Inequality as an Externality : Theoretical Insights and Public Sentiment

Morten Nyborg Støstad

Universite Paris I Panthéon-Sorbonne – École doctorale d'économie
Paris School of Economics

Ph.D Thesis

Submitted for the Degree of Doctor of Philosophy in Economics

Prepared and defended on September 8th 2023 by

Morten Nyborg Støstad

Inequality as an Externality: Theoretical Insights and Public Sentiment

Thesis Advisor: Stéphane Gauthier
Professor at the Paris School of Economics
Professor at Université Paris 1 Panthéon-Sorbonne

Members of the Jury

President	Nicolas Jacquemet, Professor, Paris 1 Panthéon-Sorbonne
Reviewer	Laurence Jacquet, Professor, CY Cergy Paris Université
Reviewer	Olof Johansson-Stenman, Professor, Univ. of Gothenburg
Member	Emmanuel Saez, Professor, U.C. Berkeley
Member	Marc Fleurbaey, Professor, Paris School of Economics

Universite Paris I Panthéon-Sorbonne – École doctorale d'économie
Paris School of Economics

THÈSE

Pour l'obtention du titre de Docteur en Économie

Présentée et soutenue publiquement le 8 Septembre 2023 par

Morten Nyborg Støstad

Essais sur l'Inégalité comme une Externalité

Sous la direction de M. Stéphane GAUTHIER
Professeur, Paris School of Economics
Professeur, Université Paris 1 Panthéon-Sorbonne

Membre du Jury

Président	Nicolas Jacquemet, Professeur, Paris 1 Panthéon-Sorbonne
Rapporteure	Laurence Jacquet, Professeur, CY Cergy Paris Université
Rapporteur	Olof Johansson-Stenman, Professeur, Univ. of Gothenburg
Examinateur	Emmanuel Saez, Professeur, U.C. Berkeley
Examinateur	Marc Fleurbaey, Professeur, Paris School of Economics

Summary

The primary topic of this doctoral thesis is the connection between the consequences of economic inequality – how inequality could affect the amount of social unrest, trust, or democratic institutions, for example – and economic redistribution. Each of the three chapters discusses this link in some way. Chapter One is theoretical, introducing such consequences into standard economic frameworks and showing why they imply that economic inequality itself is an externality. Chapters Two and Three are based on empirical large-scale surveys and explore public sentiment on the issue. As a whole, the thesis sheds light on how the consequences of economic inequality influence our willingness to redistribute, both from an optimal and real-world perspective.

In Chapter One, co-authored with Frank Cowell, the focus is on how the consequences of economic inequality impact traditional economic models. The paper considers such consequences as an externality and notes that this differs from traditional equity-based inequality concerns. It also implies a mathematically distinct rationale for government redistribution. Through the classical Mirrlees optimal income taxation model, theoretical and simulation-based analysis demonstrates that the externality particularly affects optimal top marginal tax rates. These tax rates may in our analysis exceed 90%. The paper further illustrates that the current U.S. tax system is not sufficiently redistributive to both value income at the bottom more than at the top *and* to consider inequality as having even moderate negative consequences. The overarching comment of the paper is that large swaths of economic theory has implicitly assumed that economic inequality has no meaningful consequences beyond purely distributional concerns.

Chapter Two, co-authored with Max Lobeck, presents the first known empirical research on U.S. citizens' beliefs regarding these consequences of economic inequality. This study utilizes large-scale surveys conducted by the authors with a total of 6,731 respondents in the United States. The findings reveal that a majority of individuals perceive economic inequality to have significant and diverse negative consequences. Through an information experiment, the paper demonstrates that these beliefs exert a substantial causal influence on individuals' preferences for redistribution. The study also finds indicative evidence that beliefs about the consequences of economic inequality are less divided across incomes and political parties than comparable economic fairness views, and that such fairness-based arguments induce more anger in respondents.

Chapter Three, also co-authored with Max Lobeck, develops a novel methodology to evaluate properties of classes of statements or arguments. This method is used to rigorously test whether fairness-based redistributive arguments are structurally different from inequality externality-based redistributive arguments. The underlying idea is that such differences could have lead to significant cross-country differences in policy, polarization, and public redistributive sentiment. The paper uses three surveys with a total of 4,444 respondents. The first two surveys gather and quality-check a large unbiased sample of fairness-based and externality-based arguments which are evaluated across various dimensions in the third survey. The paper strengthens the finding that fairness-based arguments are comparatively more anger-inducing, and indicates that this is because these arguments are more normatively based than arguments about inequality's consequences.

In brief, the thesis formalizes the concept of inequality as an externality and explores this efficiency-based reason to redistribute in both theoretical and real-world settings.

Résumé

Le sujet principal de cette thèse de doctorat porte sur l'impact des inégalités économique sur les individus et la société, ainsi que sur la manière dont impact pourrait ou devrait influencer le niveau de redistribution économique. Chacun des trois chapitres explore, sous différents angles, cette relation entre les conséquences des inégalités économiques et la redistribution.

Dans le Chapitre Un, réalisé en collaboration avec Frank Cowell, l'accent est mis sur l'impact des conséquences de l'inégalité économique sur les modèles économiques traditionnels. L'article introduit le concept d'inégalité économique en tant qu'externalité, ce qui constitue une nouvelle justification pour la redistribution gouvernementale en complément des arguments d'équité déjà existants. À travers une analyse théorique adossée à des simulations du modèle classique de taxation optimale des revenus de Mirrlees, l'étude démontre que cette externalité affecte particulièrement les taux marginaux supérieurs optimaux, pouvant dépasser 90%. L'article illustre en outre que le système fiscal actuel aux États-Unis n'est pas suffisamment redistributif pour valoriser à la fois la redistribution du haut vers le bas et pour prendre en compte que l'inégalité a des conséquences négatives modérées.

Le Chapitre Deux, réalisé en collaboration avec Max Lobeck, présente une nouvelle recherche empirique sur les croyances des citoyens américains concernant les conséquences de l'inégalité économique. Cette étude utilise des enquêtes à grande échelle réalisées par les auteurs, impliquant un total de 6,731 répondants aux États-Unis. Les résultats révèlent que la majorité des individus perçoivent des conséquences négatives significatives et diverses liées à l'inégalité économique. À travers une expérience d'information et d'autres méthodes, l'article démontre que ces croyances ont une influence causale importante sur les préférences des individus en matière de redistribution. L'étude constate également que les croyances concernant les conséquences de l'inégalité économique sont moins expliquées par le revenu et les partis politiques que les points de vue comparables sur l'équité économique, et que de tels arguments basés sur l'équité suscitent également plus de colère chez les répondants.

Le Chapitre Trois, également réalisé en collaboration avec Max Lobeck, teste la validité externe de plusieurs des résultats du Chapitre Deux. L'article utilise une méthodologie novatrice à travers trois enquêtes totalisant 4,444 répondants pour recueillir un large échantillon d'arguments basés sur l'équité et les conséquences des inégalités, puis demande à d'autres répondants de les évaluer selon diverses dimensions. L'article renforce la constatation selon laquelle les arguments basés sur l'équité suscitent la colère, en grande partie en raison de leur fondement plus normatif que les arguments comparables sur les conséquences de l'inégalité.

Acknowledgments

I dedicate this thesis to my maternal grandparents, Anne Alvik and Per Nyborg, who have inspired much of this work and mean the world to me.

I am deeply grateful to my advisor, Stéphane Gauthier, who has given me invaluable advice throughout my PhD. Stéphane's academic knowledge has been crucial in many problems I have faced during the thesis, and I am very thankful for all the time and effort he has devoted to making this thesis what it is today.

I could not have done the research I did without Emmanuel Saez, who invited me to Berkeley and has been extremely generous with his time, insights, and financial support. Our many conversations have shaped me as a researcher. I am indebted to Marc Fleurbaey, who helped me when I was at one of my lowest points of the thesis and has given me hours of invigorating discussions full of valuable ideas. I also owe a strong debt of gratitude to Claudia Senik, who both gave me early insights into economic research and consistently delivered unique, insightful feedback when I began to develop projects of my own.

I would like to thank Frank Cowell, co-author of Chapter One and my Master's thesis advisor at LSE, for both his valuable contributions and his push to making Chapter One more than a Master's thesis. I am grateful to Daniel Waldenström, my Master's thesis advisor at PSE, who helped me develop an understanding of how difficult it is to pin down the empirical effects of inequality on other factors. Tremendous thanks is owed to Antoine Bozio, whose kind words, advice, and extensive knowledge about Public Economics has been extremely valuable. A similar gratitude is due to Thomas Piketty, whose work originally influenced me to study inequality-related issues; receiving his insights on my work has been an extraordinary privilege and an invaluable part of my academic journey. I also particularly thank Etienne Lehmann, who has been a crucial help for many of the optimal tax-related problems I have encountered.

I am especially indebted to my two referees, Olof Johansson-Stenman and Laurence Jacquet, who have improved the final product of this thesis dramatically. I am also grateful to Nicolas Jacquemet for his advice on the behavioral aspects of the dissertation and for agreeing to be part of the thesis jury.

I would also like to thank the many professors I have discussed various aspects of this project with at different times, in no particular order; David Margolis, Fredrik Carlsson, Nathaniel Hendren, Camille Hémet, François Fontaine, Andrew Clark, Sylvie Lambert, Tanguy van Ypersele, Nicolas Gravel, Mathieu Lefebvre, Katheline Schubert, Shachar Kariv, Alan Auerbach, Stefano DellaVigna, Evan Friedman, Stefanie Stantcheva, Gabriel Zucman, Dmitry Taubinsky, Louis Kaplow, and many others. I am grateful to everyone who gave comments on my work during the academic job market, and I am particularly thankful to Angelo Secchi and Roxana Ban for their help preparing me for that experience.

Max Lobeck, who co-authored the last two Chapters of this thesis and has been a fantastic co-worker and friend, deserves a special mention. The work contained within this thesis has as much of Max' intellectual influence as anyone else's, and I value our collaborations greatly. Several others have also been both close friends and large influences on my academic work; Elif Cansu Akoğuz, Kieran Byrne, Thomas Blanchet, Amory Gethin, Sébastien Laffitte, and Eddy Zanoutene particularly so. I am also grateful to Eddy Zanoutene for his help in improving the

French parts of this dissertation.

My doctoral experience would not have been the same without all the wonderful PhD students and friends I met along the way at PSE, Berkeley, AMSE, and everywhere else across the world. It would take a book to write about everyone who have left a significant imprint on me during this time. If you find yourself on this list, please know that in one way or another I have a strong appreciation for you. I am deeply grateful in a wide variety of ways to Simon, Ander, Marie, Paula, Hugh, Luiz, Tom, Theresa, Luis, Clara, Rowaida, Alvaro, Nitin, Mélusine, Marc and Naomi, Ignacio, Mauricio, Olivia, Dario, Ana, Paolo, Yajna, Sam and Ellen, Dessie, Tiziano, Caroline, Fausto, Anika, Aliénor, Nick, Emily, Stian, Jonas, Karl-Erik, Daniel, Vilde, Camille, Eva, Wouter, Duncan, Elias, Kathy, Ruslana, Olatz, Fédé, Oscar, Rind, Manon, Martin, and many more I'm sure I've forgotten. I also particularly want to thank Hanaë and Nazim for their company and endless helpful actions in Paris, and Lyon and Mags for being who they are and welcoming me into their wonderful home in San Francisco. And of course Marion Leroutier and Caroline Coly for dragging me kicking and screaming through the academic job market.

I am grateful to the many institutions and associations who have supported my research; PSE, Paris 1, CEPREMAP, the WIL, IRLE at Berkeley, AMSE, the Agence Nationale de la Recherche, the Deutsche Forschungsgemeinschaft, Ingegerd og Arne Skaugs Forskningsfond, Lånekassen, the University of Konstanz, and the U.C. Berkeley James M. and Cathleen D. Stone Center on Wealth and Income Inequality.

Most of all, I want to thank my parents, Karine Nyborg and Jan-Erik Støstad. They have always been my closest advisors, and have been there through the highs and the lows of the dissertation with unwavering love and care. Finally, I thank my siblings Anita, Hanna, and Mads, as well as their families, who make Oslo feel like home when I am often so far away.

Table of Contents

General Introduction

"*In a state which is desirous of being saved from the greatest of all plagues, [...] here should exist among the citizens neither extreme poverty, nor, again, excess of wealth, for both are productive of these evils.*" – Plato (360 B.C.)

Why do we care about the economic differences between people? It's a surprisingly complicated question, and nearly all of us have our own answers. These different answers lead to diverging perspectives; some of us care deeply, others not so much. While certain experts tell us to reduce economic inequalities at almost any cost, others think of it as a necessary and relatively benign evil. The question of *why* we care about economic inequalities reverberate into our elections and policies, into our purchasing power, and into our societies.

This dissertation revolves around this question. It focuses on an angle which is relatively unexplored within the economic literature, namely the *societal consequences* of economic inequality. Many examples of such consequences are possible. The amount of economic differences could affect the trust between people, institutions, and communities, for example, or the constraints and quality of public governance. In what follows I will explore many such potential consequences, and how their existence affects economic theory and individuals' actual preferred levels of redistribution.

The backdrop for this work is the contentious debate on economic inequality and redistribution. This debate has long been a mainstay in the political and public arena. At its core is again the question of why economic differences matter – and the natural follow-up of why we choose not to reduce them. The traditional trade-off is that of equity and efficiency. Redistributing from the rich to the poor is generally seen as promoting equity. This comes at an efficiency cost, however, as reducing inequality through taxation carries a potentially large associated deadweight loss. This equity-efficiency trade-off has been at the core of public policy and inequality-focused economic research for decades, and is illustrated in Arthur Okun's classic *Equality and Efficiency: The Big Tradeoff* [Okun, 1975].

In this thesis I consider a different approach. As illustrated by the introductory statement by Plato, it is often suggested that economic inequality affects society beyond its purely distributional effects. These consequences of economic inequality could affect all of us. Deteriorating trust and dysfunctional political systems are only two of many examples; heightened social unrest, increased crime, and a proliferation of corruption are a few more. In economic terminology, such consequences would imply that economic inequality itself is an *externality*. We all influence the level of economic inequality through our market actions, yet we do not normally take into account how this inequality affects others. This indicates a classic public goods problem and suggests a potentially significant reason to redistribute beyond any standard equity concerns.

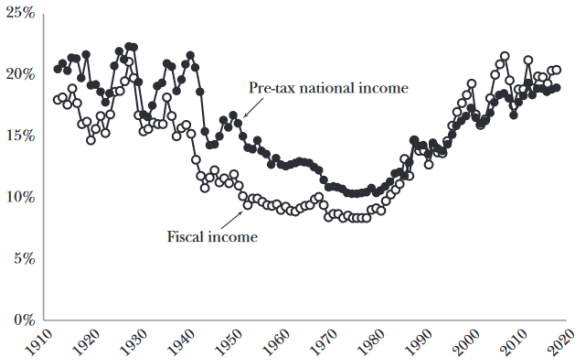
The presence of this externality modifies the traditional equity-efficiency trade-off into a more complicated balancing act, the form of which is not immediately obvious. Both the government's optimal policy path and the individual's preferred level of redistribution could be affected. Indeed they are, as we will show in due time. The main goal of this thesis is to discuss these implications to the redistributive problem; first from the optimal policy makers' point of view in Chapter One, then from the behavioral changes to the individual's preferences for redistribution

in Chapters Two and Three. I will now introduce the topic more generally before discussing the contributions of the dissertation.

The Rise of Economic Inequality

The context of this thesis is the widely publicized rise in within-country income and wealth inequality in much of the world in the last fifty years. The increase in the United States has been particularly notable, as indicated in Figure 1. Partly due to this large increase, the United States is used as the laboratory setting for the numerical simulations in Chapter One and the surveys conducted in Chapters Two and Three.

Figure 1: Share of U.S. Income Earned by the Top 1%



Note: From Saez and Zucman [2020].

Economic inequality is a global problem, however, and the recent rise of within-country inequality is not a phenomenon unique to the United States. Since 1970, the pre-tax income inequality increase in the U.S. has been outpaced by the increases in Mexico, Italy, China, India, and Russia, among others. There have also been more moderate increases across the world; Germany, Japan, South Africa, and Canada are just a few examples of countries where the top pre-tax income share has increased noticeably in the period.[1]

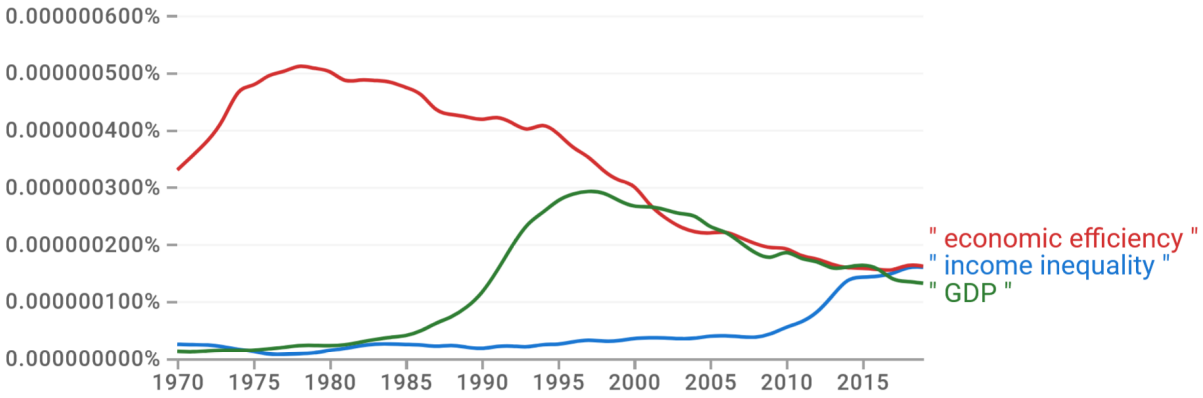
Within-country wealth inequality has risen in a similar fashion. In the United States, the top 1% wealth share has risen from 23.2% in 1980 to 35.3% in 2019. In Russia the top 1% wealth share more than doubled since 1995 to 2019, increasing from 21.5% to 47.6%. In India, the same indicator rose from 23.2% to 33.7% during the same time period.

This has not gone unnoticed. World leaders have denounced the rise in economic inequality, illustrated here by a comment of António Guterres, secretary-general of the United Nations as of this writing; "*Inequality defines our time. More than 70 per cent of the world's people are living with rising income and wealth inequality*" [Guterres, 2020]. The academic interest in the topic has exploded; a cursory search for "income inequality" on Google Scholar finds over a million related academic works.[2] Thomas Piketty's 2014 book "*Capital in the Twenty-First Century*" [Piketty, 2014] became a global phenomenon and topped The New York Times Best Seller list for hardcover nonfiction. In the Google Ngrams Books collection, the frequency of the phrase "income inequality" has risen rapidly since 2010 and had as of 2019 (barely) overtaken the frequency of both "economic efficiency" and "GDP", as shown in Figure 2.

This is not just an academic phenomenon. In the aforementioned Okun [1975], Arthur Okun noted that "*public criticism or even discussion of income inequality is rare, perhaps because differences in incomes arise so naturally.*" This is not a comment that would be as readily made today. In their 2014 Global Attitudes Survey, the Pew Research Center [2014] asked citizens of 44 countries to rank the "greatest dangers in the world". In the United States and most of

[1] Data from wid.world.
[2] The search garnered 1,090,000 results on July 7th 2023.

Figure 2: Frequency of phrases in Google Books Ngrams database



Europe, the most common answer was inequality, ahead of religious and ethnic hatred, nuclear weapons, and environmental degradation.

Still, a salient question is why we should have a problem with rising economic inequalities at all. In a review of Piketty's aforementioned bestseller, Martin Wolf of the Financial Times notes that the book "*does not deal with why soaring inequality [...] matters. Essentially, Piketty simply assumes that it does*" [Wolf, 2014]. This is an especially common criticism if there is a trade-off between economic equality and economic growth; as Nobel prize winner Robert Lucas famously put it, "*Of the tendencies that are harmful to sound economics, the most seductive, and in my opinion the most poisonous, is to focus on questions of distribution. [...] The potential for improving the lives of poor people by finding different ways of distributing current production is nothing compared to the apparently limitless potential of increasing production.*" [Lucas, 2004].

There are many ways to answer such criticisms, and many reasons to dislike inequality. However, detailing all the potential reasons in a rigorous manner is not as straight-forward as it might seem. From a certain point of view, these critics have a point; why would we care about the income distribution itself if every person's income is growing?

There are many strong counter-arguments to this point of view, but perhaps the strongest is the main topic of this dissertation, namely the idea that inequality itself could change our societies. This is an interesting but not entirely uncomplicated idea; it seems to deserve a more systematic exploration than what is currently present in the economic literature. In the remainder of the thesis I will attempt to provide this exploration and add a building block to the traditional economic toolkit, with the overall aim of introducing nuance to the general discussion of why we should care about economic inequality. Before that, however, I begin with the traditional explanation for why we should care and why it might be difficult to do so; the equity-efficiency trade-off.

Equity and Efficiency

"[Equality versus efficiency] is, in my view, our biggest socioeconomic tradeoff, and it plagues us in dozens of dimensions of social policy. We can't have our cake of market efficiency and share it equally." – Arthur Okun (1975)

Unequally distributed income and wealth leads to individual suffering due to a lack of re-

sources. In the economic sense, income is inefficiently distributed according to a utilitarian welfare maximization criterion when there are unequal marginal utilities of income between individuals. This, or equivalently a social planner that simply prefers lower-income agents to receive income over high-income agents, represent the intuition behind the mathematical formulations of the classic *equity* channel from Okun [1975].

