportions sont à peu près équivalentes dans les deux pays. Nous montrons en outre que les sujets se font souvent des idées inexactes sur la fiabilité de certains groupes ethniques, à l'exception de ceux concernant les seconds joueurs d'origine turque, qui transfèrent en effet beaucoup moins d'argent aux premiers joueurs que les autres groupes.

Bien que tous les groupes ethniques tendent à réduire leurs transferts lorsque le second joueur est riche, l'information sur le revenu atténue la discrimination ethnique. Le favoritisme ethnique disparaît presque complètement, sauf pour les Afro-Américains et les Allemands d'origine qui continuent à favoriser leur propre groupe, même si le second joueur est riche, mais dans une mesure bien moindre que lorsque les informations sur les revenus ne sont pas divulguées. De plus, nous montrons l'existence d'un premium pour les minorités riches en Allemagne. En effet, les Allemands d'origine discriminent moins les seconds joueurs d'origine turque qui sont riches que ceux qui ne le sont pas, ce qui laisse entendre que les récits de réussite au sujet des minorités ethniques pourraient contribuer à changer les stéréotypes. Cependant, nous montrons également que ce traitement peut se retourner contre eux et générer de la méfiance au sein des groupes minoritaires. Aux États-Unis, nous observons également un premium pour les minorités riches mais d'une plus faible ampleur qu'en Allemagne. Cela pourrait s'expliquer par le fait que les Américains Blancs avaient au départ un comportement finalement peu discriminatoire, donc l'intervention aurait eu de toute façon un effet limité. Aux États-Unis, le traitement ne génère aucun effet délétère au sein des minorités.

Nous constatons également que les premiers joueurs appartenant au top 20% de la distribution des revenus favorisent les plus riches car ils transfèrent davantage à leurs semblables (eux aussi dans le top 20% de la distribution), par rapport au comportement des premiers joueurs moins riches. Cependant, ce favoritisme basé sur le revenu n'est significatif que pour les Allemands de souche.

Bibliographie

Aassve, Arnstein, Pierluigi Conzo, and Francesco Mattioli. Was Banfield right ? New insights from a nationwide laboratory experiment. 2018a.

Aassve, Arnstein, Gianmarco Daniele, and Marco Le Moglie. Never Forget the First Time : The Persistent Effects of Corruption and the Rise of Populism in Italy. 2018b.

Alesina, Alberto, Armando Miano, and Stefanie Stantcheva. Immigration and Redistribution. *NBER Working Paper 24733*, 2018.

Algan, Yann, Sergei Guriev, Elias Papaioannoun, and Evgenia Passari. The European Trust Crisis and the Rise of Populism. *Brookings Papers on Economic Activity*, (Fall):309–382, 2017.

Algan, Yann, Elizabeth Beasley, Daniel Cohen, and Martial Foucault. *Les Origines du populisme*. Le Seuil, 2019.

Atkinson, A. B. After Piketty? *British Journal of Sociology*, 65(4):619–638, 2014.

Autor, David, David Dorn, Gordon Hanson, and Kaveh Majlesi. Importing Political Polarization? The Electoral Consequences of Rising Trade Exposure. *NBER Working Paper*, 2016.

Autor, David H, David Dorn, Gordon H Hanson, and Jae Song. Trade Adjustment: Worker-Level Evidence. pages 1799–1860, 2014.

Black, Dan A., Seth Sanders, and Taylor Lowell. The economic reward for studying economics. *Economic Inquiry*, 41(3):365–377, 2003.

Bonikowski, Bart and Noam Gidron. Multiple Traditions in Populism Research : Toward a Theoretical Synthesis. *APSA Comparative Politics Newsletter*, 26(12):7–14, 2016.

Carnevale, Anthony P, Jeff Strohl, and Michell Melton. What's it worth? The economic value of college majors. Technical report, 2013.

Citrin, Jack and Laura Stoker. Political Trust in a Cynical Age. *Annual Review of Political Science*, 21(1):49–70, 2018.

Gilens, Martin. *Why Americans hate welfare: Race, media, and the politics of antipoverty policy*. University of Chicago Press, 2009.

Guriev, Sergei. Economic Drivers of Populism. *AEA Papers and Proceedings*, pages 200–203, 2018.

Inglehart, Ronald F and Pippa Norris. Trump, Brexit, and the Rise of Populism: Economic Have-Nots and Cultural Backlash. 2016.

Kuznets, Simon. Economic Growth and Income Inequality. *The American Economic Review*, 45(1): 1–28, 1955.

Malamud, Ofer. Discovering one's talent: Learning from academic specialization. *Industrial and Labor Relations Review*, 64(2):375–405, 2011.

Mudde, Cas. Populism: An Ideational Approach. In *The Oxford Handbook of Populism*, number October. 2017.

Murtin, Fabrice, Lara Fleischer, Vincent Siegerink, Arnstein Aassve, Yann Algan, Santiago Gonzalez, Gianluca Grimalda, Rafael Hortala, Vallve Soonhee, Kim David, Lee Zsuzsanna, Lonti Louis, Putterman Ulrich, and Schmidt Conal. Trust and its Determinants : Evidence from the Trustlab Experiment. 2018.