This requires a small detour into mathematical formulations. Theoretical economic analysis often assumes that personal well-being (utility) is comparable across individuals, following in the Benthamite tradition. Furthermore, it is often assumed that total welfare only depends on a combination of individuals' levels of well-being. Combined, this means that total welfare can be calculated and compared across different policy options (assuming we can calculate every individual's well-being in every policy option). This comparison, in the welfarist tradition, is usually done with what is called a *social welfare function*.

This can be expressed mathematically, which is the starting point for a myriad of economic problems. A standard welfare (W) maximization problem in optimal taxation, for example, can look like;

$$max_{T(z)} \int_{\underline{z}}^{\bar{z}} W\left(U_i(x_i, z_i, ...)\right) di, \tag{1}$$

where individual i's well-being U_i is a function of pre-tax income z_i which is distributed from $\underline{z}$ to $\bar{z}$ and is taxed through the tax system $T(z)$ such that final consumption is $x_i = z_i - T(z_i)$. The individual exerts effort (disutility) to earn income z_i, such that $\frac{\partial U_i}{\partial z_i} < 0$,[3] but gains utility from consumption x_i, such that $\frac{\partial U_i}{\partial x_i} > 0$. The total welfare in this state of the world is calculated with the social welfare function W – if this is a Bergson-Samuelson social welfare function, for example, this is just a weighted sum of all the individual utilities.

There are two clear motives for inequality reduction in this framework, which also reoccur in cost-benefit analyses, macroeconomic welfare calculations, and so on.

1. Individuals' diminishing marginal utilities of income, indicating that those with already-high incomes have less need for one more unit of income. Mathematically this is formulated as $\frac{\partial^2 U_i}{\partial x_i^2} < 0$.

2. Decreasing social welfare weights, which means that the social planner (usually the government) might simply value additional welfare more when that welfare goes to low-welfare individuals. Mathematically this is formulated as a strictly concave social welfare function such that $\frac{\partial^2 W(U)}{\partial U^2} < 0$.

These two motives both punish the existence of economic inequalities and push the social planner towards designing a redistributive tax system.[4] Raising tax rates can be one solution; at the same time, high tax rates could also make people reduce their work effort z, thus *reducing* tax revenue. Together these factors represent the traditional equity-efficiency trade-off.

[3] This is the standard formulation in optimal taxation, but the problem could equally be formulated as the individual disliking work effort h which increases pre-tax incomes $z(h)$.

[4] In practice the two concepts discussed here are heavily related. From the social planner's view, the net effect of an agent receiving one more unit of income is the combination of their utility benefit and how much this benefit is discounted or amplified by the social welfare function.

This framework is relatively flexible, which is partly why it is often used in policy-related work. A classic benchmark in the literature is the "Rawlsian" min-max, where the social planner maximizes the welfare of only the worst-off agent. (The name is nominally from Rawls [1971] where Rawls describes a philosophical viewpoint that is relatively more nuanced.) This is often considered the most inequality averse a social planner can be. As such, the idea is useful as one of two benchmarks for economic policy. The second benchmark is a purely "Utilitarian" point of view, where the social planner puts the same value on every unit of utility regardless of which individual receives it.[5] These two poles of analysis has been at the core of welfarist traditions and economic theory for decades.

Still, something may be missing. The individual, with their utility (well-being/preference) function $U_i(x_i, z_i, ...)$, is indifferent to the level of economic inequality in society. Note, for example, that there is no sensible reason to prevent top incomes from growing without bounds in this framework. If x_{top} increased, the welfare of the top-income individual is increased and no other individual notices a change. This leads to an unambiguously non-negative change in welfare – regardless of how high x_{top} is. Under this formulation, the only exception is if the government prefers to actively reduce these people's well-being, which seems unreasonable.

This has lead many economic models to find solutions focused on *efficiency* where top incomes are extremely high. All agents are heavily incentivized to work, then the state redistributes (which is easier since incomes at the top are high). That way, bottom incomes are also slightly higher than they otherwise would have been. In some ways, it seems like we have achieved both efficiency (high incomes) *and* equality (high bottom incomes).

But this conclusion is immediately perturbed by the realization that economic inequality itself is high. This has happened because inequality *per se* has no place in the trade-off. In this framework there is no cost associated to the large economic differences between people. While individuals might simply prefer more equal societies, I am personally particularly concerned about the case where economic inequality changes our lives independently of what we think or feel; when the differences between us shape our societies and social interactions. I turn now to the consequences of economic inequality.

The Consequences of Economic Inequality

Concerns about the consequences of economic inequality have been present from the ancient Greeks. Beyond Plato's views, we also have similar accounts from his contemporaries. Around the 2nd century AD, Plutarch wrote in "*Parallel Lives*" that in Athens "*the disparity between the rich and the poor had culminated, as it were, and the city was in an altogether perilous condition.*"[6] In "*Politics*", Aristotle argued that "*it is clear then that those states in which the middle element is large, and stronger if possible than the other two(wealthy and poor) together, or at any rate stronger than either of them alone, have every chance of having a well-run constitution.*"[7] These ideas continue well into our own time; Obama [2011] contended that "*This kind of inequality – a level that we haven't seen since the Great Depression – hurts us all,*" and Pope

[5] Note that the Utilitarian benchmark in utility can still contain significant income inequality aversion. What is often used in practice is the Utilitarian benchmark in income, however, where every unit of income is valued equally by the social planner.

[6] Quotation from Bernadotte Perrin's translation, Plutarch [1923].

[7] Quotation from Benjamin Jowett's translation, Aristotle [1885].

Francis [2014] asserted that "*Inequality is the root of social evil.*" There are many more examples which I will not list here; to be brief, it seems as if almost every influential societal thinker has at some point noted that economic inequality has the potential to change our communities.

Indeed, economic inequality could affect individuals in a myriad of ways. Kate Pickett and Richard G. Wilkinson popularized these ideas in their 2011 book *the Spirit Level* [Wilkinson and Pickett, 2009], where they argued that inequality is correlated to (and causes) worse health outcomes, lower levels of trust, higher rates of mental illness, increased levels of violence and crime, reduced social mobility, and diminished educational performance. They further posited that these negative effects are not confined to the poor, but permeate all levels of society, arguing for the integral connection between economic inequality and societal well-being.

Wilkinson and Pickett's empirical work has been controversial. And even in the best case, correlational plots are no proof of causality. While many academic papers have attempted to establish a *causal* connection between economic inequality and various outcomes, this literature has several empirical challenges. Largest among them is the lack of obvious exogenous variation for macroeconomic inequality. In other words, any time economic inequality changes, something else has changed as well (e.g. the political and social context that lead to a tax decrease). This makes it hard to know whether inequality itself is the causal factor, and is a severe obstacle for the robust empirical detection of any consequences of economic inequality. It is my belief that convincingly detecting such effects is almost impossible, regardless of how large they might be.

Detecting correlational relationships between inequality and negative outcomes is possible, as noted, even if it faces additional challenges – measurement issues, the question of which inequality metric to use, the lack of large variation in economic inequality over time, the question of whether perceived or actual inequality is more impactful, potentially time-lagged and non-linear effects, the ecological fallacy (the need to control for individual income), and so on. There are large literatures attempting to achieve this correlational goal. For reviews of such literatures see Rufrancos et al. [2013] on inequality and crime or Bergh et al. [2016] on inequality and individual health.

A more precise avenue is laboratory experiments. In recent years, a substantial body of experimental and microeconomic research has demonstrated that economic inequality among workers or participants in experiments has significant effects on stated life satisfaction [Card et al., 2012], productivity [Breza et al., 2018], trust [Fehr et al., 2020b], and cooperation [Xu and Marandola, 2022]. Although potentially precise, such methods have limited external validity and are often unsuitable for channels that are difficult to model in microeconomic settings (such as the effect of inequality on social unrest or economic growth).

Empirical detection aside, how would such consequences of inequality occur? I will first illustrate a few examples where the perception of income differences changes social behavior. Differences that are perceived as too large could induce protest movements and social unrest from those who feel left behind. A distrust towards others who are perceived to have a very different economic status could create social cleavages; cynicism could emerge towards formal establishments, driving political dysfunction and heightening political polarization. Frustrations with a perceived loss of status could drive unhealthy habits and instigate health problems, and a lack of cooperation could deteriorate shared culture. Technological *improvements* could emerge,

as perceived high levels of inequality could function as an incentive to increase labor supply and induce high risk-taking among innovators.

But it's not only perceived inequality that could influence our lives. Although economic inequality is in some sense an abstract concept, it can directly influence societal factors we care about in an almost mechanical fashion. To illustrate this I create micro-foundations of how economic inequality could affect political polarization, demands for public goods, trust, crime, and more in Chapter One.

As an example, think about the willingness of individuals to contribute to public goods. Suppose there is an interaction between income and public good preferences such that the rich and poor have different tastes for public projects. The rich could ideally want opera houses, for example, while the poor prefer public housing. An unrelated rise in economic inequality would lead to less agreement and thus fewer funded public goods. This hurts everyone; we are less able to band together to commit to investments that benefit us all. In more unequal societies the same concept could lead to severe conflicts of interest. These conflicts can culminate in resource-powerful lobbying interests subverting public opinion and distorting democratic processes.

In general, coordination difficulties and conflicts of interest both increase when economic inequality increases. Simply put, when individuals are different it is more difficult to do good policy design. In economic theory, models with heterogeneous agents are more challenging to solve. Inequality's consequences is a real-world application of the same concept.

The preceding paragraphs has made two points clear. First, the consequences of inequality could be significant. Second, it is not empirically trivial to measure what these consequences are. This poses an issue for academic explorations on the consequences of inequality. It is possible to examine a specific channel – a few well-known works are Benabou [1996] for inequality and economic growth or Bourguignon [1999] for inequality and crime, for example – but the more general problem of how to model income or wealth inequality's consequences in economic frameworks has remained unclear. This has, in practice, lead to the consequences of inequality being largely ignored in both optimal policy frameworks and in the empirical literature on individuals' preferences for redistribution.

Inequality as an Externality

This introduces the main focus of Chapter One and an overarching theme of this thesis, which is the concept of economic inequality as an externality. In the economic literature, an externality is present when an individual's market action affects other individuals despite them not being involved in the transaction [Buchanan and Stubblebine, 1962]. The classic example is pollution. When companies pollute, they impose an externality on the rest of society which require third-party action (e.g. a tax) for compensation. In this thesis, and particularly in Chapter One, I discuss why economic inequality itself is an externality.

This notion is built on the following intuition. We all affect economic inequality through our market actions, as such actions affect our own incomes or wealth and thus economic inequality. A simple example of such actions is our education and labor choices. If economic inequality in turn affects something else we care about, then our market actions also affect others through these consequences. The net effect is that our individual actions – which may be optimal for us, as is often assumed in economic models – have an externality effect on others through the level

of inequality. In sum, economic inequality itself is an externality.

This formulation is beneficial for two main reasons. First, it simplifies what is a complex and multi-faceted issue into a single tractable concept. Second, it allows us to merge two well-known ideas within the economic literature, namely economic inequality and the concept of externalities. Combining Marx and Piketty with Pigou, as it were.

This can be illustrated mathematically. When inequality is an externality, the social planner's maximization problem can be written as,

$$max_{T(z)} \int_{\underline{z}}^{\bar{z}} W\left(U_i(x_i, z_i, \Gamma(\theta(\mathbf{x})), ...)\right) di, \tag{2}$$

where income inequality $\theta(\mathbf{x})$ is a function of the distribution of post-tax incomes $x = z - T(z)$ of every individual denoted by $\mathbf{x}$, which affects other factors the individual cares about (e.g. crime) through the function Γ.

The core difference of this formulation to that in (1) is that individuals' incomes also affect others directly through $\theta(\mathbf{x})$. Whereas the standard equity-efficiency trade-off as pertains to optimal taxation is in truth a selection problem,[8] the introduction of $\Gamma(\theta)$ also imposes an externality dimension; a wedge has been created between privately optimal decisions and the social optimum. In practice this changes the associated trade-offs drastically. These ramifications are discussed in Chapter One. I will now discuss this paper and its contribution, first taking a short detour to discuss the optimal taxation literature.

Optimal Taxation

The economic literature on redistribution often focuses on the concept of *optimal taxation*, or the optimal tax rate the government should set under various assumptions. The modern optimal income taxation literature is heavily influenced by Mirrlees [1971]. Mirrlees evaluated the equity-efficiency trade-off under a nonlinear income tax schedule; the framework builds on what I discussed in the section on equity and efficiency.

A short review of major results in optimal taxation is in order. The first era of optimal taxation results often justified what now looks like regressive tax policy. Sadka [1976] and Seade [1977] found that the Mirrlees model implies that the optimal marginal tax rate at the top should be zero, for example.[9] This was true even under a Rawlsian social planner (seemingly the most inequality-averse one could be). This was politically and academically controversial, but did not become a mainstay in the literature; the result is both fragile [Stiglitz, 1982] and very local [Saez, 2001]. Other results have been more significant in shaping actual policy. The Atkinson and Stiglitz [1976] theorem states that direct income taxation and other forms of indirect taxation are equivalent under what seems like mild separability assumptions. This famous result provided a theoretical foundation for arguments against capital taxation. Meanwhile, the famous "Laffer curve" – drawn on a napkin by Arthur Laffer for president Ronald Reagan, illustrating how tax revenue is maximized somewhere between 0% and 100% – popularized the idea that there is a level above which taxation reduces tax revenues and is thus seemingly irrational. This latter

[8] This means that government's main constraint is to incentivize individuals to self-select into the work effort that maximizes tax revenue without observing those individuals' intrinsic ability – see [Stiglitz, 1982] for more.

[9] They also found that the optimal marginal tax rate at the bottom should be zero and that the optimal marginal tax rate is bounded between zero and one.

argument, based on the same framework as the Mirrlees model, has had a long-lasting legacy as an argument for cutting top tax rates.

The modern literature on optimal taxation, by contrast, arguably favors more progressive tax policies. This modern literature began with the seminal contributions of Diamond [1998] and particularly Saez [2001], which brought Mirrlees' theoretical idea closer to real-world tax design. Saez [2001] reduces the complex mathematical problem into a search for empirical elasticities that are estimable in real-world settings; as a result, economists can calculate revenue-maximizing tax rates under relatively few assumptions. The revenue-maximizing top marginal tax rate has been found to be relatively high, with estimates roughly ranging from 65% to 75% [Saez, 2001, Piketty et al., 2014].

The model as a whole is very popular in the academic literature. In recent years, optimal taxation has been used to academically study the trade-offs implicit in top income taxation [Piketty et al., 2014], the taxation of couples [Kleven et al., 2009], optimal capital taxation [Saez and Stantcheva, 2018], the trade-offs between capital and wealth taxation [Guvenen et al., 2019], the optimal reaction of governments to migration responses [Lehmann et al., 2014], the effect of extensive and intensive margins on the optimal tax rate [Saez, 2002, Jacquet et al., 2013], and much more. Optimal taxation ideas have been discussed in the New York Times[10] and the Washington Post,[11] and economists have frequently advised governments or political hopefuls with these ideas in mind.

The workhorse model, which most of the above papers and many more are based on, makes many assumptions for simplicity. Two assumptions are crucial as it pertains to Chapter One.

First, the assumption of no relevant externalities. While the literature has extensively explored the introduction of externalities into optimal taxation models – see Sandmo [1975], Oswald [1983], Bovenberg and van der Ploeg [1994], Cremer et al. [1998] and Kopczuk [2003] for examples, many of which are related to Chapter One – the majority of works in the field assume no externalities unless the topic is explicitly externality-related.

Second, the assumption that inequality itself does not affect the individual's well-being beyond their individual income. This is usually justified on philosophical grounds, as it is not clear that the social planner should take into account preferences that directly include other individuals' well-being or incomes, also known as other-regarding preferences [Harsanyi, 1977, Goodin, 1986]. In short, although relative income concerns, altruism and jealousy are empirically established to matter for individuals [Cooper and Kagel, 2016], these preferences might not be relevant when crafting optimal policy. The easiest examples come from negative emotions – it is not clear that we should tax a billionaire simply because others are jealous, for example.

The two assumptions are related. In fact, the mathematical analysis developed in the case of other-regarding preferences implies an externality dimension. In any case, either of these assumptions are problematic when economic inequality is considered an externality.

The remaining question is whether these assumptions have been impactful enough for an inequality externality to change model conclusions. The theoretical aspect of this question is detailed in Chapter One, co-authored with Frank Cowell, where we discuss the effect of treating income inequality as an externality on the optimal income tax rates deduced from the Mirrlees

[10] https://www.nytimes.com/2019/01/05/opinion/alexandria-ocasio-cortez-tax-policy-dance.html
[11] https://www.washingtonpost.com/news/wonk/wp/2012/11/27/should-the-top-tax-rate-be-73-percent/

model.

Chapter One: *Inequality as an Externality: Consequences for Tax Design*

Chapter One, *Inequality as an Externality: Consequences for Tax Design*, was co-written with Frank Cowell from the London School of Economics.

The paper introduces various inequality externalities, focusing on a post-tax income inequality externality, into the Mirrlees model. The addition means that there is a direct motive for income equality in the model. This changes the mathematical structure and the implicit trade-offs of the model, and has a particularly strong effect on optimal top tax rates. The strong effect on top tax rates is due to an interesting mathematical feature. When trying to maximize a sum of incomes, there is always a trade-off in setting top tax rates; the classical equity-efficiency dilemma. But when optimizing for equality itself, top taxation is now doubly effective. The normal equality channel (mechanical redistribution) is still beneficial, as redistributing from the top to the bottom is inequality-reducing. But what was previously an efficiency channel – top-income individuals reducing their work effort – is now also beneficial because it reduces income inequality. This changes the trade-offs of the model dramatically.

This theoretical finding is confirmed in numerical simulations, where optimal tax rates are particularly affected. Optimal tax rates can reach above 90% under our estimates, approaching the tax rates in the United States and United Kingdom after World War II. Such high tax rates are inefficient and irrational in the standard model; when inequality is a significant negative externality they can be rationalized quite easily.

As a general point, the existence of a negative inequality externality creates a new incentive for the government to redistribute. This creates an immediate link to the *inverse optimum* literature. This literature uses the actual tax schedules of countries to estimate the governments' implied social welfare weights – how much they value an additional dollar across the income distribution – assuming that the government has optimally set tax rates according to these social welfare weights and the assumptions within the Mirrlees model. In other words, the method calculates – under the Mirrlees assumptions – the value that governments' tax schedules imply for an additional dollar at every income percentile. In the classic literature [Lockwood and Weinzierl, 2016, Hendren, 2020], this value is decreasing across the distribution (except the very top), indicating that governments generally value additional income at the bottom more than additional income at the top.

Importantly, however, this method assumes that the government has not considered inequality as a negative in itself when designing the tax schedule. This means that there is no inequality externality implicit in the government's tax design priorities. If the government took a negative inequality externality into account when designing the tax schedule – due to a concern that high inequality could lead to social unrest, for example – there is a redistributive incentive independent of any potentially decreasing social welfare weights.

For our purposes, we are interested in what potential inequality externalities and social welfare weights the real-world U.S. tax schedule could accommodate. The externality and the welfare weights are both free variables, which means that we must assume one to estimate the other. We thus assume that the government has considered inequality as a negative externality of various strengths to estimate the resulting social welfare weighs. Through this exercise we

find that even a small negative inequality externality leads to social welfare weights that *increase* with income.

This leads to our second main finding; under our assumptions and optimal design, the U.S. government cannot have designed the income tax system with *both* a higher value for income among the poor *and* a significant concern for inequality's consequences. One or the other is possible, but not both. This questions to what extent the U.S. government – and other governments, given that the same result would likely be true for most developed countries' tax schedules – in practice (i) considers inequality as an externality, (ii) prioritizes "pure" economic redistribution, or (iii) sets tax schedules optimally. Overall, the finding reaffirms the point that absent strong opposing forces (e.g. migration responses), optimal tax schedules are likely much more redistributive than we currently see in practice.

Beyond the real-world implications, the inverse optimum literature is one example of an academic sub-field where conclusions change drastically if economic inequality is an externality. In the paper, we discuss how this is likely true for many other academic works as well. As examples, Thurow [1971] shows that the famous First Welfare Theorem no longer holds if the income distribution is a pure public good (if inequality is an externality, in effect), and Støstad [2019] shows that the aforementioned [Atkinson and Stiglitz, 1976] theorem no longer holds if inequality is an externality. In general, the overarching comment of Chapter One is that models in the welfarist tradition – and thus much of the economic literature – have assumed that economic inequality has no significant effects on society.

The broader implications are naturally intriguing, which brings us back to the equity-efficiency framework. When equality is itself efficient, optimal public policy changes. Looking beyond income taxation, what Chapter One emphasizes is a powerful and efficiency-based motive for inequality reduction.

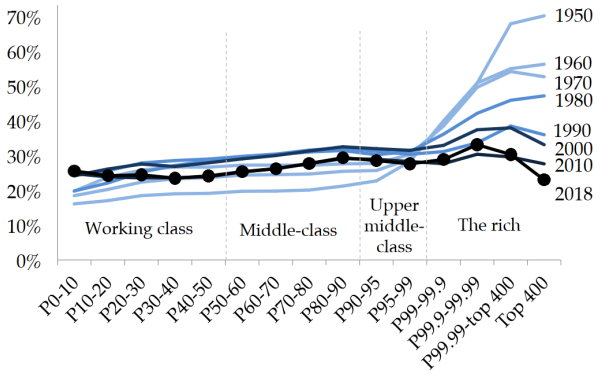
Before moving to Chapters Two and Three, I will briefly discuss how a reduction of economic inequality could be achieved in actual policy settings.

On Actual Methods to Reduce Economic Inequality The policy investigated in Chapter One is income taxation, and many assumptions are made in the model framework. As a consequence, the mathematical solution should not be interpreted as an unambiguous policy proposal. Indeed, there are good reasons for why one might not want a 90% income tax rate. These are largely related to what the model is too simplistic to include; migration issues, that people prefer money they earn over money given by the government, that we are not truly utilitarian, that different forms of taxation can be used as substitutes, tax avoidance and evasion, and so on.

For realistic inequality-reducing policy making, a mixture of approaches is necessary. *Inequality: What Can Be Done* by Tony Atkinson [Atkinson, 2014] sets out one such agenda with fifteen policy suggestions designed to reduce inequalities. Atkinson's main point is that we possess the necessary knowledge and strategies to reduce inequality, and that the real issue lies in our resolve to put these ideas into action. Among the policy suggestions are guaranteed public jobs at minimum wage, a universal capital endowment for all adults, and the creation of sovereign wealth funds. Progressive taxation policies also feature prominently, with suggestions for levying taxes on inheritance, gifts, and property, based on current assessments.

In *the Triump of Inequality*, Emmanuel Saez and Gabriel Zucman summarize the evolution of the U.S. tax system over time and note that it has become increasingly regressive (Figure 3, Saez and Zucman [2019]). They suggest that this is due to a collapse in capital taxation, a proliferation of tax avoidance and tax evasion, and globalization's effects on tax competition. Three main approaches are suggested to reduce the regressivity of the system. First, a plan to stop corporate tax evasion through both domestic reform and a global corporate minimum tax.[12] Second, a top-end wealth tax levied above $1 billion (which, as an aside, could be very intuitively justified by a wealth inequality externality). Third, and most related to Chapter One, they suggest a reinvention of the income tax through taxing national income. In a later work, the authors also suggest a 0.2% wealth tax on corporations headquartered in G20 countries based on stock market share prices [Saez and Zucman, 2022].

Figure 3: Tax Rates by Income Group



Note: Average tax rates by income group as a percentage of pre-tax income in the United States, from Saez and Zucman [2019].

These two sets of proposals are only a small subset of potentially inequality-reducing policies. Prominent economists have also argued for inequality reduction in the form of a higher minimum wage [Dube, 2019], increased early childhood education investments [Heckman, 2011], an inheritance tax [Piketty et al., 2014], increased refundable tax credits that promote work (e.g. the EITC) [Hoynes et al., 2017], and more. Global coordination might be necessary to implement the tax-side reforms; while the OECD recently instituted a relatively low global minimum corporate tax [OECD, 2023], this is also a relatively unexplored avenue to minimize concerns about tax migration, evasion, and avoidance.

The purpose of the above discussion is to show that well-known and rigorous proposals for reducing economic inequality do exist. Still, most of them have not been put in place. This leads to the salient question of why this is so. Do individuals want their governments to redistribute more, or are they relatively happy with the situation as it is? And if they do want to redistribute, is it purely for the traditional equity reasons – or are they also worried about the consequences of inequality? I study these questions in Chapters Two and Three.

Preferences for Redistribution

The traditional starting point of the literature on preferences for redistribution is the theoretical Meltzer and Richard [1981]. This paper uses the median-voter theorem to argue that self-interested individuals' preferences for higher incomes sets the size of government, as there is a trade-off between redistribution (benefiting the low-income agents) and lower taxes (benefiting high-income agents). The median voter's self-interested preference determines the amount of redistribution.

Other early theoretical papers discussed other reasons to like or dislike inequality from an

[12] A smaller version of a global minimum corporate tax rate was recently proposed by the OECD and ratified by 136 countries.

individual point of view. Individuals could reject inequalities from a motivation to insure themselves against future income shocks [Harsanyi, 1955], for example. Or they could reject inequalities because they see themselves one day benefiting from the high top incomes at the top of the distribution; this is the essence of Hirschman and Rothschild [1973]'s so-called tunnel effect. A related idea was developed in Bénabou and Ok [2001], which argues that high inequality may be positively viewed because it signals a future increase in income.

These papers begin with the starting point that our preferences for redistribution are largely set by self-interested income motives. This idea has since been strongly criticized. Two seminal early additions are Fehr and Schmidt [1999] and Bolton and Ockenfels [2000]. Both papers are experimental and show that individuals have preferences over other individuals' incomes. In short, in laboratory games many people are willing to give up some of their own income to change other people's incomes. These motives could for example be about inequality aversion, as in Fehr and Schmidt [1999], or relative income concerns, as in Bolton and Ockenfels [2000].

These papers were early works in what is now a large empirical literature on individuals' preferences for redistribution. Self-interest is still a strong determinant, echoing Meltzer and Richard [1981] (see Cruces et al. [2013], Durante et al. [2014], for example). But there are also other important factors. Perhaps the most impactful of these are individuals' fairness views, as indicated by Fehr and Schmidt [1999]. An early contribution to our understanding of the topic was Cappelen et al. [2007]. The authors point out that there is a pluralism of fairness ideals given the same (known) allocation of resources; individuals could be strict egalitarians, liberal egalitarians, or libertarians. In effect, even if we all agreed on who produced and received what, we could still disagree on the ideal *fair* income distribution. In other words, philosophical ideals are different. Broad fairness views are thus a combination of two factors; first our understanding of the production process, and second our philosophical ideals. Various aspects of broad and narrow fairness views have been extensively explored by among others Tyran and Sausgruber [2006], Cappelen et al. [2013], Durante et al. [2014], Almås et al. [2020], Epper et al. [2020], Cappelen et al. [2020] and Almås et al. [2022].

There are several methods to study these questions. Most works use surveys to gather at least part of the relevant information. Large-scale international surveys, for example the Gallup World Poll or the World Values Survey, are also often used as supplementary evidence; these surveys generally ask questions on individuals' preferences for redistribution that offer both a time and a cross-country dimension. This allows for broad correlational studies. However, causality usually remains unclear, and academic papers usually employ their own specifically-designed surveys to supplement other methods. Many of the papers listed above include laboratory experiments where participants are offered small sums of real money, for example.

Another well-known method is information experiments. These experiments provide individuals with information through for example a short video to see how this information affects their preferences for redistribution. This is a powerful method as it allows researchers to assign *causality* to information provision. Some examples are Kuziemko et al. [2015], who find that preferences for redistribution are relatively inelastic to information about the income distribution, and Stantcheva [2021] who finds that explaining the workings and consequences of inequality-reducing tax policies influences individuals' willingness to support such policies. Both

papers also emphasize the importance of general fairness views as well as trust in the government. There is also a large literature informing individuals of their position in the economic distribution and exploring whether this changes their preferences for redistribution with mixed results [e.g. Cruces et al., 2013, Hvidberg et al., 2022].

These types of works have increased with the proliferation of online survey, and many more determinants have been suggested. It has for instance been explored whether redistributive preferences are influenced by immigration beliefs [Alesina et al., 2018a], social mobility beliefs [Alesina et al., 2018b], experienced inequality [Roth and Wohlfart, 2018], over-confidence [Buser et al., 2020], individual risk preference [Gärtner et al., 2017], and more.

It is this large literature we wish to add to in Chapter Two. The main goal of Chapter Two is to introduce *inequality externality beliefs* as a potentially important determinant for individuals' preferences for redistribution. Despite the large literature referenced above, there was very little known about these beliefs before we launched the project. There was essentially no information in existing large-scale surveys about how individuals believe economic inequality affects the amount of crime, social unrest, political polarization, or generalized trust, for example.

We believed this information would be useful for two main reasons. First, individuals' preferences for redistribution could be partly determined by these beliefs. This could have ramifications for voting choices and public policy, and explain part of the cross-country variation in inequality reduction. Second, the information would give an indication for how relevant it is to consider inequality as an externality in policy modeling exercises. Although individuals' beliefs are questionable evidence for whether economic inequality truly is an externality, such beliefs can be a barometer for whether the assumption that inequality is not an externality is widely accepted among the populace. If individuals strongly believe that inequality is a negative externality, for example – as we find – it seems more reasonable to argue that public policy and economic models should account for this possibility.

Chapter Two, co-written with Max Lobeck, studies these topics. We first explore individuals' beliefs about the consequences of economic inequality, then link these beliefs to their preferences for redistribution. This is a largely empirical paper which builds on the same themes as Chapter One.

Chapter Two: The Consequences of Inequality: Beliefs and Redistributive Preferences

Chapter Two is based on two specifically designed surveys in the United States with $4,371$ and $2,360$ respondents respectively. The respondents, who approximate a representative sample of the U.S. population, were found through the professional survey companies *Lucid* and *Dynata*. We explore U.S. citizens' beliefs in the consequences of inequality and the connection of these beliefs to redistributive preferences.

Our first major finding is that the majority of U.S. citizens think economic inequality has significant negative consequences. In general, almost all people agree that inequality impacts society in one way or another, with about 60% of respondents asserting that economic disparity causes detrimental effects on society as a whole. Notably, there are profound convictions regarding particular aspects; for instance, 76% of the respondents believe that heightened economic inequality escalates crime rates, while 68% are of the opinion that it erodes the overall trust

within society.

Similar conclusions hold across incomes and party affiliation. Economic status does not correlate with inequality externality beliefs, and Democrats, Independents, and Republicans all strongly believe in the negative consequences of inequality. This is true for every outcome we elicit; even Republican are more likely to believe that more economic inequality *decreases* rather than *increases* the amount of economic growth and innovation, for example. This contrasts with broad fairness views, which are more polarized across incomes and party affiliation in our sample.

Using an information experiment with five different short videos, we link these beliefs to the individuals' preferences for redistribution. Each respondent is assigned to a different survey group, and most are shown a video regarding either (i) inequality's externality properties (three different groups are shown differing videos on inequality's correlations with trust and crime), (ii) the evolution of the wage-productivity gap and the top 1% income share (a fairness video as a point of comparison), or (iii) neutral information about inequality metrics as a control group (combined with another control group that saw no video). By comparing each group's willingness to redistribute later in the survey, we can measure how information on these different topics causally affect preferences for redistribution.

We find that individuals who watched the most comprehensive inequality externality video and the fairness video both increase their preferences for redistribution. The former of these points yields an intuitive but crucial finding; individuals' beliefs about the consequences of inequality causally affect their preferences for redistribution. In robustly establishing this fact, we add empirical evidence to the hypothesized efficiency-based reason to redistribute discussed in Chapter One. We also introduce this determinant to the empirical literature on redistributive preferences.

We estimate that the size of this effect is relatively large. Using three distinct methods, we find an effect on the same order of magnitude (but somewhat smaller) as that from broad fairness views. This indicates that inequality externality beliefs might play a large role in the public debate on redistribution – and thus on the amount of redistribution itself. Indeed, differential beliefs about these consequences of inequality could partly determine how we design our societies; after all, these beliefs determine what we believe to be *efficient*. Differences in these beliefs, then, could have caused cross-country differences in public policy and inequality reduction. This was indicated by the many quotes from public figures in the earlier sections, but our work presents the first academic evidence on the topic.

We also find suggestive evidence that these beliefs are structurally different from standard fairness views. Some of this is described above; in descriptive questions, fairness views are generally more polarized across political affiliation and income. Further, and importantly for Chapter Three, the fairness treatment arm from the information experiment lead to respondents self-reporting significantly more anger than the externality-based treatment arms. We also find that the fairness-based treatment arm has a more polarized effect across incomes than the externality-based treatment arms.

These structural differences are intriguing. Suppose fairness-based arguments for redistribution really are generally more polarizing than inequality externality-based arguments, leading

to more anger and splitting the population into partisan or income-based factions. If so, the extent to which the redistributive conversation focuses on fairness or inequality's consequences could explain differences in polarization across countries. But there is reason to be cautious; the findings from Chapter Two that motivate this idea are generally dependent on survey and video design choices. The different party and income polarizations in descriptive questions could be due to the questions we pose, and the differences across treatment arms could be due to video design choices. It is thus difficult to draw broader implications without further exploration. In Chapter Three, again co-written with Max Lobeck, these ideas are further explored.

On External Validity

Chapter Three focuses on the potential structural differences between inequality externality-based arguments and fairness-based arguments. To summarize, the indicative evidence from Chapter Two on this topic suggests that (i) fairness-based arguments for redistribution cause more anger than inequality externality-based arguments, (ii), fairness-based arguments for redistribution are more polarized across political affiliation than inequality externality-based arguments, and (iii) fairness-based arguments for redistribution are more polarized across economic status than inequality externality-based arguments.

Studying these questions rigorously is not trivial. This is due to the same problem that made the evidence from Chapter Two only indicative; survey design choices could affect results. Internal validity does not guarantee external validity.

While information experiments are generally useful at showing causality, they are also problematic in drawing broader conclusions. As an example, Chapter Two found that learning about the consequences of inequality can increase preferences for redistribution (and thus that inequality externality beliefs causally affect redistributive preferences). As this is true for the specifically chosen information we provided, the general principle must be true; it is a proof by example. The problem arises when the researcher wishes to say something general about a *class* of statements. For example, the fairness treatment arm in Chapter Two was stronger than the inequality externality treatment arms, but this does not necessarily mean that fairness-based information will affect redistributive preferences more than inequality externality-based information on average. Such a statement would require universal proof and a different method of approach.

In order to study these question in a more robust manner, then, we thus needed a different method than standard information experiments or laboratory games. This is why we created a new three-step experimental method for Chapter Three, which we call "*A Universe of Arguments*".

Chapter Three: A Universe of Arguments

Chapter Three introduces what we call the "Universe of Arguments" methodology, which is designed to extract empirical information about classes of statements. The idea centers on collecting an unbiased sample of statements (arguments) from the desired statement (arguments) class, then eliciting evaluations on this sample from other individuals. We believe this is a useful method to inch closer towards external validity in survey settings. The method is designed around three surveys; one to collect arguments, one to quality check the arguments, and one to

evaluate the arguments.

We employ this method to arguments about redistribution based on either fairness ideas or inequality externality ideas. We collect a total of 160 arguments in favor of redistribution, 80 of each type, and elicit a total of 32,300 argument evaluations. The three surveys include a total of 4,523 individuals.

The work strengthens some findings from Chapter Two and questions others. The main finding is that fairness-based redistributive arguments fuel more anger in respondents than inequality externality-based arguments. Due to the methodology, this finding does not rely on any arbitrarily chosen arguments from the researcher side and, we believe, is generally robust across the "universe" of such arguments used in public discourse. We also find that this anger stems from individuals who *agree* with the argument in question, and is largely due to the more normative nature of fairness-based arguments.

We also find indications that the support for externality-based arguments is less polarized across incomes. This could have significant real-world ramifications, as recent academic works indicate that high-income individuals have more political influence [e.g. Gilens, 2012, Mathisen et al., 2021]. We hypothesize that a differential focus on inequality's externality properties could have lead to different redistributive regimes across countries.

At the same time, indicative findings from Chapter Two are also questioned. Specifically, while Democrats are more likely to be convinced by pro-redistributive arguments than Republicans, there is no difference in this gap depending on whether the argument is fairness-based or externality-based. In other words, unlike the descriptive evidence from Chapter Two, the party polarization between Democrats and Republicans is equal for fairness-based and externality-based arguments.

Overall, Chapter Three adds nuance and external validity to the topic of the redistributive effects of fairness views and inequality's consequences.

As a whole, the contribution of this dissertation is to introduce the concept of *inequality as an externality* and to further our theoretical and empirical understanding of this efficiency-based reason to redistribute. I now proceed to the three Chapters.

Chapter 1

Inequality as an Externality: Consequences for Tax Design

Inequality as an Externality: Consequences for Tax Design*

Morten Nyborg Støstad† and Frank Cowell‡

Abstract

Economic inequality may change a wide range of societal outcomes, for example crime rates, economic growth, and political polarization. In this paper we discuss how to model such effects in welfarist frameworks. Our main suggestion is to treat economic inequality itself as an externality, which has wide-ranging implications for classical economic theory. We show this through the Mirrlees [1971] optimal non-linear income taxation model, where we focus on a post-tax income inequality externality. Top tax rates are particularly affected by the externality; in our main specification the optimal top marginal tax rate increases from 63% to 81%. Our model also provides a theoretical basis for real-world governmental tax choices that are irrational under standard optimal taxation methods. Finally, we find that the total inequality aversion implied by the current U.S. tax system is insufficient to accommodate both social welfare weights that are decreasing in income and a significant concern for inequality's externality effects. *JEL* Codes: H21, H23, D62, D63

*We thank Stéphane Gauthier, Marc Fleurbaey, Emmanuel Saez, Daniel Waldenström, Olof Johansson-Stenman, Fredrik Carlsson, and Karine Nyborg for helpful comments and discussions. We have also benefited from suggestions from Etienne Lehmann, Marie Young Brun, Max Lobeck, Elif Cansu Akoğuz, Stefanie Stantcheva, Thomas Blanchet, Antoine Bozio, François Fontaine, Damián Vergara, Eddy Zanoutene, Thomas Piketty, and seminar participants at the Paris School of Economics, UC Berkeley, the University of Oslo, the GT Économie de la Fiscalité, ECINEQ 2019, the 2021 EEA Congress, LAGV 2021, the 2021 IIPF Annual Congress, and the 2021 NTA Annual Conference on Taxation. Finally, we are deeply grateful to the Journal of Public Economics' Editor Nathaniel Hendren and two anonymous referees for invaluable comments and suggestions. Version: November 30, 2023.

†Paris School of Economics, 48 Boulevard Jourdan, 75014 Paris, France. Phone: +33766142152. Email: morten.stostad@psemail.eu (corresponding author).

‡London School of Economics, Houghton Street, London, W2CA 2AE, UK. Email: f.cowell@lse.ac.uk.

1. Introduction

In the last sixty years, economic modeling has regularly used individualistic utility functions and social welfare functions to evaluate policy options. The real-world influence of these models has been considerable; as such, how they treat economic inequality is also of great significance. There are several well-formulated reasons to prevent economic differences in the standard framework, which we will return to shortly, but a crucial factor has also remained neglected; the consequences of economic inequality on society and thus individuals' well-being. Suppose, for example, that higher income or wealth inequality causally changes the crime rate, the amount of social unrest, or the political polarization in a society. If so, even purely self-interested individuals are affected by the economic differences between people – regardless of whether their individual incomes change. Given that virtually all market activities affect the extent of economic inequality, it follows that economic inequality itself could be an externality. This paper explores the consequences of this idea.

The analysis we present can be divided into two primary components. The first component is on the overarching theme of the paper; the concept of economic inequality as an externality. This concept is explored in a general fashion. We discuss why an economic inequality term in the utility function is the most appropriate way to model the effects of inequality on society, following both Thurow [1971] and Alesina and Giuliano [2011], and why such a term cannot be mathematically approximated by appropriate social welfare functions (SWF) or concave utility functions. As such, most models which preclude externalities also assume that economic inequality does not significantly change society. As this is a potentially large assumption, we discuss how weakening it and allowing for various inequality externalities affects both general economic intuition and optimal taxation frameworks. We allow for either positive or negative inequality externalities and establish potential micro-foundations, which are often simple. The externality can exist in the presence of fully self-interested, rational individuals.[1]

The second component of the paper focuses on the Mirrlees [1971] optimal non-linear income taxation model, where we calculate optimal marginal tax rates analytically and numerically in the presence of various types of inequality externalities. While we focus on a post-tax income inequality externality, we also introduce other types of inequality externalities into the model (pre-tax income, utility) and vary the inequality metric itself. To pin down plausible magnitudes of a real-world income inequality externality we utilize three distinct methods, the primary of which uses survey data from Carlsson et al. [2005] and all of which imply similar magnitude ranges. Finally, we perform an inverse-optimum exercise to examine how implied social welfare weights in the U.S. tax system change if the tax schedule design was influenced by an income inequality externality.

The principal insight of our paper is that the large majority of welfare-based economic frameworks implicitly assume that economic inequality has no meaningful effects on societal outcomes, and that softening this assumption changes model conclusions drastically. We explicitly show these changes in optimal income taxation (OIT), where both theoretical and simulation-based

[1] That self-interested individuals are affected by the externality is the main difference between our concept and other-regarding preferences. Such preferences are philosophically problematic for policy design as they are based on individuals' emotions [Harsanyi, 1977, Goodin, 1986].

findings are affected, and discuss which other frameworks could be similarly fragile. Within the context of OIT we find two main results. First, the presence of an inequality externality has a particularly pronounced impact on *top* optimal marginal tax rates. This is a theoretical finding that is borne through in our numerical simulations; in our main specification the optimal top marginal tax rate changes from 63% to 81% when introducing our median post-tax income inequality externality.[2] Second, our analysis reveals that the total inequality aversion in the current U.S. tax system is insufficient to accommodate both social welfare weights that are decreasing in income and a significant concern for inequality's externality effects. While the current tax system could be rationalized as prioritizing income transfers to lower-income individuals [Hendren, 2020], it cannot also contain a realistic concern for inequality's externality effects given the aggregate capacity of the tax schedule to mitigate inequality.