Norris, Pippa and Ronald Inglehart. *Cultural backlash: Trump, Brexit, and authoritarian populism*. Cambridge University Press, 2019.

OECD, . *OECD Guidelines on Measuring Trust*. 2017.

Ostry, Jonathan D, Prakash Loungani, and Andrew Berg. *Confronting inequality: how societies can choose inclusive growth*. Columbia University Press, 2019.

Piketty, Thomas. *Le capital au XXIe siècle*. Le Seuil, 2013.

Saez, Emmanuel and Gabriel Zucman. *Le Triomphe de l'injustice. Richesse, évasion fiscale et démocratie*. Le Seuil, 2020.

Stephens-Davidowitz, Seth. The cost of racial animus on a black candidate : Evidence using Google search data. *Journal of Public Economics*, 118:26–40, 2014.

List of Figures

1 Distribution of individual give rates 8

1.1 Evolution of academic tastes between high school and second year 39
1.2 Ability signal and strength of dominant high-school taste 48
1.A1 Proportion of match between keyword-based and hand-labeled dominant taste (**accuracy rate**) 66
1.A2 Proportion of correctly identified (hand-labeled) positives using keywords (**sensitivity rate**) 67
1.A3 Proportion of correctly identified (hand-labeled) negatives using keywords (**specificity rate**) 68
1.B1 Distribution of local peer group sizes 70
1.B2 Randomization tests for tutorial groups 72
1.B3 Randomization tests for tutorial groups – Paris and Nancy campuses 72
1.B4 Composition of dominant high-school tastes in peer groups 73
1.C1 Typology of students 76

2.1 Choices Behind the Veil of Ignorance by Treatment 92
2.2 Decisions with Known Position - Constant Efficiency Choices 94
2.3 Decisions with Known Position - Project A is Pareto-Dominant 95
2.A1 Distribution of the Choices Made for Lottery 1 110
2.A2 Differences between Choice 1 and the payoff-equivalent lottery 110
2.A3 Distribution of the Choices Made Behind the Veil of Ignorance - **Constant Efficiency Choices** 111
2.A4 Distribution of the Choices Made Behind the Veil of Ignorance - **A Pareto-Dominant project** 111
2.A5 Choice 1 (Constant Efficiency Choice) with Known Rank - by Rank Order 112
2.A6 Choice 2 (Constant Efficiency Choice) with Known Rank - by Rank Order 112
2.A7 Choice 3 (Constant Efficiency Choice) with Known Rank - by Rank Order 113
2.A8 Choice 5 (A Pareto-Dominant) with Known Rank - by Rank Order 114
2.A9 Choice 6 (A Pareto-Dominant) with Known Rank - by Rank Order 115
2.A10 Choice 7 (A Pareto-Dominant) with Known Rank - by Rank Order 115
2.C1 Screenshot of Choice Behind the Veil of Ignorance(Choice 2) 117
2.C2 Screenshot of Risk Aversion Elicitation 118
2.C3 Screenshot of Choice with Known Position (Choice 6, Rank 2) 119
2.C4 Screenshot of the Real Effort Task 119
2.C5 Comprehension test 120

3.1 Contract trade-offs assuming best responses 143
3.2 Workers' stated effort and principals' expected effort by piece rate wage 146
3.3 Principals' belief-based contract trade-offs 148
3.4 Principals' contract choices by treatment groups 149

3.5 Firm's profit if agents care about equal procedure and principals anticipate it correctly 162
3.6 Firm's profit if agents care about equal procedure and principal is naive 164
3.7 Firm's relative profit if agents care about potential ex-post income 165
3.8 Firm's relative profit if agents care about potential ex post income 166
3.A1 Distribution of individual ex post probabilities to be part of a given class 177
3.A2 The production and cost function per effort level and agent. 177
3.A3 Screenshot of a decision made by agent B. 178
3.A4 Screenshot of a decision made by the principal. 179
3.D1 Effort-production table *[Stakeholder treatment]* 182
3.D2 Effort-production table *[Spectator treatment]* 183
3.D3 Effort-production-cost table 184
3.D4 Effort-production table 185
3.D5 Effort-production-cost table 185
3.D6 effort-production-cost-income table 186
3.E1 Effort-production-cost table *[Stakeholder treatment]* 187
3.E2 Effort-production-cost table *[Spectator treatment]* 187
3.E3 Effort-production-cost table *[Stakeholder treatment]* 188
3.E4 Effort-production-cost table *[Spectator treatment]* 188
3.E5 Effort-production-cost table 190
3.E6 Effort-production-cost table 190
3.E7 Effort-production-cost table 191
3.E8 Principals' comprehension tests 192
3.E9 Workers' comprehension tests 192