Before further discussing our results we will briefly explore what we know about how economic inequality affects various facets of society and individuals' lives. It is difficult to establish causality on the topic for several reasons, the first among them being the lack of exogenous variation in macroeconomic inequality.[3] There is no shortage of empirical papers on the subject, however, and there are overall strong indications that economic inequality acts as an externality in various ways. First, considerable experimental and microeconomic evidence has in recent years indicated that economic inequality between workers or experimental subjects impacts life satisfaction [Card et al., 2012], productivity [Breza et al., 2018], trust [Fehr et al., 2020a], and cooperation [Xu and Marandola, 2022]. Second, while correlation by no means implies causation, there are robust cross-country correlations between income inequality and various negative societal outcomes.[4] We show two such correlations for general trust and homicides in Figure 1. Third, both laypeople and experts often express the belief that inequality does change society; in the United States, the large majority of citizens believe that economic inequality negatively affects a wide range of societal outcomes [Lobeck and Støstad, 2023]. Similar concerns have been raised by prominent politicians, philosophers, and economists.[5] Laboratory experiments also indicate that a majority of individuals would forego part of their income to live in more macroeconomically equal societies [Carlsson et al., 2005, Fisman et al., 2021, Bergolo et al., 2022]. Fourth, it is trivial to create realistic microfoundations of various inequality externalities, which we show in Section 5. Other papers have given more attention to specific potential channels; Benabou [1996], Auclert and Rognlie [2018], Mian et al. [2020], and Jones [2022] are just a few examples.[6]

[2]The median inequality externality is estimated with the survey data from Carlsson et al. [2005], which asks respondents for their income-inequality trade-off in a hypothetical setting.

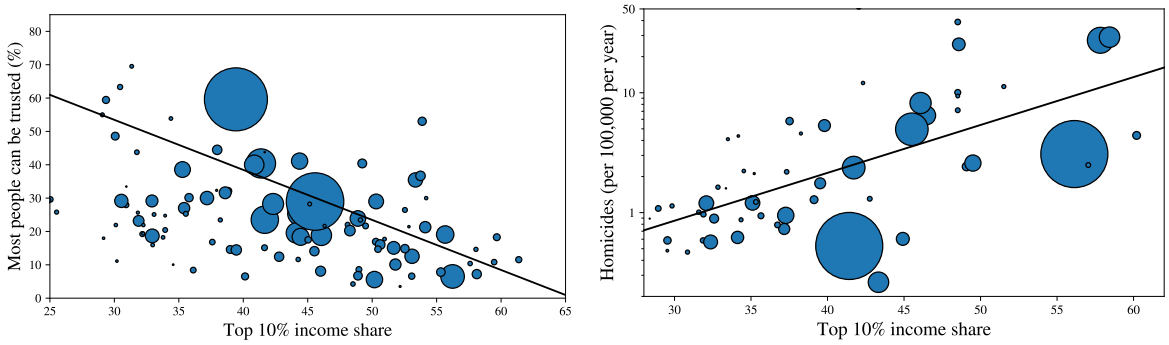
[3]Other concerns include measurement error and missing data on economic inequality, a generally low variability in inequality over time, reverse causality (where outcomes also affect inequality), non-linear effects of inequality on outcomes, the intertwined nature of inequality effects and poverty or individual income effects, the question of *which type* of inequality matters, and so on.

[4]The related literatures are too large to summarize here. Examples of relevant reviews can be found in Rufrancos et al. [2013] for crime and Bergh et al. [2016] for individual health.

[5]For example Plato (est. 360): "In a state which is desirous of being saved from the greatest of all plagues [...] here should exist among the citizens neither extreme poverty, nor, again, excess of wealth, for both are productive of both these evils," translated in Plato [2016]. More recent examples include Greenspan [2014]: "You can see the deteriorating impact of [inequality] on our current political system," or Obama [2011]: "This kind of inequality – a level that we haven't seen since the Great Depression – hurts us all."

[6]While most of the evidence presented in this paragraph indicates that economic inequality is a negative externality, we do not assume this in general. Jones [2022] discusses how top incomes could drive innovation, for

Figure 1: The Correlations of Inequality



Note: Left: The cross-country correlation of generalized trust (World Values Survey) and the top 10% income share (World Inequality Database). Right: The cross-country correlation of homicides (World Bank) and the top 10% pre-tax national income share (World Inequality Database). Both correlations are relatively unaffected by standard controls. Data point area is proportional to population. Note the logarithmic scaling of homicide rate. Calculations by the authors.

So let us assume such effects exist and are welfare-relevant. How would one consider their overall consequences in a welfarist framework? The most intuitive approach is simply to model every externality individually. It is difficult to imagine that such a model could remain tractable, however, and this strategy would also require an empirically problematic assessment of the actual importance of every externality channel. For general use, then, this route appears overall infeasible. Meanwhile, a standard inequality-averse or Rawlsian (maximin) social planner is generally insufficient to model inequality externalities based on income or wealth because the externality introduces a wedge between what is individually and societally optimal.[7]

Instead, a more appropriate solution is the inclusion of an inequality metric in the individual's utility function. Such an externality can exist even when individuals are fully rational and selfish, unlike models with traditional other-regarding preferences (ORP). As an example, imagine a perfectly self-interested individual in a society where income inequality increases crime through a Becker [1968]-type opportunity cost framework. Suppose income inequality and thus crime increases, and that the person's bike is stolen as a consequence. The individual experiences a negative shock and would undoubtedly, absent any other changes, prefer the prior (more equal) state of the world. Thus, if inequality leads to more crime, inequality should enter her utility function. A similar argument could be made with any other variable affected by inequality.

We are not the first to note that economic inequality's effects on society can imply an inequality term in the utility function. The idea was first developed by Thurow [1971], who shows that the First Welfare Theorem fails if the income distribution is a pure public good. Since then the idea has periodically resurfaced. Kaplow [2010], for example, mentions that the economic distribution could affect variables such as crime, which could imply optimal taxation effects. Alesina and Giuliano [2011] briefly consider how inequality's effects on society might affect consumption and thus utility, and Rueda and Stegmueller [2016] discuss how inequality can act as an

instance. One could imagine such concerns being more prevalent in societies that are more economically equal than we see today.

[7]Non-altruistic individuals will choose their own labor effort without taking into account how this effort impacts the global level of income inequality, for example; if income inequality affects societal factors, there is an externality dimension to this choice that is not well-modeled by simply discounting individual utility.

externality in the case of crime. We add to this literature by detailing the effect of an inequality externality in a specific, well-known economic model – the Mirrlees model – and by furthering the analysis presented in these works. To advance our understanding of the overarching idea, we (i) clarify the mathematical structure of the externality, classify its key components, and develop a sufficient statistic for the magnitude of the externality given an inequality metric, (ii) make a first approach at estimating the magnitude of a post-tax income inequality externality based on the Gini coefficient, (iii) formulate a set of theoretical consequences of income- and wealth-based inequality externalities that are relevant for the broader economic literature, and (iv) create micro-foundations for various ways in which economic inequality can change pertinent societal factors.

Our case study uses the Mirrlees [1971] model, where we introduce various types of inequality terms into the individual's utility function. As a widely used model describing OIT, the Mirrlees model represents an important pillar of public economics [Diamond and Mirrlees, 1971, Diamond, 1998, Saez, 2001]. The original Mirrlees model assumes no externalities, an assumption which has been examined by a wide range of papers. Though we will return to how our work contrasts with the existing literature, the briefest way to describe our technical contribution is that we are the first paper to explore the effect of an income inequality term in the individual utility function in the continuous Mirrlees model. The social planner is faced with a trade-off between maximizing tax revenue and setting the preferred level of post-tax income inequality, which expands on models from Oswald [1983], Kanbur and Tuomala [2013], and Aronsson and Johansson-Stenman [2020]. We explore several types of inequality externalities in this framework, focusing mainly on a post-tax income inequality externality, and solve the problem both analytically and numerically. Several assumptions are made for simplicity; there are no income effects in the base case, there is no extensive margin of labor supply, there is separability in income, work effort, and the inequality externality, and the only instrument available to the social planner is that of income taxation.

As mentioned, our OIT exercise yields two main results. First, top marginal tax rates are particularly sensitive to the inequality externality. The intuition follows from how a small marginal tax increase at a given tax bracket affects tax revenue and post-tax income inequality respectively. In general, the effects on post-tax income inequality – which are welfare-relevant in our case – are more heavily influenced by the distributional location of the tax bracket than the standard revenue effects. This can be seen through the framework detailed in Saez [2001], which discusses the consequences of a small tax increase as (i) the mechanical effect on tax revenue, and (ii) agents' behavioral responses. In the standard revenue-maximizing case, both effects always oppose each other. A tax increase leads to mechanically higher tax revenue, which is welfare-positive. Meanwhile, agents' behavioral shifts away from labor supply is distortionary and decreases tax revenue, which is welfare-negative. This creates the classical equity-efficiency trade-off. In contrast, the two effects can also harmonize in their impact on post-tax income inequality. The mechanical effect is similar to the revenue case, as gathering tax revenue from those above a given tax bracket to redistribute uniformly always decreases post-tax income inequality (except at the very bottom tax bracket). Agents' behavioral responses, however, increase or decrease post-tax income inequality depending on the location of the tax hike. At

the bottom, a behavioral shift away from work effort *in*creases income inequality. At the top, a behavioral shift away from work effort *de*creases income inequality. This creates a distributional asymmetry, where the behavioral responses' effect on income inequality (and thus on the optimal tax rate) is opposite at the top and bottom of the distribution. This means that, as pertains to post-tax income inequality, the optimal tax effects of the mechanical effect and agents' behavioral responses always oppose each other at the bottom of the distribution and harmonize at the top. Top marginal tax rates are thus particularly sensitive to the inequality externality.

In our numerical simulations, applying the median externality estimate results in the optimal top marginal tax rate increasing from 63% to 81%. Given standard parameter values and reasonable magnitudes of the externality we find a very wide range of possible optimal marginal top tax rates, ranging from negative (<0%) marginal top tax rates if inequality is a positive externality to extremely high (>90%) marginal top tax rates if inequality is a negative externality. This range of optimal top tax rates is wider than what is supported by standard parameter values in the no-externality case, where optimal top marginal tax rates usually range between 50% and 80%. This has arguably decreased the focus on the "equality dimension" in optimal top tax rate analysis – which we show can be highly relevant as long as inequality itself affects the individual.[8] The *individual* inequality concerns that arise from an inequality externality thus differ from the *social* inequality concerns modeled by an inequality-averse SWF. We naturally find optimal top tax rates above the revenue-maximizing Laffer rate, as direct equality effects imply that the social planner might trade off some revenue for changed equality levels, and our results also provide a theoretical basis for previously unsupported policy arguments – such as the high post-war top marginal tax rates in the United States and the United Kingdom if inequality is a negative externality.

Our second main result is related to this last point and comes from the inverse-optimum exercise popularized by Bourguignon and Spadaro [2012]. This method calculates the implied social welfare weights (SWWs) of real-world tax systems under the assumption that the tax schedule was set optimally. As shown in Lockwood and Weinzierl [2016] and Hendren [2020], SWWs from the U.S. tax schedule are generally decreasing in income in the no-externality case. We introduce an inequality externality into this framework. At the bottom of this exercise is a substitution effect between these two redistributive motives. If we suppose that the social planner considered inequality as an externality when designing the tax schedule, this must imply less progressive SWWs *given the same (actual) tax schedule*. This allows us to explore what motives the "total" inequality aversion in the U.S. tax schedule could contain. Through this exercise we find our second main result; the 2019 U.S. tax system is not sufficiently inequality averse to accomodate both SWWs that decrease in incomes and a realistic concern about the societal effects of economic inequality. While the tax system can accommodate either redistributive motive, it cannot accommodate both. If the U.S. social planner considered inequality as an externality to our median value, implied SWWs are sharply increasing in income – indicating that one dollar at the bottom of the distribution is worth five dollars at the top. We conclude that the current U.S. tax schedule is either not progressive in transfers or has an implied concern for inequality's externality effects that is significantly lower than both our empirical estimates

[8] This is not to say that the inequality externality is an equity concern in the standard equity and efficiency-framework, where it is clearly an efficiency concern.

and U.S. citizens' concerns as detailed by Lobeck and Støstad [2023].

We will now briefly outline how our work differs from the existing OIT literature. As our approach assumes separability between the externality and the remainder of the utility function for simplicity,[9] our framework is mathematically speaking a specific case of the models presented in Oswald [1983] and Kanbur and Tuomala [2013].Both these papers examine an average income-based externality. We note three main technical novelties as compared to the existing literature. First, we introduce a new and simple way to account for inequality terms in individual utility functions in optimal taxation frameworks. This is possible due to the family of inequality metrics we use, which simplifies an analytically intractable externality[10] into a linear combination of consumption externalities with varying marginal effects that depend on the income-rank of the individual.[11] As such, we can use much of the existing externality framework, including the aforementioned Oswald [1983] and Kanbur and Tuomala [2013], to evaluate what would otherwise be a challenging analytical problem. The second contribution to the literature is to explore the ramifications of a specific case of these models where we allow the marginal externality to depend on the location of the individual in the distribution (and thus also the individual's income). Although both Oswald [1983] and Kanbur and Tuomala [2013] mention this as a possibility, neither paper explicitly explores the issue. We focus on a small-perturbation framework to build intutition, unlike both these papers which use mechanism design frameworks (the modified version of which we also solve). Our analysis leads to novel insights relating to the effect of distributional externalities on optimal income taxation, and particularly on optimal *top* tax rates. Third, we solve the inverse-optimum problem [Bourguignon and Spadaro, 2012] in the presence of a global externality and illustrate the consequences for implied SWWs of the 2019 U.S. tax system. Global externalities are rarely discussed in this literature – we are only aware of Tsyvinski and Werquin [2017], which discusses the compensation principle in a general equilibrium-based framework and is thus both conceptually and mathematically different from our work. Given the large focus on inequality's effects on society in political rhetoric, we believe this is a particularly interesting exercise in our framework.

In general, our work adds to the already large literature on externalities in optimal taxation. This literature has been particularly developed for relative income concerns or ORP [Boskin and Sheshinski, 1978, Persson, 1995, Aronsson and Johansson-Stenman, 2008, 2015, 2018a,b, 2020] and environmental externalities [e.g. Sandmo, 1975, Bovenberg and van der Ploeg, 1994, Cremer et al., 1998]. Our analysis is particularly related to Aronsson and Johansson-Stenman [2020], which discusses various types of ORP including classical Fehr and Schmidt [1999]-type inequality aversion in a three-agent OIT model. We further this analysis by using a broader set of inequality-related specifications in a full continuous Mirrlees-type model. We also note Aronsson and Johansson-Stenman [2018a], which explores the first-best Pareto-efficient marginal tax structures when people are inequality averse in four different models. The potential for a

[9] This is a large assumption due to how it constrains how the externality magnitude relates to the marginal utilities of income and labor. The assumption of separability in externalities is weakened by among others Pirttilä and Tuomala [1997] and Jacobs and De Mooij [2015]. We also assume separability between utility from income and labor, again for simplicity. For more on the separability assumption see Gauthier and Laroque [2009].

[10] Typical inequality metrics often use absolute values and multiple integrals that depend on endogenous model variables.

[11] This is the same family used in Simula and Trannoy [2022], developed concurrently with this paper. The family itself is general and allows for various types of income inequality metrics.

direct focus on distributional concerns in the OIT model is also found in Kanbur et al. [1994] in terms of poverty concerns in the social welfare function, which contrasts to our continuous distributional metric inside the individual's utility function.

The paper is organized as follows. Section 2 examines the concept of inequality as an externality and how it differs from other ways in which distributional concerns are modeled in conventional OIT analysis. Section 3 incorporates an inequality externality in a standard OIT model and investigates the impact of the externality on optimal tax rates. Section 4 conducts numerical simulations in the OIT model. Section 5 discusses the inequality externality concept further, creating micro-foundations and discussing other potential mathemetical formulations. In total, Sections 2 and 5 discusses the concept of inequality as an externality while Sections 3 and 4 examine the OIT case. Section 6 concludes.

2. Inequality and Social Welfare: An Externality Approach

Suppose that economic inequality causally affects non-consumption goods individuals care about, the relevant of which we capture in an vector $\overrightarrow{\Psi_i}$. The most natural example of such goods are public goods (such as the amount of political polarization), but they might also be individual-specific (such as individual health) – see Section 5.1 for a further discussion on various channels. Suppose further that economic inequality can affect individual consumption x_i [Alesina and Giuliano, 2011],[12] and that individuals may have other-regarding preferences over economic inequality $\bar{\theta}$ [Cooper and Kagel, 2016].[13] The individual's utility can thus be written as,

$$U_i(x_i(\bar{\theta}), \bar{\theta}, \overrightarrow{\Psi_i}(\bar{\theta}), ...). \tag{1.1}$$

Detailed information on each component in the specification (1.1) is unlikely to be available; such complexity would also be unrealistically cumbersome for most models. We propose a simplification, noting that the separate contributions are less important than the overall impact of inequality in the utility function. The specification (1.1) could be written more compactly as the simplified form:

$$\tilde{U}_i(\tilde{x_i}, \bar{\theta}, ...) \tag{1.2}$$

where $\tilde{U}_i$ is the modified utility function, $\tilde{x_i}$ is the portion of consumption which is not determined by economic inequality,[14] and the term $\bar{\theta}$ represents the total impact of the inequality externality on the individual.[15]

The simplification from (1.1)→(1.2) does not rely on the existence of any of the three components we show in (1.1). The externality exists as soon as one of the three components enters

[12] Alesina and Giuliano [2011] discusses how income inequality could affect the income of individuals through three channels; externalities in education, crime and property rights, and incentive effects. One could also imagine that individual income is affected through some of the other channels we discuss in this work (political capture, innovation, social unrest, and so on).

[13] The overbar indicates a society-wide variable.

[14] This assumes this portion is separable. If not, one could equally write $\tilde{U}_i(x_i(\bar{\theta}), \bar{\theta}, ...)$ – the larger point is to simplify $\overrightarrow{\Psi_i}$.

[15] As this is a simplification, it may seem like an imperfect way to analyze implications of inequality's externality effects. We discuss this further in Appendix I.A.

the utility function and is deemed policy-relevant. For instance, individuals could be wholly self-serving and still have a utility function that is strongly dependent on economic inequality if economic inequality affects some pertinent public good. Given the many philosophical problems with introducing ORP and thus emotions into the welfarist framework – as discussed by Harsanyi [1977] and Goodin [1986], among others – this scenario may often be appropriate, and we focus on it for the remainder of the article. Before we continue, however, it is worth noting that as expressed in the form (1.2), the inequality externality as a whole is mathematically equivalent to an ORP term in the utility function. It follows that many of the results from the ORP literature can be applied to our framework. This immediately hints at the potential practical significance of the inequality externality, as ORP modifications often have large impacts on standard model conclusions [e.g. Oswald, 1983, Kanbur and Tuomala, 2013].

The concept also needs a well-defined inequality metric $\bar{\theta}$. We return to this later in the paper, but we note that the main type of inequality we will focus on is *income* inequality.[16] For simplicity we avoid other concerns that, while nonetheless important, complicate a first approach to an inequality externality. These issues include questions related to perceived inequality, inequality in different regions, (non-)meritocratic inequality, and so on.

We also note that the inequality externality could be heterogeneous. Various inequality externality channels could affect people in different ways, perhaps depending on their individual income or their position in the income distribution. We will return to this in Section 3 and 4.

We will now make a short detour to discuss how the inequality externality fits into the general utilitarian framework. In such models the social planner maximizes a social welfare function consisting of some weighted sum of every individual's utility. In addition to the inequality externality, there are thus two other channels through which inequality-related concerns can enter into the formulation of social welfare comparisons. These are (i) the cumulative effect of diminishing marginal utilities of income (DMUI), and (ii) social welfare weights (SWWs). We summarize this framework in Table 1.

Table 1: The Three Welfarist Consequences of Inequality

	Diminishing marginal utility of income	Social welfare weights	Inequality externality
Formulation	$\int_i g_i U_i(\underbrace{x_i}, \bar{\theta}, ...)di$	$\int_i \underbrace{g_i} U_i(x_i, \bar{\theta}, ...)di$	$\int_i g_i U_i(x_i, \underbrace{\bar{\theta}}, ...)di$
Causes	The decreased value of a dollar with increased income	Societal considerations of fairness, philosophical concerns	The societal effects of inequality, other-regarding preferences

Note: The three channels through which inequality could influence welfarist modeling. For each channel the key expression is highlighted by an underbrace. Individual consumption is denoted by x_i, resource inequality is denoted by $\bar{\theta}$, and the utility-based SWW is denoted by g_i.

We posit that the inequality externality is mathematically and intuitively distinct from these

[16] It could also be intriguing to consider $\bar{\theta}$ as *wealth* inequality (or some combination of the two). This would be a particularly insightful approach in optimal wealth taxation models, where a key practical motivation for additional taxation is arguably a concern for the societal effects of high wealth inequality.

other two channels; except for special cases,[17] an inequality externality cannot be mathematically captured by the other formulations.[18] The intuition is trivial: as with any other externality, an inequality externality introduces a gap between the socially and individually optimal decisions. The sub-optimality of individual decisions cannot be approximated by suitable SWWs, as discounting *utility* is dissimilar from discounting *income*,[19] and also cannot be approximated by modifications to an individualistic utility function as such modifications would have to depend on other agents' incomes. We further discuss why standard methods are insufficient to model an inequality externality in Appendix I.B.

We will now explore the effect of introducing three types of inequality externalities – pre-tax income, post-tax income, and utility – into the Mirrlees [1971] framework.

3. Optimal Income Taxation: Theory

We consider the second-best solution for a non-linear optimal income taxation schedule with a continuum of individuals in the presence of an inequality externality. The continuum of agents is indexed by i and normalized to one. Individual i derives utility from consumption $x_i \geq 0$, incurs disutility from work effort $l_i \geq 0$, and is affected by the society-wide post-tax income inequality $\bar{\theta} > 0$. To keep the relevant intuition simple, we assume separability between consumption, work effort, and the inequality externality throughout. As such, individuals' work decisions are independent on the level of income inequality which they thus do not need to know or estimate.

In the main text we will discuss a simplified version of the problem in the small-perturbations framework [Saez, 2001]. We assume quasi-linearity in consumption and in the inequality externality. This implies that there are no behavioral responses to average income changes (often discussed as "no income effects" in the literature). The resulting utility function is,

$$U_i = \mathcal{U}\left(x_i - v(z_i) - \eta_i \bar{\theta}\right),$$

where the function v is identical for all individuals and we have replaced work effort l_i with pre-tax income $z_i \geq 0$,which the individual incurs disutility from earning. This z is distributed with a strictly positive density $h(z)$ and cumulative density $H(z)$ across the population. v is increasing and strictly convex in z_i and identical for all individuals. The social planner has SWWs g_i indicating the benefit of one more unit of income to individual i; for the remainder of the paper we will assume that $\mathcal{U}$ is taken into account by the social planner's social welfare function. In Appendix I.C we use a mechanism design framework to solve the problem with a more general utility function.[20]

The inequality externality is formalized as an inequality term $\bar{\theta}$ of individual-specific magnitude η_i, where this magnitude represents the marginal rate of substitution between post-tax

[17] We discuss this further in Section I.B.

[18] This differs from how DMUI *can* be approximated by appropriate SWWs as long as the remaining (individualistic) utility function is appropriately modified to keep the individual's work choice unaffected.

[19] We discuss income-based SWWs [e.g. Saez and Stantcheva, 2016] in Section I.B.

[20] The utility function in Appendix I.C allows for behavioral income effects and is,

$$U_i = \tilde{\mathcal{U}}\left(u(x_i) - V(l_i) - \Gamma_i(\bar{\theta})\right),$$

where U_i is the cardinal utility of individual i as viewed by the social planner, post-tax income inequality is represented by $\bar{\theta}$ in an individual-specific function Γ_i, and the functions $\tilde{\mathcal{U}}$, u, and V are identical across all individuals.

income inequality and individual income $\eta_i = MRS_{x_i\bar{\theta}} = -\frac{dU_i/d\bar{\theta}}{dU_i/dx_i}$. In other words, η_i measures how much consumption the individual would give up for or pay for one unit decrease in the relevant inequality metric. We allow for potentially heterogeneous inequality externalities; as we will later show, the net inequality externality effect for a given inequality metric is $\eta = \int_j \eta_j g_j dj$. The main discussion will be for a post-tax income (consumption) inequality externality, with extensions for pre-tax income inequality and utility inequality in Section 4.6.

We also need to choose the inequality metric $\bar{\theta}$. Inequality metrics are often analytically difficult, and to simplify the problem we use a particular family of absolute inequality metrics discussed in Cowell [2000]. For post-tax income inequality, which will be used in the main specification, this family has the form,

$$\bar{\theta}\left(\boldsymbol{z}, H\right) = \int_{\underline{x}}^{\bar{x}} \kappa(z) x(z) dH(z), \tag{1.3}$$

where $\kappa(z)$ is the weight of the agent in the inequality metric. This weight is crucially only dependent on the *rank* of the individual in the distribution. We have used the rank-invariance between pre-tax income z and post-tax income x to specify the weight in terms of z. As x is endogeneous to the tax system and thus difficult to deal with, this key mathematical trick simplifies the problem.[21] We propose that these rank-dependent inequality metrics could represent an important simplification in similar problems.

The inequality weight $\kappa(z)$ is non-decreasing, continuous, positive near the top of the income distribution and negative near the bottom, and otherwise general. For example, the (absolute) Gini coefficient in post-tax income has a weight $\kappa_G(z) = 2H(z) - 1$. In the numerical simulations we will also explore other post-tax income inequality metrics based on other types of rank-specific weights $\kappa(z)$ where $\int_{\underline{z}}^{\bar{z}} \kappa(z) dH(z) = 0$, such as those in the Lorenz [Aaberge, 2000] or S-Gini families [Donaldson and Weymark, 1980]. Absolute inequality metrics are used to keep scale invariance.[22]

It is worth mentioning that the true inequality metric for measuring the inequality externality accurately is likely to be a function of several different inequality metrics. To show an example of this, suppose that inequality's effect on crime is dependent on relative poverty and that inequality's effect on political capture is dependent on the proliferation of top incomes. Both relative poverty $\bar{\theta}_p$ and top income proliferation $\bar{\theta}_t$ are distributional metrics, which we represent in our framework by the distributional weights κ_p and κ_t for their respective inequality measurements $\bar{\theta}_p$ and $\bar{\theta}_t$. Take then an example with separability and homogeneity in these externality effects, such as in the simple example of $U = x - \eta_p\bar{\theta}_p - \eta_t\bar{\theta}_t$ where η_p and η_t indicate externality magnitudes. The total externality effect is $-\eta_p\bar{\theta}_p - \eta_t\bar{\theta}_t = -\left(\eta_p + \eta_t\right)\int_{\underline{z}}^{\bar{z}} \left(\frac{\eta_p}{\eta_p+\eta_t}\kappa_p(z) + \frac{\eta_t}{\eta_p+\eta_t}\kappa_t(z)\right) x(z) dH(z)$. The modified inequality metric is thus $\bar{\theta}_{true} = \int_{\underline{z}}^{\bar{z}} \left(\frac{\eta_p}{\eta_p+\eta_t}\kappa_p(z) + \frac{\eta_t}{\eta_p+\eta_t}\kappa_t(z)\right) x(z) dH(z)$, a weighted sum of the two inequality metrics, with an externality magnitude of $\eta_{true} = \eta_p + \eta_t$. As such, the inequality metrics and externality magnitudes we use could be seen as a combination of potentially several externality-

[21] A similar trick is also crucial in the more general mechanism design approach in Appendix I.C, where we use the rank-invariance of x_i and wage-earning ability n_i.

[22] Absolute inequality metrics are equal to the standard inequality metrics multiplied by the average income. We use these metrics to keep scale independence in the additive utility function.

determining inequality metrics.

Combined with the potentially heterogeneous inequality magnitudes, this allows for both heterogeneous inequality metrics, heterogeneous inequality magnitudes, or a combination of both. Under a combination of both, the net inequality externality is $\eta_{net}\bar{\theta}_{net} = \int_j g_j \sum_t \left(\eta_{jt}\bar{\theta}_{jt}\right) dj$ where t indicates the type of externality and j indicates the individual. This allows for individual-specific inequality metrics and externality magnitudes for a flexible number of inequality externalities.

The social planner sets an income tax $T(z)$ dependent on pre-tax incomes z such that $x_i = z_i - T(z_i)$. This is done through finding the tax schedule $T(z)$ from which no given small perturbation ϵ which changes the tax schedule as $T(z) + \epsilon\Delta T(z)$ leads to welfare improvements. We denote the resulting change in the inequality metric from the small tax increase by $\Delta\bar{\theta}$. The local optimal tax criterion is thus defined as the tax schedule $T(z)$ for which any small budget neutral tax reform in direction $\Delta T(z)$ has $\int_i g_i \left[\Delta T(z_i) + \eta_i \Delta\bar{\theta}\right] di = 0$, where g_i is the SWW of individual i.

We will find the optimal tax system by calculating the revenue changes, individual income changes, and inequality externality effects of such a small reform, then assuming that all these effects equal to zero at the optimum. Although first-order conditions are only *necessary* criteria for the tax system to be optimal, we assume here that they are also *sufficient*; in every numerical simulation we check that the second-order conditions also hold. Before continuing we also note that our model neglects many factors that are important in setting true optimal tax rates; migration responses [Lehmann et al., 2014], the extensive margin of labor supply [Jacquet et al., 2013], and rent-seeking [Piketty et al., 2014], among others.

3.1. Optimal marginal tax schedules

The small perturbation method is discussed in Saez [2001] and is based on calculating the welfare effects of a small tax increase around the social optimum. This allows each effect of a small tax increase – and thus the optimal tax formula – to be numerically pinned down in terms of the relevant sufficient statistics, usually presented as various elasticities.

We show the full calculation with a post-tax income inequality externality in Appendix I.D. Before discussing the full formula we will discuss the intuition of the solution, following Saez [2001] and diverging when necessary.

Consider an infinitesimally small marginal tax rate increase $\partial\tau(z)$ for individuals in a small band of income between z and $z + dz$ that leaves marginal tax rates unchanged at all other income levels. We first note that there are two channels through which a tax increase affects incomes and thus social welfare. The first are agents' behavioral responses, the second is the mechanical revenue effect. The behavioral responses capture how a small tax increase leads each agent located at that tax bracket shift their work decision towards leisure (recall that there are no behavioral income effects outside of Appendix I.C). The mechanical effect captures how every agent above the tax bracket is taxed more in the absence of any behavioral response. These two channels both affect both revenue and incomes, as in the standard case, and post-tax income inequality.

There are five welfare-pertinent effects of such a change. Three of these are well-known from the previous literature and discussed in Saez [2001]. These are (i) the mechanical effect of higher

tax revenue, dM (ii) the behavioral responses of agents reducing their work effort, dB, (iii) and the welfare-relevant income losses of those who are taxed more, dW. There are also two new two equality consequences; (iv) the inequality impact of the mechanical effect, dI_M, and (v) the inequality impact of the behavioral responses, dI_B. At the optimum, the sum of the welfare effect of these five changes must equal to zero:

$$dM + dB + dW + dI_M + dI_B = 0 \tag{1.4}$$

In the literature, the key trade-off is represented by dM and dB, which together determine how the tax increase changes total revenue. We will discuss these consequences as *revenue effects*. As this calculation is well-known, we will discuss it quickly. The behavioral response dB represents a tax revenue loss, while the mechanical effect dM represents a tax revenue gain. In our set-up the revenue gain from those above z is $[1 - H(z)]\, dz\partial\tau$. The revenue loss from those in the band is denoted by $-dz\partial\tau \cdot \epsilon(z)zh(z)\tau(z)/\left(1 - \tau(z)\right)$, where $\epsilon(z)$ is the elasticity of earnings $\epsilon(z)$ with respect to $1 - \tau(z)$ (see Appendix I.D for derivation). The two terms together represent a revenue collection trade-off and together form the basis for the calculation of the revenue-maximizing tax rate. In non-Rawlsian SWFs there is also a pertinent welfare loss from the agents above the tax bracket who have their individual incomes reduced, dW. This effect dampens, but cannot cancel, the revenue-based benefit of the mechanical effect due to the assumption of SWWs that are non-increasing in income, and equals $-dz\partial\tau \int_{\{j:z_j \geq z\}} g_j dj$.

The mechanical effect and behavioral responses also impact post-tax income inequality directly. This is not considered welfare-relevant in traditional models, as the welfare effect of individual income changes is already taken into account through dM, dB and dW. In our model the inequality externality creates a welfare-pertinent effect. In the following we will assume a negative inequality externality for simplicity.[23]

The mechanical (in)equality effect is denoted as dI_M and the (in)equality effect of the behavioral responses is denoted by dI_B. Before calculating the welfare effect, it is convenient to first calculate the effects of these channels on post-tax income inequality; we denote these effects as $d\theta_M$ and $d\theta_B$ respectively.

Recall that the mechanical effect implies a collection of income from those above a certain tax bracket and a redistribution of this income as a flat dividend to all individuals.[24] The flat dividend does not affect absolute inequality metrics, and so we can focus on where income is reduced.[25] The effect of this income reduction on post-tax income inequality is the same as the classical mechanical revenue effect weighted by the importance of the individuals above z in the inequality metric. In other words, each dollar of additional revenue from an agent at income $z' > z$ from the mechanical effect corresponds to an inequality reduction of $\kappa(z')$. The mechanical effect thus changes income inequality by $d\bar{\theta}_M = -\overline{\kappa}(z)\left[1 - H(z)\right] dz\partial\tau$, where we have defined the average inequality metric $\kappa(z')$ above z as $\overline{\kappa}(z)\left[1 - H(z)\right] = \int_{\{j:z'>z\}} \kappa(z')h(z')dj$.[26] As

[23] The same intuition with the opposite welfare direction holds for a positive externality.

[24] Any change from a flat dividend would be equivalent to changing the marginal tax schedule.

[25] Note as well that if we were to use *non-absolute* inequality metrics (where flat income increases change the relevant statistic), the intuition would be overall similar with additional terms to correct for changes in average income.

[26] In the absolute Gini, $\overline{\kappa}(z) = H(z)$.

$\bar{\kappa}(z) \geq 0$, this effect always reduces income inequality regardless of the tax bracket in question except at the very bottom.[27]

The behavioral responses indicate a reduction in the work effort and thus the income of agents at z. This also affects post-tax income inequality. The (in)equality effect depends on the weight of these individuals in the inequality metric $\kappa(z)$, how much they change their work effort represented by $\epsilon(z)$, and the change in their post-tax income. We show in Appendix I.D that this is equal to $d\bar{\theta}_B = -\kappa(z) \cdot dz\partial\tau \cdot \epsilon(z)zh(z)$. As $\kappa(z)$ changes signs across the distribution, so does $d\bar{\theta}_B$. At the bottom, behavioral responses increase income inequality. At the top, behavioral responses decrease income inequality. Notably, this means that behavioral responses are welfare-*positive* at the top under a negative externality. The changing sign of $d\bar{\theta}_B$ across the distribution contrasts with the always negative dB, and is a key difference between the traditional revenue effects and the new equality impacts. We also note that the reliance on the elasticity $\epsilon(z)$ reverses the standard intuition from the revenue channel, where a high elasticity leads to a low tax rate (as the state should keep tax rates low to collect what little revenue they can). In our case, the state might instead prefer to place high tax rates (or subsidies) at the ends of the distribution to increase or decrease inequality as they see fit.

We summarize the discussion of the revenue and inequality effects in Table J1.

In terms of utility, each (in)equality effect impacts individuals as $\int_j \frac{\partial U_j}{\partial \bar{\theta}} \cdot \partial\bar{\theta} \cdot dj$. We have that $\eta_i = MRS_{x\bar{\theta}} = -\frac{\partial U/\partial\bar{\theta}}{\partial U/\partial x_i}$, and thus the total welfare effect is $dI = \int_j(-g_j\eta_j) \cdot (d\bar{\theta}_B + d\bar{\theta}_M) \cdot dj = -(d\bar{\theta}_B + d\bar{\theta}_M) \cdot \int_j \eta_j g_j dj$. This illustrates what was previously asserted, which is that heterogeneous inequality externalities can be weighted by SWWs g_j to become functionally equivalent to a homogeneous inequality externality. As such we denote $\eta = \int_j \eta_j g_j dj$ as the net inequality externality effect.

We can now consider the externality-induced sign change to optimal marginal tax rates as compared to the standard case. At the bottom, where dI_M and dI_B are in opposition (regardless of whether the externality is positive or negative), the welfare effect of a tax increase through the externality dimension is ambiguous. The change to the optimal marginal tax rate due to the externality is thus also ambiguous. At the top, where the signs of dI_M and dI_B harmonize – both are positive (negative externality) or negative (positive externality) – the change to the optimal tax rates is also unambiguous. Under a negative (positive) inequality externality there are unambiguously higher (lower) welfare benefits from increasing the marginal tax rate as compared to the standard case. Compared to the standard case, it follows that resulting top optimal rates are higher with a negative post-tax income inequality externality and lower with a positive post-tax income inequality externality.

Inserting the values from the preceding discussion into (1.4) allows us solve for $\frac{\tau(z)}{1-\tau(z)}$ and find that

$$\frac{\tau(z)}{1-\tau(z)} = \eta\kappa(z) + \frac{\eta\bar{\kappa}(z)}{\alpha(z)\epsilon(z)} + \frac{\big(1-\bar{G}(z)\big)}{\alpha(z)\epsilon(z)}, \tag{1.5}$$

where we use the local Pareto parameter $\alpha(z) = \frac{zh(z)}{1-H(z)}$ and the average SWW above z

[27] Formally this is due to $\kappa(z)$ being non-decreasing and the assumption that $\int_{\underline{z}}^{\bar{z}} \kappa(z)dH(z) = 0$.

denoted by $\bar{G}(z)$.[28] The last term corresponds to the traditional Saez [2001] result under our assumptions. The two former terms are due to the inequality externality and correspond to dI_B and dI_M respectively, and correspond to a Pigouvian correction of the no-externality tax schedule.

The behavioral response: $\eta\kappa(z)$ The first term comes from dI_B in (1.4) and represents the behavioral responses of the individuals who are located at income z.[29] This term corresponds to the equality impact from the substitution effect of a price change (the price of leisure, in this case). Agents at income z work less due to the tax increase as they substitute into leisure.[30] The revenue consequence is that tax revenue is reduced no matter the location of the tax increase. The direction of the equality impact, on the other hand, is conditional on the location of the tax bracket. The new term incentivizes individuals who make socially suboptimal labor choices to substitute into leisure, keeping their utility relatively high.[31] The term directly depends on (i) the inequality externality magnitude η, as a larger externality leads to a larger welfare gain from reducing inequality, and (ii) the relative weight of the agents at z in the inequality metric $\kappa(z)$; how these individuals' incomes contribute to the inequality metric determines how their income losses influence inequality and thus social welfare.

This term cannot be approximated by non-negative income-based SWWs g_i (or any utility-based SWWs, see Appendix I.C). It invalidates three classic results from the literature based on Mirrlees [1971] noted by Sadka [1976] and Seade [1977] – (i) that the optimal marginal tax rate at the top of a bounded distribution should be zero, as it is instead $\tau(z) = \frac{\eta\kappa(z)}{1+\eta\kappa(z)}$,[32] (ii) that the optimal marginal tax rate at the bottom should be zero, and (iii) that the optimal marginal tax rate is bounded between zero and one. These are not new findings in a mathematical sense, as the same is shown for relative income concerns by Oswald [1983].[33] Still, the modifications to the classic OIT results are intuitively appealing given the simplicity of the inequality externality, and as such we mention them here. One could see these previously controversial results as a consequence of the Mirrlees [1971] model assuming that economic equality in itself has no concrete value.

The mechanical effect: $\frac{\eta\bar{\kappa}(z)}{\alpha(z)\epsilon(z)}$ The second term comes from dI_M in (1.4) and represents the mechanical effect on the agents located above income z. As agents above z face an additional lump-sum tax and do not change their labor decisions, their post-tax incomes decrease. Tax revenue is increased no matter the location z of the tax increase – except at the very top – with an accompanying welfare loss of those above z who have their incomes reduced. The tax

[28] $\alpha(z) = \frac{zh(z)}{1-H(z)}$ is a distributional measure which becomes constant in a Pareto distribution. $\bar{G}(z)$ is defined as $\bar{G}(z)\,(1-H(z)) = \int_{\{j:z_j \geq z\}} g_j dj \big/ \int_j g_j dj$. In the Rawlsian min-max framework, $\bar{G}(z) = 0$. See Saez [2001] for further discussion on these variables.

[29] Agents above z do not change their labor choice due to the assumption of no income effects.

[30] We note that $\epsilon(z)$ and part of $\alpha(z)$ originate from this substitution effect despite not entering into the term in (1.5).

[31] This does not imply that the social planner wants to punish certain individuals. While the social marginal welfare of *income* can be negative, the social marginal welfare of *utility* is never negative, all else equal (upholding the Pareto principle).

[32] Reducing the income of the top-earner has become a social cost or benefit in itself, and should be a subsidy or tax depending on the direction of the inequality externality.

[33] Generally speaking these three results are fragile and change with many small modifications to the model – see Stiglitz [1982] and Saez [2001] for examples.

revenue is redistributed uniformly to every agent, so the post-tax income of every agent below z increases. Post-tax income inequality, as a result, decreases no matter the location of z except at the very top (where no revenue is gathered) and at the very bottom (where the uniform tax equals the uniform transfer). This provides the government an additional incentive to increase tax rates in addition to the standard revenue considerations; assuming a negative (positive) inequality externality, this term unambiguously increases (decreases) the marginal rate in every tax bracket except at the very top and at the very bottom.

How much this increases optimal marginal tax rates at any z depends notably on (i) the total magnitude η of the inequality externality, as before, and (ii) the relative average weight of the agents above the tax bracket z in the inequality metric, represented by $\bar{\kappa}(z)$; if these individuals' incomes contribute heavily to the inequality metric on average, their income losses also heavily reduce inequality and thus improve welfare. In addition, the standard model parameter values $\epsilon(z)$ and $\alpha(z)$ also appear here in this formulation.[34]

The externality thus introduces two new terms to the optimal tax formula. In general, the new key model variables are the size of the inequality externality (represented by η) and the choice of the relevant inequality metric (represented by κ).

4. Optimal Income Taxation: Numerical Simulations

In this section we use numerical calculations to find optimal marginal tax rates in the presence of a post-tax income inequality externality.