4.1 Size of the ethnic ingroup bias by first mover ethnic group and experimental block . . 208
4.2 Expected and actual transfer from second mover to first mover by ethnic group of the second mover 215
4.3 Normalized High-Income Bias by first mover ethnicity i – US 221
4.4 Normalized High-Income Bias by first mover ethnicity i – Germany 221
4.5 Transfer decisions from first mover to second mover in the US and Germany by first mover income 223
4.6 Time to complete the survey by type of transfer decisions 229
4.A1 Transfer decisions from first mover to second mover in the US and Germany by second mover ethnic group 239
4.A2 Transfer decisions from first mover to second mover in the US by first mover ethnicity 240
4.A3 Transfer decisions from first mover to second mover in Germany by first mover ethnicity 240

List of Tables

1.1 Transition matrix of dominant tastes . . . 40
1.2 Master's degree choice and dominant tastes . . . 42
1.3 High-school taste intensity by (hand-labeled) dominant taste . . . 46
1.4 Determinants of change in taste: individual-level characteristics . . . 49
1.5 Determinants of change in taste: individual-level characteristics – heterogeneity by gender and socio-economic status . . . 51
1.6 Determinants of change in taste: performance position among peers . . . 52
1.A1 Hand-labeled field categories . . . 59
1.A2 Dominant taste and hand-labeled rank in high-school letters . . . 60
1.A3 Dominant taste and hand-labeled rank in second-year letters . . . 61
1.A4 Keywords . . . 62
1.A5 Average number of keywords per letter . . . 63
1.A6 Average number of keywords in high-school letters per (hand-labeled) dominant taste 64
1.A7 Average number of keywords in second-year letters per (hand-labeled) dominant taste 65
1.B1 Correlation between the students' and their tutorial group peers' dominant high-school tastes . . . 71
1.C1 Typology – category definitions . . . 75
1.D1 Students characteristics . . . 77
1.D2 Sciences Po Schools and master's degrees . . . 78
1.D3 Choice of specialist courses in the second year and dominant high-school tastes . . . 79
1.D4 Choice of elective courses in 2nd year and dominant high-school tastes . . . 80
1.D5 Determinants of change in taste: performance position among peers – students with tutorial groups only and students in dual degrees are also dropped . . . 81
1.D6 Determinants of change in taste: performance position within peers – heterogeneity by gender and socio-economic status . . . 82

2.1 Constant Efficiency Choices . . . 88
2.2 Pareto Comparable Choices. Project A is Pareto-Dominant . . . 88
2.3 Drivers of Inequality Aversion (Choice of Project B) - Logit Regressions on Pooled Data 93
2.A1 Balance Table . . . 104
2.A2 Who are the money burners? . . . 105
2.A3 Who are the people that are willing to pay for more equality? . . . 106
2.A4 Is choice of Project B *only* explained by risk? . . . 107
2.A5 Confidence and Over-Confidence do not Predict Choices . . . 108
2.A6 Choice table to elicit risk aversion using Eckel-Grossman method (Lottery 1) . . . 109
2.A7 Choice between two lotteries . . . 109
2.B1 Hypothetical situations . . . 116

3.1 Individualized wage raises and managers' distributive preferences . . . 138

3.2 Set of decisions made by the principal assuming workers' choose effort to maximize their own income . . . 142
3.3 Regressions that characterize Contract decisions . . . 151
3.4 Results from a pooled specification . . . 156
3.5 Results from the finite mixture model with three classes . . . 159
3.A1 Description of the main variables used in the REPONSE survey . . . 171
3.A2 Performance-based bonuses and managers' distributive preferences . . . 172
3.A3 Summary statistics Agents vs principal . . . 173
3.A4 Summary statistics principals Spectator vs Stakeholder . . . 173
3.A5 Regressions that characterize contract decisions using belief-based trade-offs and individual fixed effects . . . 174
3.A6 Regressions that characterize contract decisions . . . 175
3.A7 Regressions that characterize contract decisions . . . 176
3.B1 Results from the finite mixture model with three classes including outlier beliefs . . . 180
3.E1 True-False average score . . . 193

4.1 Trust game transfers in the US - Only second mover ethnic group is known . . . 211
4.2 Trust game transfers in Germany - Only second mover ethnic group is known . . . 212
4.3 Statistical vs taste-based discrimination . . . 214
4.4 Trust game transfers in the US - Second mover is rich (Ethnic plus Income block) . . . 218
4.5 Trust game transfers in Germany - Second mover is rich (Ethnic plus Income block) . 219
4.6 Comparison of the ingroup ethnic bias and the ingroup income bias when 2nd mover is in top 20% by income . . . 225
4.7 In group bias based on income . . . 226
4.B1 Summary statistics - US . . . 241
4.B2 Summary statistics - Germany . . . 242
4.B3 Ingroup bias at the country level in the Ethnic block . . . 243
4.B4 Individual determinants of ingroup ethnic bias . . . 244
4.C1 Description of the variables in *Trustlab*'s survey module . . . 246
4.C2 Attitudes by type of transfer decision - US . . . 247
4.C3 Attitudes by type of transfer decision - Germany . . . 248