4.1. Numerical specification

We use the mechanism design solution from Appendix I.C for the numerical specifications, where individuals are on a continuum of wage-earning abilities n with associated density distribution function $f(n)$ and cumulative distribution function $F(n)$. The associated weight in the post-tax income inequality metric is $\kappa(n)$. We assume quasi-linear utility in consumption, a constant labor elasticity E_L, and a separable linear homogeneous inequality externality. This implies the utility function

$$U(x, l, \bar{\theta}) = x - \frac{l^{(1+\frac{1}{E_L})}}{\left(1+\frac{1}{E_L}\right)} - \eta\bar{\theta}, \tag{1.6}$$

where l is individual labor supply. We will logarithmically scale in the SWF to introduce a classical inequality-aversion motive.[35] The resulting optimal marginal tax rates $t(n)$ at each productivity level n are,

[34] It would be misleading to consider these two parameters as "part of" the mechanical effect. If the tax rate was equivalently written as a function of $\tau(z)$ such that

$$\tau(z) = \frac{1 + \eta\kappa(z)\alpha(z)\epsilon(z) + \eta\bar{\kappa}(z) - \bar{G}(z)}{1 + \eta\kappa(z)\alpha(z)\epsilon(z) + \eta\bar{\kappa}(z) + \alpha(z)\epsilon(z) - \bar{G}(z)},$$

then $\alpha(z)\epsilon(z)$ occurs in the term for the behavioral responses. This is more intuitive for $\epsilon(z)$, which directly determines how individuals react to the tax change. The local Pareto parameter $\alpha(z)$ indicates the number of individuals who are in the tax bracket z as a proportion of the individuals above z; as this proportion changes, the relative strength of the mechanical effect and the behavioral responses changes with a resulting effect on optimal tax rates.

[35] Such that the Utilitarian case is equal to $W = \int_i log(U_i)di$, for example.

$$\frac{t(n)}{1-t(n)} = \eta\kappa(n) + \eta\left(1+\frac{1}{E_L}\right)\frac{\bar{\kappa}(n)}{\alpha(n)} + \left(1+\frac{1}{E_L}\right)\frac{1}{f(n)n}\int_n^\infty \left[1-\frac{W_{U(p)}}{\lambda}\right] dF(p),$$

where λ is the marginal value of public funds, $\alpha(n)$ is the local Pareto parameter, and $W_{U(p)}$ is the derivative of the SWF with respect to utility (capturing the aforementioned inequality aversion).

The underlying wage-earning ability distribution n is found through inverting the observed income distribution using the solution to the individual's maximization problem, following Saez [2001] and others. We use the US Distributional National Accounts micro-files to measure the 2019 U.S. labor income distribution.[36] The NBER TAXSIM model was used to find marginal tax rates on labor income for any given tax unit in the DINA files.[37] The main focus of the numerical simulations will be on how the inequality externality changes the results from the no-externality case; we thus largely follow the existing literature for the remaining model specification. For more details on the simulation procedure see Appendix I.E.1.

There are two further choices that are crucial for the simulations that are specific to the inequality externality. These are the choice of the relevant inequality metric $\bar{\theta}$ (e.g. the Gini coefficient in post-tax income) and the magnitude and direction η of the inequality externality. We detail these choices below.

4.1.1. Inequality metric

The inequality metrics we use follow the general form in (1.3), using wage-earning ability n instead of pre-tax earnings z.[38] In the main specification (Section 4.2) we use the Gini, which has the following form:

$$\kappa_G(n) = 2H(n) - 1. \tag{1.7}$$

We also show results for a generalized Gini with weights of the following form (see Section 4.3),

$$\kappa_T(n) = (q+1)H(n)^q - 1, \tag{1.8}$$

which was designed to analytically approximate top income shares (which have a discrete jump and are thus analytically intractable). The Gini corresponds to $q = 1$ in this specification, while larger q approximates top income share inequality metrics by increasingly weighting incomes at the top of the distribution while equalizing the weights of other agents' incomes. The weights of the Gini and the generalized Gini with $q = 4$ are plotted in Figure 2.[39] We also show the weights used in the top 10% income share for comparison, which is discontinuous and thus not usable in an analytical setting. Other inequality metrics are examined in Appendix I.E.3.

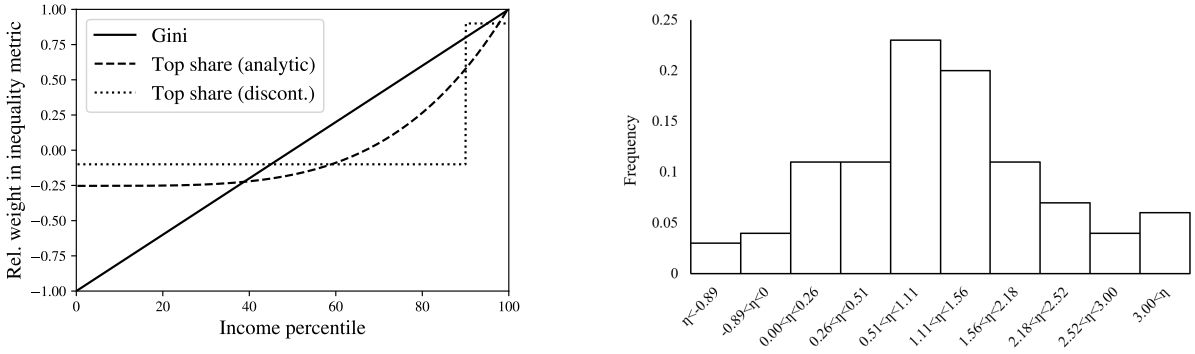
[36] Described in Piketty et al. [2018], accessed at https://gabriel-zucman.eu/usdina/ on March 22nd 2023.

[37] Described in Feenberg and Coutts [1993], accessed at https://taxsim.nber.org/ on April 20th 2023. More details in Appendix I.E.1.

[38] These are equivalent as $\kappa(z) = \kappa(n)$ by assumption.

[39] The figure shows the relative weight of the income of any agent when calculating the specified inequality metric.

Figure 2: Weights of Inequality Metrics (left), estimated η_G (right)



Note: Right: The relative weights of individuals' income in the inequality metrics we primarily use (the Gini and the analytic top share metric are used in Figures 4 and E4, respectively). This corresponds to $\kappa(z)$ in the general expression $\bar{\theta} = \int_{\underline{z}}^{\bar{z}} \kappa(z)x(z)dH(z)$. More inequality metrics are explored in Appendix I.E.3. Left: The estimated magnitudes of the inequality externality magnitude η_G from the survey experiment in Carlsson et al. [2005]. In the following numerical simulations we restrict η_G between -0.5 and 2.0 (and equivalent values for other inequality metrics).

4.1.2. Inequality externality magnitude

Given the inequality metric we need to choose values for the inequality externality magnitude. The values of η depend on which inequality metric is chosen to be relevant for the externality, and we denote the values calculated for the Gini coefficient as η_G. As there are unavoidable empirical challenges in calibrating such a number,[40] we do not aim to strongly argue for any one value. We instead use a range of realistic η_G to illustrate the potential tax policy consequences of various income inequality externalities. We present three different methods to understand the magnitudes of these η_G.

Correlation-based estimates To make a reasonable first-pass at an order of magnitude of η_G one could take the cross-country correlation between income inequality and externality dimensions – naively taking the correlation after controlling for observables as causal – and use willingness-to-pay estimates for each externality dimension to find the dimension's contribution to the total η_G. We do this for intentional homicides as an illustrative example. We use data from the World Bank for homicides, the World Inequality Database for the Gini coefficient, and Cohen et al. [2004] for the societal willingness to pay for fewer homicides.[41] The correlation between income inequality and intentional homicide is strongly positive, and through this very simple and likely biased approach we find $\eta_{G,homicides} \approx 0.07$.

This only represents a single externality channel, and the full η_G estimate would be found as $\eta_G = \sum_k \eta_{G,k}$. Extending this method to find all $\eta_{G,k}$, however, requires internationally comparable outcome data.[42] This is not a trivial requirement, and precludes the use of more detailed crime data.[43] Other internationally comparable outcomes usually lack willingness-

[40] Beyond specific empirical challenges relating to the existence and quality of the available data, it is very challenging – perhaps impossible – to find true exogenous variation in macroeconomic inequality.

[41] Cohen et al. [2004] estimates the total social cost of a homicide as $9.7 million, or $12.8 million corrected for inflation to 2018.

[42] We note that this approach assumes that the inequality externality operates on the country-level.

[43] Harrendorf [2018] notes the following: "Crime levels are not a valid measure of crime in different countries, with the possible exception of completed intentional homicide. Total crime rates depend mainly on the internationally differing quality of police work."

to-pay estimates, making further use of this approach complicated even under the stringent assumptions we impose.

Experimental estimate To find a range of η_G that takes into account *all* externality dimensions we present estimates based on data from Carlsson et al. [2005]. The work uses a survey design to estimate macroeconomic inequality aversion in Swedish university students.[44] The survey, which asks respondents to decide what income-inequality trade-off their hypothetical grandchildren would prefer, allows us to find individual preferences for η determined to an interval.[45]

The distribution is presented in Figure 2. The median respondent in the survey has approximately $\eta_G = 1.00$. A majority of respondents have $0.26 < \eta_G < 2.18$.[46] A negative η_G – indicating a preference for inequality, or that inequality is a positive externality – is only observed in 7% of respondents. The equivalent externality magnitude values for top income shares, η_T, are calculated from the same experiment. As a general rule of thumb, $\eta_G \approx 2\eta_T$ when externality magnitudes are equal.

Hypothetical exercises As these numbers are rather abstract, we present an alternative way of understanding the magnitudes through equivalent incomes. Answering the following question pins down either η: *What multiple of their current income should an average agent require to move from Denmark-like to United States-like inequality?*[47]

Answering the question creates equivalent incomes for the two differing inequality levels, which allows us to pin down an inequality externality magnitude.[48] These equivalent incomes for Denmark and the United States, and their corresponding η_G, are shown in Table 2. As an example, if we have an inequality externality of $\eta_G = 1.0$, the average individual in a society with Denmark's inequality level would require 13% more income to be indifferent if inequality increased to the U.S. level.

Based on these techniques we use the range $-0.5 \leq \eta_G \leq 2.0$ for the Gini-based externality and $-0.15 \leq \eta_G \leq 1.0$ for the top share-based externality in the main numerical simulations.

Evaluating these externality values in the simulations also gives us a way to measure the significance of the inequality externality. In Figure 3 we show the cost of the inequality externality as the percentage of consumption each income percentile would be willing to give up to remove the externality for each η_G used in the main specification. Although the distribution of the ex-

[44] Bergolo et al. [2022] finds comparable numbers for Uruguayan university students.

[45] Using a survey experiment instead of a direct externality estimate means that we are relying on potentially biased beliefs to proxy for inequality's externality effects. There is also selection bias in the survey respondents and, because the only degree of freedom is being used to estimate the extent of inequality aversion, it is not possible to know how well our assumed utility function matches the respondents' perceived utility functions. All these reasons contribute to why we are using a *range* of η_G.

[46] Due to the design of the experiment, any one individual's inequality aversion is only pinned down to a range.

[47] Assuming the same leisure, that the mean income difference between the two countries is negligible, and that relative position is irrelevant. According to the 2017 World Economic Outlook database GDP per capita is \$61,803 in Denmark, and \$59,707 in the United States. Calculations are based on Gini coefficients of 0.410 for the United States and 0.285 for Denmark.

[48] These numbers are significantly dependent on the income specified (average income in the above case) under a homogeneous inequality externality. They can also be interpreted more generally, however. Under Utilitarianism, quasi-linearity in consumption, and heterogeneous linear and separable inequality externalities, the same percentage can be thought of as the total society-level income increase that would be required for indifference when $\eta = \frac{1}{n}\sum_i \eta_i$. These assumptions correspond to a social welfare function such that $W = \sum_i \left(x_i - \eta_i \bar{\theta}\right)$, as in Sen [1976].

Table 2

The Magnitude of Inequality Externalities η_G

	$\eta = -0.5$	$\eta = 0.0$	$\eta = 0.5$	$\eta = 1.0$	$\eta = 2.0$	$\eta = 3.0$
U.S. Income Multiplier	0.94	1.00	1.06	1.13	1.25	1.38

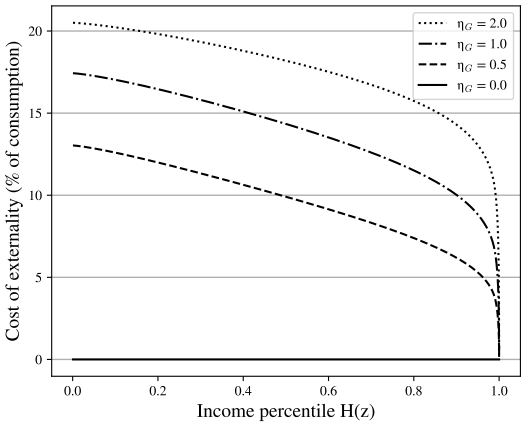
Note: Which multiple of their current income would an average-income agent need to move from Denmark-like to U.S.-like inequality? Above are these equivalent incomes for various levels of the inequality externality η_G from the utility function in (1.6).

ternality impact is not particularly meaningful in our case – the net welfare effect is the policy-determining variable – we believe the illustration gives the reader an idea of the magnitudes we introduce. There are two further caveats to this figure. First, these values are endogenous, as the social planner has already reduced inequality due to the externality. Second, inequality levels are generally very low in optimal income taxation simulations even without an inequality externality; applying the same η_G to real-world inequality levels would mean much higher externality costs.

4.2. Main results: The Gini externalities

Our main specifications, using the Gini as the post-tax income inequality metric, are presented in Figure 4. To remain general we show both positive and negative inequality externalities. The introduction of even a small post-tax income inequality externality substantially changes the optimal tax structure. The effect is larger towards the top of the income distribution.

Figure 3: The cost of the Utilitarian-based inequality externalities used in Figure 4 for each income percentile.



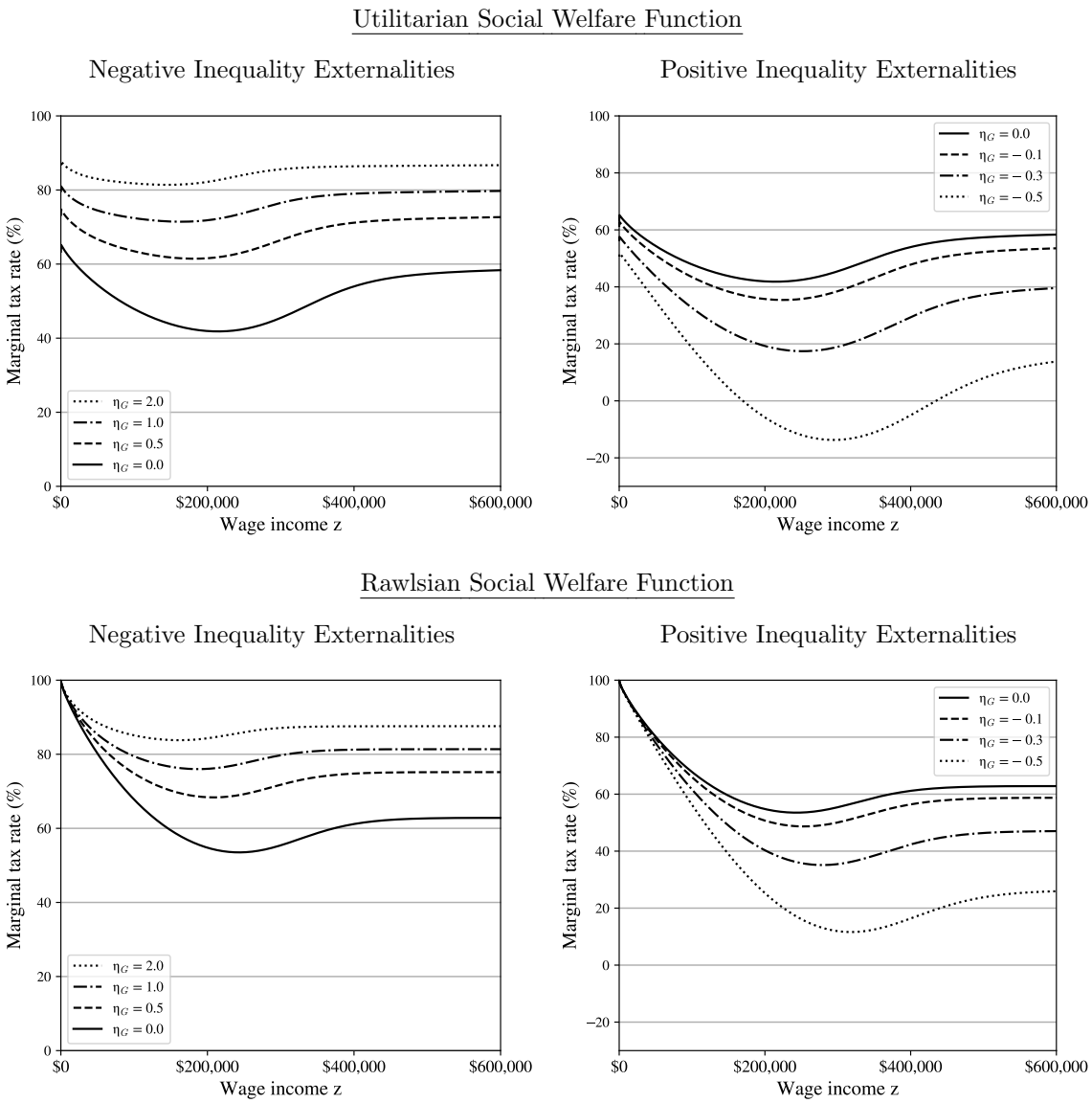
First, note that at the very top the Utilitarian and Rawlsian results converge under any externality value, as in the classical literature.[49] The magnitude of the inequality externality, however, is naturally impactful for the optimal top tax rate. This illustrates that a Rawlsian SWF, in itself, does not imply a maximum dislike of inequality.

We thus begin by discussing optimal top marginal tax rates, which are the same in both the Rawlsian and the Utilitarian case. With no inequality externality, the optimal top rate is 63%. For $\eta_G = 1.00$, the value closest to the empirical externality estimate taken from Carlsson et al. [2005], it is 81%. When assuming a larger negative inequality externality, $\eta_G = 2.0$, the top rate increases to 88%. With a small positive inequality externality ($\eta_G = -0.5$), the optimal top marginal tax rate is only 26%. The inequality externality magnitude thus has a large impact on the optimal top tax rate; we will discuss this further in Section 4.4.

In the Utilitarian case, the marginal tax rates are shifted up or down from the no-externality case across the entire distribution. This is due to the empirical strength of the mechanical effect

[49] This is due to the assumptions of separability and (specific to the simulations) a homogeneous inequality externality.

Figure 4: Optimal Marginal Income Tax Schedules with Gini Inequality Externalities



Notes: Optimal marginal tax rates for various Gini-based inequality externalities with magnitudes η_G, where inequality is either a negative externality (left) or a positive externality (right). The social planner is Utilitarian (above) and Rawlsian (below). The Utilitarian and Rawlsian cases converge when moving towards the top for a given externality value. Empirical estimates indicate $\eta_G = 1.0$. The solid line, $\eta = 0$, is the standard no-externality case. Further explanation of η is in Table 2. Note the different scales of the vertical axes between the negative and positive externalities.

(which increases/decreases optimal rates across the entire distribution for a negative/positive externality), which dominates that of the behavioral responses (which increases or decreases optimal rates differentially at the top and bottom) under our parameter choices.[50] The effects are larger near the top, which is particularly noticeable around the 95^{th} percentile. The larger effects near the top of the distribution is due to the equality effects of the mechanical and behavioral channels working in the same direction in this region, as discussed in Section 3.1.[51] We also note that all simulations have lower optimal tax rates around the 90^{th}–95^{th} percentiles due to the well-known decrease of the local Pareto parameter around these values, which leads to the classical U-shape found in the literature [Diamond, 1998]. We return to this shortly.

The Rawlsian externalities we introduce have negligible impacts near the bottom of the distribution, where marginal tax rates are very high in the no-externality case. This is driven by a very high mechanical revenue benefit of taxation near the bottom (which is also found in the classical literature) drowning out any effect of the externality.[52] The effects of the inequality externality are mostly located above the 90^{th} percentile for both negative and positive externalities. Under a positive externality, top marginal tax rates approach zero around the 97^{th} percentile.

The extent of the classical U-shape varies across simulations. It is most striking in the positive externality and no-externality simulations, and is difficult to notice in the negative externality simulations. As the U-shape has been widely discussed as having potential implications for practical tax design it is relevant to ask why this occurs. The U-shape emerges from the empirically estimated wage-earning (or income) distribution, as the local Pareto parameter α is high around these wage (or income) percentiles. In short this implies a relative over-density of individuals *in* these tax brackets compared to those *above* these tax bracket, which in turn implies that the relative strength of the behavioral channel is high in this bracket (as compared to the relatively low strength of the mechanical effect). In other words, optimal tax policy in these brackets is increasingly set by the welfare consequences of agents' behavioral responses. This decreases the no-externality optimal tax rates in the region. How does this change when one introduces an inequality externality? In the negative externality case, there is a welfare-positive dimension to the behavioral responses (namely decreased inequality). It follows that

[50] This result is not universal, and the effect of the externality at the bottom is usually smaller than in this case due to the counteracting behavioral response. Indeed, the Utilitarian case with no income effects has among the least top-heavy distributional optimal policy effects of any of our simulations. It is notable that the effects are largest at the top even in this case. Using certain skill distributions, such as the full Pareto distribution in Appendix I.E.2, a negative externality *decreases* optimal marginal tax rates at the bottom. We also find this result with any pre-tax income inequality externality (see Section 4.6).

[51] We observe negative optimal marginal tax rates for income earners between the 84^{th} and 98^{th} percentiles when $\eta_G = -0.5$. These negative optimal top rates come from the social planner's incentive to increase income inequality when inequality is a positive externality, even if this comes at a significant revenue cost – to the extent that a tax subsidy at the top can be optimal.

[52] The high optimal rates at the bottom of all the Rawlsian simulations are due to the large positive mechanical revenue effects of increasing bottom marginal tax rates. When one only cares about the very bottom agent, as in the Rawlsian case, redistributing away from any other agent is a net positive absent changed labor choices. Since we do not consider income effects, these labor choices do not occur for anyone above the tax bracket in question. The mechanical revenue effect is thus very large at the bottom and leads to very high marginal tax rates in this region. The introduced equality effects are not large enough to change this substantially. In contrast, the Utilitarian simulations take into account the income losses from agents above the tax bracket, which discounts the mechanical benefits of tax increases near the bottom. Very high bottom marginal tax rates are thus less appealing, and the effects of the inequality externality are more visible.

an increased importance of the behavioral responses does not necessarily imply a U-shape and lower optimal tax rates – as we can see in the simulations.[53] In the positive externality case, meanwhile, the shift towards a concern for behavioral responses is still highly relevant, as the behavioral responses remain entirely welfare-negative (through decreased revenue and decreased inequality). To summarize, the classical U-shape from the optimal taxation literature may depend on the absence of a negative income inequality externality.

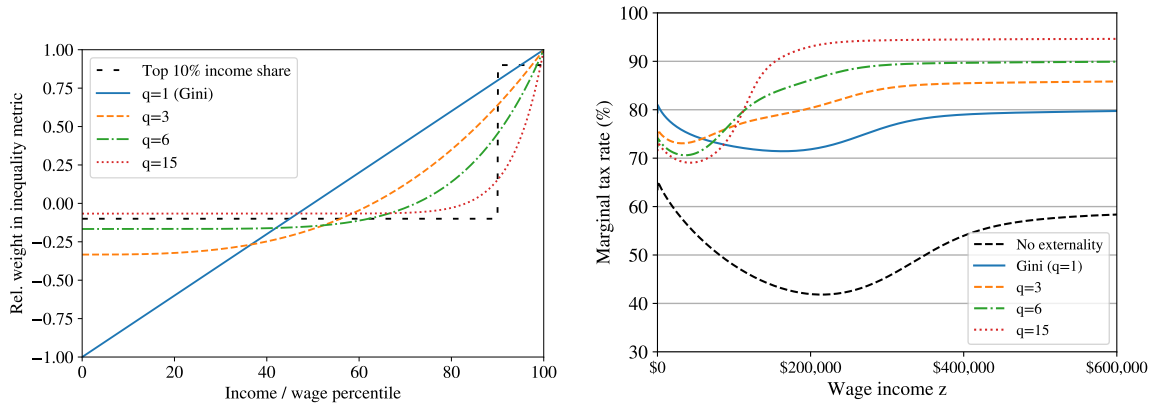
The exact optimal tax structure depends heavily on the model specification, so the numerical simulations should be interpreted with caution.

4.3. Robustness: Top income share externalities

The choice of the inequality metric naturally influences our results. And while the Gini coefficient is analytically appealing, it is often considered to over-weight middle-income inequalities. To address this concern we present a robustness check of our main findings where we use the general top income share metric family $\kappa(z) = (q+1)H(z)^q - 1, q \in \mathbb{N}$ as the relevant inequality measurement, with increasingly larger q. After $q = 1$, which defines the Gini coefficient, this inequality metric becomes increasingly top-focused and approximates top income share metrics.

We show a set of such inequality metrics and the effect of using them in the optimal tax calculation in Figure 5. The externality in the optimal tax calculation is kept constant at the median result from Carlsson et al. [2005].[54]

Figure 5: Varying the Inequality Metric with a Fixed Externality Magnitude



Note: Left: The income weights over the distribution of various inequality metrics in the family where $\kappa(z) = (q+1)F(n)^q - 1, q \in \mathbb{N}$. The top 10% income share is also plotted. Larger q indicates that top incomes are increasingly weighted. Right: Optimal marginal tax rates for these inequality metrics, keeping the magnitude of the inequality externality constant for all q at the upper bound of the median value from the empirical inequality aversion estimates in Carlsson et al. [2005]. The no-externality case is shown as a reference in dotted black. The wage-earning ability distribution is the empirical income distribution, and the SWF is Utilitarian.

When we move away from the Gini towards a top income share, the effects of the externality are increasingly concentrated towards the top of the distribution. This should not be surprising given the increasing weight of top incomes in the inequality metric, although the *magnitude* of

[53] Optimal marginal tax rates can even increase in the region under different specifications. In Section I.H.1 this occurs under a negative pre-tax income inequality externality.

[54] The actual values of η change, as estimating η from the Carlsson et al. [2005] data requires an assumption about which inequality metric to use. Changing this inequality metric also changes the calculated η.

the effect is large. The inequality metric defined by $q = 15$ coupled with the median inequality externality from Carlsson et al. [2005] leads to optimal top marginal tax rates above 95%.

It is also noticeable that the effects near the bottom are reduced. This is not as obvious, as lower inequality metric weights near the bottom have opposite optimal tax effects through the behavioral channel (through which lower κ_{bottom} leads to higher τ) and the mechanical effect (through which lower κ_{bottom} generally leads to lower τ through a higher $\bar{\kappa}_{bottom}$).[55] In the numerical simulations, the mechanical effect is more powerful, indicating that the average marginal externality above is more impactful than marginal externality of the tax bracket itself. Due to this, tax rates for the majority of Americans would be closer to the no-externality case under inequality metrics that focus more on top income shares.

Overall, using top income shares further concentrates the effect of the externality towards the top of the tax schedule. With other inequality metrics, such as those in the S-Gini family, results are overall similar. This is further discussed in Appendix I.E.3. In sum, the Gini is a conservative choice which dampens effects at the top in return for larger changes across the rest of the distribution. We will now discuss implications for top tax rates specifically.

4.4. Equality concerns: Top tax rates

As we have discussed in the preceding sections, the new equality concerns have a particularly large effect on the optimal top tax rate. The optimal tax rate near the top in the small-perturbation framework with a Gini post-tax income inequality externality is,

$$\tau(z) = \frac{1 + \eta + \eta\alpha(z)\epsilon(z)}{1 + \eta + \eta\alpha(z)\epsilon(z) + \alpha(z)\epsilon(z)}, \tag{1.9}$$

which is strongly dependent on the inequality externality magnitude η. It is useful to discuss why this occurs.

Revenue considerations, which in this context implies the direct individual effects from the redistribution of income, have few distributional biases. In a Rawlsian set-up, for instance, one tax dollar raised remains one tax dollar raised, regardless of which tax-payer pays it (if not taken from the very bottom).[56] Equality concerns are naturally different: *where* the income is taken from is of key importance. And, as we have seen, the tax policy effects of these equality concerns generally increase as one approaches the top of the distribution.

This implies that the optimal tax rate can be above the revenue-maximizing rate (the so-called "Laffer rate"). The revenue-maximizing rate is occasionally used as an upper bound for sensible tax rates. For example, Piketty et al. [2014] states that they "focused on the revenue-maximizing top tax rate, which provides an upper bound on top tax rates". This position would need to be modified in a model with societal effects of inequality. We discuss this further in Appendix I.F.

[55] In the case of the behavioral channel, the bottom-earner imposes less of an externality and the negative Pigouvian term is thus smaller. In the case of the mechanical effect, redistributing from everyone above is less impactful for inequality-reduction if everyone in the lower half is weighted relatively equally.

[56] In general the welfare changes from a tax and its associated revenue across the distribution is dependent on the SWF. However, the net distributional biases are mechanically constrained due to the non-negativity of the SWWs.

4.4.1. Large variation in top rates: A maximum income, or the Rawlsian Conservative?

Some of the variation in international tax brackets, particularly at the top, could be due to policy setters' differing considerations of the inequality externality. Two Rawlsian governments might agree on the elasticity of earnings and revenue-maximizing tax rates and still strongly disagree on optimal top tax rates – *if* they disagree on how inequality changes society. Indeed, varying the value of the inequality-sensitivity parameter η_G has a larger effect on optimal top tax rates than varying the standard parameter values $1/\alpha$ or E_L, which we show in Tables J2 and J3. By changing η_G within reasonable bounds, the same Rawlsian social planner can find optimal top tax rates from close to zero to over 90%. Under stronger positive externalities the same social planner can even find negative optimal top rates. In other words, a wide range of top tax rates can be optimal depending on the magnitude of the inequality externality.

This contrasts with standard OIT models, where top marginal income tax rates usually converge to around $60-70\%$ regardless of the underlying SWF. Although these numbers depend heavily on parameter specifications, heterodox assumptions are required for optimal rates below 50% or above 80%.[57] Our model thus rationalizes a wider array of tax schedules. We use two real-world examples to illustrate the power of such a finding.

First, the idea of extremely high top tax rates (a "maximum income"). If one believes in a large negative inequality externality, the negative effect of top income earners on the rest of society is sufficient to argue for top tax rates above 90%. These are similar to tax rates from the post-war period in the United Kingdom, Germany, and the United States. The disincentive for high earners at this stage begins to approach a maximum income.

Second, the idea of a Rawlsian government with low tax rates on the highest income-earners. If one believes in even a small positive inequality externality, here represented by $\eta_G = -0.5$, marginal rates at the top quickly fall below 50% and begin approaching zero. We call this the Rawlsian conservative; the argument that a low top tax rate will lead to the highest possible utility for the worst-off agent.

Both of these intuitive arguments are occasionally discussed in the public sphere. In the standard OIT literature, however, they are unfounded. One strength of our model is that such arguments can be logically substantiated, and disagreements can be traced back to the variable η. Individual opinions on η could be related to (or even determinants of) political leanings and policy preferences, as argued by Lobeck and Støstad [2023].

4.5. U.S. social welfare weights with an inequality externality

As shown in Bourguignon and Spadaro [2012], it is possible to calculate the implied SWWs of the observed tax schedule given the relatively large assumption that the social planner is welfare-maximizing under the constraints of the optimal income tax problem.[58] This method is applied to the U.S. in Lockwood and Weinzierl [2016] and Hendren [2020], both of which generally find decreasing SWWs with income. Hendren [2020], which has more granular data,

[57] Piketty et al. [2014] finds revenue-maximizing rates varying from 57% to 83% with differing elasticity compositions, for instance.

[58] This is an unlikely assumption, as discussed in Lockwood and Weinzierl [2016]. Nonetheless, it is useful to see how current tax systems can be rationalized in the framework of optimal taxation.

also notes an increase in SWWs towards the very top of the distribution.

These methods implicitly assume that no inequality externality is taken into account by the social planner when setting the tax schedule. However, U.S. citizens generally believe that inequality has negative consequences [Lobeck and Støstad, 2023]. Such beliefs have also been voiced by prominent U.S. politicians.[59] It is thus natural to think that some concern for inequality itself could be included in the income tax schedule design. If so, under the same assumptions from Section 3, we show in Appendix I.G that the implied SWW $g(z)$ is,

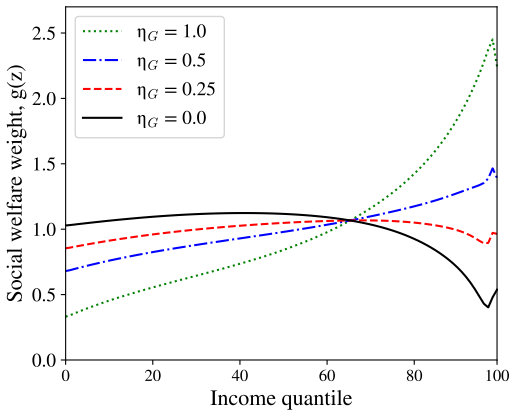
$$g(z) = -\frac{1}{h(z)}\frac{d}{dz}\left[(1-H(z))\,(1+\Upsilon(z)) - \frac{\tau(z)}{(1-\tau(z))}zh(z)\epsilon(z)\right],$$

which differs from the standard case by $\Upsilon(z) = \eta\alpha(z)\epsilon(z)\kappa(z) + \eta\bar{\kappa}(z)$.[60] Intuitively, the implied inequality aversion in a given tax system can come from either the SWF $g(z)$ or externality motivations $\Upsilon(z)$, and there is a substitution effect between these two motivations. If externality motivations to avoid inequality were greater when designing a given tax schedule, the same tax schedule will imply that the SWWs in the design process were less progressive.

In Figure 6 we show $g(z)$ of the 2019 U.S. tax system under standard specifications, assuming the social planner has taken into account various negative post-tax Gini income inequality externalities. The model specification is further discussed in Appendix I.G.

The standard case of no inequality externality ($\eta_G = 0$) has generally decreasing welfare weights with income with an upward bend towards the top of the distribution, similar to Hendren [2020]. Introducing a negative inequality externality ($\eta_G > 0$) changes implied SWWs quickly, however. Implied SWWs are relatively flat for $\eta_G = 0.25$, indicating that all the inequality aversion in the tax system is accounted for by such an inequality externality.[61] The implied SWWs are increasing for $\eta_G = 0.5$, and even more so for $\eta_G = 1.0$. For $\eta_G = 1.0$, the social planner values one dollar at the top equally to five dollars at the bottom.[62]

Figure 6: Implied social welfare weights $g(z)$ from the 2019 U.S. tax system under various negative inequality externalities η_G.



This illustrates our second main finding. The cur-

[59] For example Obama [2011]: "This kind of inequality – a level that we haven't seen since the Great Depression – hurts us all."

[60] A few technical points: We use the income density directly, as in Lockwood and Weinzierl [2016], instead of the "virtual" earnings density, as employed in Hendren [2020] and the rest of this work. Due to this the elasticity we use is technically defined to include the circularity between the "virtual" earnings density and the observed income density [Jacquet et al., 2013]. This is unlikely to significantly change results due to the absence of pronounced bunching in the actual U.S. income distribution [Saez, 2010]. We assume no income effects and no extensive margin behavioral responses for simplicity. A more detailed approach for the no-externality case can be found in Jacobs et al. [2017], which also notes that these factors are empirically small.

[61] It is useful to find the η_G above which social welfare weights become regressive. There are various ways to do this. The full linear trend is flat at roughly $\eta \approx 0.21$. As $g(z)$ is slightly increasing below the median, it is also useful to note the set of η_G which correspond to $\bar{G}(z_{median}) > g(z_{median})$, which indicates that the average social welfare weight above the median is higher than that of the median. The corresponding externality magnitudes are $\eta_G > 0.28$.

[62] For $\eta = 2.0$ we find negative SWWs at the bottom, indicating that the social planner would want to remove income at the bottom if this did not also increase inequality itself.

rent U.S. tax schedule cannot accomodate both a socially progressive transfer motive and be significantly concerned with inequality's societal effects. The social planner may have progressive $g(z)$, implying that the government prefers to transfer one dollar from the poor to the rich ceteris paribus (as in Lockwood and Weinzierl, 2016, and Hendren, 2020). The social planner may also have $\eta_G \geq 0.25$, implying a negative inequality externality of a potentially sizable magnitude. However, it cannot have both. The inequality aversion in the system as a whole is simply too small for this to be the case. It should be noted that this is, again, subject to our assumptions – particularly relevant here are the assumption of optimal tax design, Utilitarianism [Weinzierl, 2014], and the absence of migration responses [Lehmann et al., 2014].

The U.S. social planner may also have a *positive* inequality externality in mind. An inequality externality focusing on positive benefits from top-incomes could explain the puzzle of increasing SWWs at the top from Hendren [2020] (a result which is also visible in Figure 6). If the social planner believes top-income inequality is strongly beneficial for society – through increasing innovation, economic growth, or charitable giving, for example – the implied SWWs may still be everywhere decreasing. We illustrate this graphically in Figure G7.

Several other conclusions from the inverse optimal tax literature could change if inequality externality beliefs are a salient feature of policy-making. Lockwood and Weinzierl [2016] note that TRA86 implies a substantial change in SWWs over a short time period, which could be resolved if TRA86 instead represented a change in the *inequality externality belief* of the social planner – beliefs that are arguably more malleable than the SWWs themselves. Both Lockwood and Weinzierl [2016] and Hendren [2020] also create welfare estimates that depend on inequality not being an externality (or having been considered an externality in the tax design process).[63] More generally, the inverse-optimum literature is an example of a welfare-based framework that is relatively fragile to the inclusion of an inequality externality.

4.6. Other types of inequality externalities

The preceding sections have discussed a *post-tax income* inequality externality. While such an externality could be reasonable through several motivations – some of which we outline in Section 5.1 – there is no *a priori* reason to exclude the possibility of other inequality externalities. Here we consider how the theoretical intuition changes with different types of inequality externalities in the optimal non-linear income taxation problem. Note that the optimal marginal tax formula with a post-tax income inequality externality from (1.5) can be written as,

$$\tau(z) = \frac{1 + \eta\kappa(z)\alpha(z)\epsilon(z) + \eta\bar{\kappa}(z) - \bar{G}(z)}{1 + \eta\kappa(z)\alpha(z)\epsilon(z) + \eta\bar{\kappa}(z) + \alpha(z)\epsilon(z) - \bar{G}(z)}. \tag{1.10}$$

Pre-tax income inequality externality A pre-tax income inequality externality implies different equality impacts of the behavioral and mechanical effects. To start with the behavioral responses, note that any behavioral shift that follows from a tax increase would lead to a larger pre-tax income reduction than post-tax income reduction; pre-tax income being reduced by one

[63] Lockwood and Weinzierl [2016] calcualte the welfare cost of the inequality in income growth between 1980 and 2010 as 4.3% of total economic growth in the period. Similarly, Hendren [2020] creates a preference ordering of countries' income distributions based on implied SWWs. Two parts of these calculations would be affected by an inequality externality. First, the implied SWWs from the inverse-optimum method would change under an inequality externality, as shown in this section. Second, the total welfare implications of income changes would be affected by an inequality externality.

unit reduces post-tax income by only $1-\tau(z)$ units, which is between zero and one (excluding the extreme case of negative marginal rates). As such the effect of any behavioral response on pre-tax income inequality is generally larger than that on post-tax income inequality. Subsequently the pre-tax externality is more heavily affected by this channel than we saw in the post-tax case.

The mechanical effect, meanwhile, no longer has any impact on the externality. This follows from pre-tax income inequality being unchanged by the mechanical (post-tax) redistribution of income from those above the perturbation.

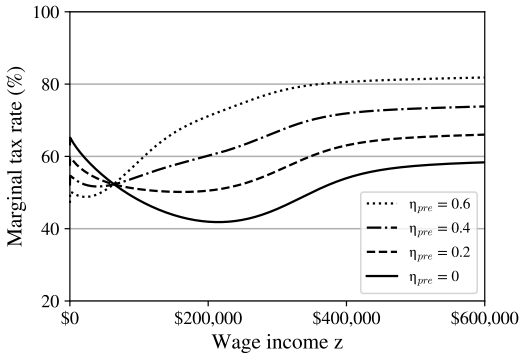
The optimal income tax rates in this case are

$$\tau(z) = \frac{1 + \eta_{pre} \cdot \kappa(z)\alpha(z)\epsilon(z) - \bar{G}(z)}{1 + \alpha(z)\epsilon(z) - \bar{G}(z)},$$

where η_{pre} is the pre-tax income inequality externality magnitude.[64] The full derivation is in Appendix I.H.1.

This result implies that a pre-tax income inequality externality could lead to a progressive modification of the standard Mirrlees tax rates (where we mean progressive in the traditional sense; marginal tax rates which increase with income). We see this in Figure 7, which shows negative pre-tax inequality externalities in the Utilitarian framework with the same specifications as in our main specification. Bottom tax rates are lower and top tax rates are higher than in the no-externality case, which is a general finding under separability. The marginal tax rates increase from 47% at the bottom to 85% at the top when $\eta_{pre} = 0.6$.[65]

Figure 7: Optimal income tax rates with a pre-tax income inequality externality. The social planner is Utilitarian, and the remaining specification is identical to Figure 4.



Interestingly, the pre-tax income inequality externality almost removes the well-known U-shape of optimal marginal tax rates from the classical literature. Instead, the marginal tax rates generally increase in income. Compared to the classical literature (or the case of a post-tax income inequality externality), this new optimal marginal income tax schedule is closer to that observed in most developed countries. One might wonder whether governments have, to some extent, considered pre-tax inequality as an ill in itself when designing tax schedules. If so, this could explain some of the differences between the numerical simulations from optimal tax theory and real-world tax schedules.

Utility inequality externality When considering a utility inequality externality, the behavioral channel no longer has an inequality impact. This follows from a miniscule tax perturbation from the optimum only leading to second-order utility effects. The mechanical effect would function similarly as in the post-tax income inequality case, as increasing the marginal tax rate

[64] There is a subtle point to be made here about the magnitude of η_{pre}. Pre-tax income inequality is generally higher than post-tax income inequality, which influences the shadow price of each unit of inequality and hence η. To keep externality sizes similar we thus use a lower set of η_{pre} in Figure 7 than the corresponding η_G in Figure 4.

[65] This corresponds roughly to $\eta_G = 2.0$ in Figure 4.

reduces utility inequality by lowering the utility of those above the tax bracket.[66]

The optimal income tax rates in such a case are

$$\tau(z) = \frac{1 + \eta_U \cdot \bar{\kappa}(z) - \bar{G}(z)}{1 + \alpha(z)\epsilon(z) + \eta_U \cdot \bar{\kappa}(z) - \bar{G}(z)},$$

where η_U is the utility inequality externality magnitude. The full derivation is in Appendix I.H.2. Assuming that negative weights are acceptable, using the modified SWWs $\bar{G}'(z) = \bar{G}(z) - \eta_U \cdot \bar{\kappa}(z)$ allows this result to be simplified to the standard Mirrlees case without the need for empirical variables in the modified income-based welfare weights.[67] Further, this result can be approximated in the mechanism design case through utility-based SWWs, unlike both the pre-tax and post-tax externality results.

Simply put, a utility inequality externality brings the problem closer to the standard no-externality case. Specifically, the utility problem can often be approximated by changing the inequality aversion of the SWF in the traditional Atkinson [1970] sense.[68] This is because the net effect of the utility inequality externality is to change the social benefit of each individuals' utility, which can be achieved through simply changing the standard SWWs.[69]

Table 3 summarizes these results.

Table 3

Optimal Income Taxation Effects of Various Inequality Externalities

	Mechanical effect	Behavioral responses	Optimal tax rates $\tau(z)$
Post-tax income inequality	✓	✓	$\frac{1+\eta\alpha(z)\epsilon(z)\kappa(z)+\eta\bar{\kappa}(z)-\bar{G}(z)}{1+\eta\alpha(z)\epsilon(z)\kappa(z)+\eta\bar{\kappa}(z)+\alpha(z)\epsilon(z)-\bar{G}(z)}$
Pre-tax income inequality	-	✓ (stronger)	$\frac{1+\eta_{pre}\kappa(z)\alpha(z)\epsilon(z)-\bar{G}(z)}{1+\alpha(z)\epsilon(z)-\bar{G}(z)}$
Utility inequality	✓	-	$\frac{1+\eta_U\cdot\bar{\kappa}(z)-\bar{G}(z)}{1+\alpha(z)\epsilon(z)+\eta_U\cdot\bar{\kappa}(z)-\bar{G}(z)}$

Note: The table describes how each type of inequality externality functions in the optimal income taxation framework.

5. Further Theoretical Discussion

We now turn to the more general implications of an inequality externality. The reframing of inequality as an externality leads to several simple conclusions:

- Equality itself becomes policy-relevant and has an associated shadow price.[70] The trade-

[66] This is more complicated outside the simple quasi-linear case, see Appendix I.H.2.

[67] To the extent that η_U is not an empirical variable, of course. A similar modification can be made to the income-based welfare weights in the post-tax income inequality case. However, there $\bar{G}''(z) = \eta\alpha(z)\epsilon(z)\kappa(z) + \eta\bar{\kappa}(z) - \bar{G}(z)$, indicating that the modified welfare weights are dependent on $\alpha(z)$ and $\epsilon(z)$.

[68] The exception is when separability does not hold such that individuals' behavior is directly affected by the externality.

[69] There is a notable complication to this problem, namely that utility has to be carefully defined. Standard inequality metrics, such as those discussed in the post-tax income case, would not remain the same through monotonic transformations of utility. This complicates the problem both philosophically and analytically. The natural simplification we have used above is a quasi-linear utility function, in which case income changes have a one-to-one relationship with utility changes.

[70] This shadow price corresponds to η in (1.5) and γ in (A.29).

off between income maximization at the bottom and the preferred inequality level becomes relevant.

- Introducing an inequality externality presents an efficiency-based reason for the state to distributionally interfere in otherwise well-functioning markets.

- A Rawlsian min-max is not the most inequality-averse modeling exercise. Similarly, a Utilitarian SWF is not the least inequality-averse modeling exercise if one restricts oneself to non-increasing SWWs.

- A change in marginal tax rates can lead to a "double dividend" of both more revenue *and* an inequality level closer to what is considered optimal, both of which are welfare-relevant.

- The marginal social welfare of income at the top can be negative [Carlsson et al., 2005]. In a utilitarian framework with homogeneous agents and a negative inequality externality, the total welfare effect of additional income at the top is:

$$\frac{d\sum_j g_j U(x_j, \bar{\theta})}{dx_i} = g_i \frac{\partial U(x_i, \bar{\theta})}{\partial x_i} + \sum_j g_j \frac{\partial U(x_j, \bar{\theta})}{\partial \bar{\theta}} \frac{\partial \bar{\theta}}{\partial x_i}$$

  The second term on the right-hand side comes from the inequality externality and can have significant magnitudes, as the results in Section 3 showed. The total effect depends on the relative importance of equality and consumption, a version of the familiar equity-efficiency trade-off.

This last result is particularly notable in the context of concentrated income gains. Extremely concentrated income gains – which are potentially becoming more prevalent with globalization and technical progress – are unambiguously good in standard models. The few agents receiving the additional income increase their utility, while every other agent's utility remains the same. If increased income inequality changes society, however, the other agents may be affected, positively or negatively, despite constant income levels. This is captured by an inequality externality, which illustrates a potential ambiguity in such cases. See Appendix I.B for further discussion.

5.1. Micro-foundations

Generally, very few assumptions are needed for an inequality externality to exist. Several different channels can be directly created from simple and mechanical microfoundations that do not rely on agents' emotional reactions, as we show in the following three simplistic examples:[71]

- Political polarization: Assume that political opinions O_i are a linearly increasing function of individual income x_i and no other factors (for simplicity). Political polarization, denoted as $\bar{P} = \varphi(\boldsymbol{O})$, is defined as an increasing function of a distributional metric φ of all opinions in the population $\boldsymbol{O}$. We assume that $\bar{P}$ enters into the individual's utility function $U_i(x_i, \bar{P}, ...)$. If income inequality increases, differences of opinion within the population mechanically increase as well, generally increasing $\bar{P}$ and affecting $U_i(...)$. Thus,

[71] An overbar indicates a society-wide variable. Bold indicates a population-sized vector.

inequality leads to more pronounced political polarization and subsequent individual utility impacts.[72]

- Innovation / Economic growth: Assume that agents view high inequality as an incentive to work such that l_i and thus x_i are increasing functions of income inequality $\bar{\theta} = I(\boldsymbol{x})$. If so, utility can be written as $U_i(x_i(\bar{\theta}), l_i(\bar{\theta}), ...)$ and inequality is an externality. Further, assume that there exists some societal variable which is positively increasing in total labor supply, such as economic growth rates $\bar{g}$ or innovation levels $\bar{L}$. If this variable has an independent effect on either individual utility $U_i(...)$ or productivity n_i, then income inequality has an additional welfare-relevant externality effect through $\bar{g}$ and/or $\bar{L}$.

- Income-sensitive taste for public goods: Consider the funding required for a public good project to be undertaken as $\widetilde{Q}_j$. Individual utility is defined as $U_i(x_i, \sum_j p_{i,j} q_{i,j}, ...)$, where the individual-specific quantity of public good j is $q_{i,j}$. Assume further that either the quantity $q_{i,j}$ or the taste variable $p_{i,j}$ varies with income levels x_i. As an example, a new youth center may be most beneficial for low-income earners, whereas an expensive opera house could be preferred by high-income earners. If income inequality $\bar{\theta}$ increases, there is less agreement on which public goods to fund and fewer projects reach $\widetilde{Q}_j$. Larger income differences in this context leads to fewer completed public projects and lower individual utility in more unequal societies.

The above examples illustrate that inequality externality channels can be mechanical in nature and can exist under only mild assumptions.[73] We also create micro-foundations for inequality effects on trust, crime, and political capture in Appendix I.I. Before we move on, we note that these channels may imply cascading effects. For instance, increasing political polarization could increase crime rates and hamper economic activity. We present one specific case of such secondary effects;

- Social unrest: Assume that one of the channels discussed above decreases the utility of a subset of individuals. These individuals might then prefer a high fixed cost of social unrest to living in a society with high economic inequality. If these events affect the utility of all individuals, inequality can lead to individual utility losses even for agents who were not initially negatively affected by the inequality externality. On this point we note that high economic inequality commonly precedes notable social uprisings; the French Revolution, the Russian Revolution, and the Arab Spring are some examples.

This last point illustrates that the impacts of inequality externality effects can be starkly discontinuous. In such events the externality itself would have complex optimal policy consequences as a low-probability, high-impact catastrophe event in the vein of Weitzman [2009].

[72] The same argument also holds for diversity of opinions more generally. A different perspective is that increased income inequality could lead to a broader diversity of opinions, carrying a positive utility impact.

[73] Three qualifications should be noted here. First, it is not self-evident which types of inequality (income, wealth, status...) and which domains (neighborhood, country, global...) are relevant, nor which effects are likely to be large on which agents. For this paper we do not go beyond some illustrative calculations in fairly simple cases. Second, the transmission of some inequality effects are clear, such as the effect of inequality on the provision of public goods, while others are dependent on social context or perceived inequality. This implies that inequality effects can differ across societies that are equally unequal. Third, some effects are time-dependent: although not well-captured in single-period models, the basic argument remains the same.

5.2. Consequences in the literature

Given that the inequality externality is harder to ignore than many other externalities, a natural question is how other optimal policy models would be affected by the inclusion of an inequality externality. While this is too large of a question to fully answer in this paper, we present a few thoughts below.

First, our results question the external validity of models which rely on utility functions that only take into account individuals' income and work hours in large-scale settings. This is particularly true for numerical solutions in models focusing on inequality-related issues. As a recent example of how policy discussion can be modified through the introduction of an inequality externality we examine the model in Heathcote et al. [2020], the 2019 *EEA Presidential Address* titled "*How should tax progressivity respond to rising income inequality?*". The work analyzes an optimal taxation model in a general equilibrium framework where the main benefit of higher progressivity is as insurance for idiosyncratic shocks. The authors find that tax progressivity should remain approximately unchanged given rising U.S. inequality levels, a result which is robust in both a Rawlsian and Utilitarian framework. Introducing an inequality externality would likely affect these results. Following our results (which admittedly come from a simpler model), a negative (positive) inequality externality would likely yield a more progressive (regressive) optimal tax rate when income inequality increases. The methodology in Heathcote et al. [2020] is relatively standard, and similar models are common in the economic literature. In general, we believe it would be prudent to check such results for robustness in the face of various inequality externalities or mention the no-externality assumption explicitly.

Second, theoretical models focusing on the trade-offs between different forms of taxation such as Guvenen et al. [2019] and Jacquet and Lehmann [2021] could also be affected by an inequality externality. With an inequality externality the social planner has an added incentive to set the inequality level itself, which may be easier with one instrument or the other. Take the example of wealth taxation versus capital income taxation in Guvenen et al. [2019], where one instrument taxes a stock and the other a flow – if the externality itself is more dependent on either the stock or the flow, the relevant trade-off could be modified.

Third, cost-benefit analysis-type results that depend on income-based SWWs may be fragile to the inclusion of an inequality externality. If an inequality externality is not explicitly taken into account through either modified SWWs or through a cost estimate of income inequality itself, these frameworks implicitly assume that income inequality itself has no effect on society.

6. Conclusion

This paper has introduced the concept of an *inequality externality* and has particularly focused on an *income* inequality externality.

Most standard models of welfarist policy design implicitly assume that income inequality has no societal effects. But as we have shown with microfounded examples, such effects likely exist and could be both numerous and important. They are often independent from individuals' personal feelings; if inequality increases crime, for example, even a selfish individual would prefer equality in the absence of other changes. Including such effects into simple welfarist models with only a combination of diminishing marginal utilities of income and social welfare

weights is generally not possible. The inequality externality is thus intended as a simple and generalizable way to model these side-effects of economic inequality without having to specify the potentially numerous causal channels independently. The concept itself is tractable and does not assume a direction to the externality, can include other-regarding preferences but does not require them, and can easily be extended to other dimensions such as wealth inequality or heterogeneous utility functions.

Introducing an income inequality externality to the welfarist framework leads income (in)equality itself to become a policy goal. Individual labor decisions become socially suboptimal, and the marginal social welfare of individual income can become negative. Frameworks known for only being self-selection problems – including the optimal taxation problem – take on a new externality dimension.

In the Mirrlees [1971] optimal income taxation model, the externality introduces an additional incentive to reduce income inequality. Given that policy makers believe that income inequality itself is concerning, the analysis presented here thus recommends more progressive taxes than those previously suggested by Saez [2001], Piketty et al. [2014], and others. We present two main new insights to the optimal income taxation literature, both of which are relevant for tax design.

First: Optimal top marginal tax rates are largely determined by the magnitude of the inequality externality. We observe both very high top marginal tax rates (above 90%) when inequality is a significant social bad and very low optimal top tax rates (<30%) when inequality is a social good. Our median estimate is an 81% optimal top marginal tax rate. We thus find theoretical support for several policy arguments previously unsupported by economic theory, including a near-maximum income (with a large negative externality) or low top tax rates under a Rawlsian social planner (with a large positive externality). The findings also imply that different beliefs about the magnitude of the inequality externality could be a potential source of political disagreement.

Second: The inequality aversion implied by the current U.S. income tax system is insufficient to explain both progressive social welfare weights *and* a realistic concern for inequality's effects on society. While the tax system may imply a preference for progressive redistribution *or* a negative inequality externality of a substantial magnitude, it is currently not able to accommodate both objectives effectively if designed optimally under our assumptions.

Finally, we briefly discuss how our results could have policy implications beyond optimal income taxation. Given that many economic models rely on the assumption of no externalities, the idea of considering inequality's societal effects as an externality that cannot be captured by standard SWFs could have widespread implications. We encourage further work on the topic.

Chapter 2

The Consequences of Inequality: Beliefs and Redistributive Preferences

The Consequences of Inequality: Beliefs and Redistributive Preferences*

Max Lobeck† and Morten Nyborg Støstad‡

Abstract

What matters for individuals' preferences for redistribution? In this paper we show that consequentialist beliefs about inequality – beliefs about how economic inequality changes the crime rate or the quality of democratic institutions, for example – have a large causal impact on individuals' redistributive preferences. Using two representative surveys of a combined 6,731 U.S. citizens, we show that a majority of respondents believe that inequality leads to a wide range of negative societal outcomes. We establish a causal link from such beliefs to individuals' redistributive preferences by using exogenously provided video information treatments. With this and other methods we show that these "inequality externality beliefs" affect redistributive preferences on the same order of magnitude as broad economic fairness views. These beliefs also have various unique properties when compared to other determinants for redistributive preferences. As such, we discuss whether a focus on inequality's consequences could shape a distinct conversation about redistribution.

*We thank Emmanuel Saez, Marc Fleurbaey, Urs Fischbacher, Stéphane Gauthier, Claudia Senik and Nicolas Jacquemet for invaluable comments and discussions. We have also benefited from suggestions from Regina Anselm, Viola Asri, Luis Bauluz, Marcelo Bergolo, Thomas Blanchet, Antoine Bozio, Stefano DellaVigna, Amory Gethin, Shachar Kariv, Sébastien Laffitte, Clara Martínez-Toledano, Karine Nyborg, Thomas Piketty, Katrin Schmelz, Maj-Britt Sterba, Robert Stüber, Julia Werner, and Irenaeus Wolff. We also thank participants at the 2022 EEA-ESEM Congress, EWMES 2022, the 2022 IIPF Annual Congress, the 2022 LAGV Conference, the 2022 European ESA Meeting, the 2022 ASFEE Conference, Doctorissimes 2022, the 2022 ESA Job Market Sessions, the JEM workshop at the University of Milan, the Fairness and the Moral Mind workshop at the Norwegian School of Economics (NHH), and seminar participants at UC Berkeley, the Paris School of Economics, the University of Konstanz, the Burgundy School of Business and the Thurgau Institute of Economics, for helpful comments and suggestions. This work has been funded by CEPREMAP program 3, the UC Berkeley James M. and Cathleen D. Stone Center on Wealth and Income Inequality, two French government subsidies managed by the Agence Nationale de la Recherche under the framework of the Investissements d'avenir programme reference ANR-17-EURE-001, and the Deutsche Forschungsgemeinschaft (DFG - German Research Foundation) under Germany's Excellence Strategy - EXC-2035/1 - 390681379. This study was pre-registered under AsPredicted #82083 and #104271. Version: November 30, 2023.

†University of Konstanz, Universitätsstrasse 10, 78464 Konstanz, Germany, Cluster of Excellence "The Politics of Inequality", Thurgau Institute of Economics, e-mail: max.lobeck@uni-konstanz.de

‡Paris School of Economics, 48 Boulevard Jourdan, 75014 Paris, France. Phone: +33766142152. e-mail: morten.stostad@psemail.eu.

1. Introduction

Understanding what drives people's willingness to redistribute is an empirical question with far-reaching implications for economic policy and social equity. A large academic literature has explored the question, proposing various determinants such as individual income maximization, economic fairness concerns, and economic inefficiency considerations [Cappelen et al., 2007, Durante et al., 2014, Stantcheva, 2021]. In this paper we quantify a previously unexplored motive within this literature, namely individuals' beliefs about the *consequences of inequality*. Such consequences occur when economic inequality affects something that in turn affects individuals, for example the level of social unrest, the economic growth rate, or the general trust between people. Importantly for the present paper, people's beliefs about these consequences can vary, which can in turn affect overall demands for redistribution. If there is a societal consensus that large economic differences lead to violent revolutions, for example, then a consensus for redistributive policy may be achieved simply by highlighting these risks.

We call these consequences *inequality externalities*, based on the idea that they occur as a side effect of the economic inequality that all of us contribute to through market actions [Støstad and Cowell, 2021].[1] Beliefs in these ideas can affect people's willingness to redistribute without requiring any altruism or other-regarding preferences, as even selfish people can be concerned about inequality's consequences. It follows that beliefs about inequality's consequences may have unique implications for the redistributive debate. In particular, varying *inequality externality beliefs* could affect distributional equilibria across societies.

Perhaps surprisingly, then, individuals' beliefs about these consequences have barely been elicited in survey contexts. This contrasts to the vast amounts of information we have about individuals' other attitudes to inequality. We only know of two prior externality-focused questions in the United States, the most relevant of which comes from the General Social Survey (GSS). Respondents were asked whether they agreed that "*large differences in income are necessary for America's prosperity*" over five survey waves between 1987 and 2021. The share who agreed with the statement steadily fell from 34% in 1987 to 12% in 2021.[2] Beyond this, very little has been uncovered about individuals' inequality externality beliefs in either the United States or the wider world. As such, any empirical connection to individuals' preferences for redistribution has also not been explored.

Following these observations, this paper poses two main questions. First, do U.S. citizens expect economic inequality to change society – and if so, how? Second, to what extent do such beliefs causally impact citizens' redistributive preferences? To answer these questions we conduct two representative surveys of the U.S. population, sampling a total of 4,371 and 2,360 distinct U.S. citizens with the professional survey providers Lucid and Dynata. These two surveys allow us to create the first comprehensive data sets of U.S. citizens' inequality externality beliefs. We explore the link between these beliefs and redistributive preferences using

[1] Note that both inequality itself and the resulting outcomes can be defined as externalities.

[2] The trend is monotonically decreasing over five waves of the GSS (which contrasts to relatively flat trends for economic fairness-related questions). The second question is from the 1991 wave of the International Social Justice Project (ISJP), which asked respondents whether they agreed to the statement that "*there is an incentive for individual effort only if differences in income are large enough*". 63% of respondents agreed. See Appendix II.A for more details and other countries.

several methods, the most important of which is a video-based information experiment. This information experiment is designed to isolate the causal effect of inequality externality beliefs on redistributive preferences. Finally, we discuss the structural differences between inequality externality beliefs and other determinants of redistributive preferences, focusing specifically on economic fairness views as a contrast.

Our results show that most U.S. citizens believe in the *negative* externality dimension. This contrasts to the way these issues have often been discussed; the GSS question mentioned earlier assumes a positive externality dimension, for example. Nearly every individual ($\sim 90\%$) believes inequality affects society in some way, and a consistent majority ($\sim 60\%$) believes that economic inequality has overall harmful societal consequences.[3] We delve into the potential reasons and find strong beliefs in specific channels; 74% of respondents think more economic inequality increases the amount of crime, for example, and 67% think it deteriorates the overall level of societal trust. These results extend into the economic dimension. Whereas only 23% think more economic inequality *in*creases the amount of economic growth (reminiscent of the GSS question), 52% think the converse, namely that more economic inequality *de*creases the amount of economic growth. Respondents have similarly negative beliefs about the effect of economic inequality on the prevalence of social unrest, the amount of innovation, the quality of democratic institutions, and more. We find a striking consensus across demographic groups and political affiliations; Democrat- and Republican-leaning voters are both more likely to believe more inequality leads to *less* and not *more* economic growth, for example. This represents our first main finding; most U.S. citizens believe that economic inequality has a wide range of negative societal consequences.[4]

Having established the existence of widespread inequality externality beliefs in the U.S. population we move to the implications of these beliefs. We first test whether such beliefs constitute a causal determinant for redistributive preferences. To establish this connection we use an information experiment which aims to shift individuals' inequality externality beliefs through short, exogenously provided video treatments. Several novel design choices are made to reduce survey demand and priming effects.[5] We find that shifting individuals' inequality externality beliefs has a strong causal effects on redistributive preferences ($p < 0.01$). This result is robust to an array of different specifications, and first-stage effects and mediation analysis indicate that the treatment mechanism is as expected (learning about the intended belief) with limited spillovers and priming. As such, we establish that inequality externality beliefs are a causal determinant of redistributive preferences.

To understand whether inequality externality beliefs are meaningful in the wider redistributive debate we then estimate the *size* of this determinant. We primarily do this through comparisons with broad economic fairness views, which are a well-known powerful determinant of redistributive preferences [Cappelen et al., 2007, Durante et al., 2014, Almås et al., 2020]. We

[3] $\sim 10-15\%$ believe the net effect to be positive, $\sim 15-20\%$ believe the positive and negative effects "cancel each other out", and $\sim 5-15\%$ do not believe inequality affects society.

[4] Results are robust to different methodologies, question phrasings, and are nearly identical across different representative samples. Substituting the word "inequality" for "equality" or "differences in income and wealth" does not change overall results.

[5] We introduce the concept of *dual control groups*, which indicates using both an active and a passive control group and merging them on pre-specified criteria to reduce priming and attention effects. We also use what we call a *secondary survey*, which is a structural, well-explained gap between the treatment and outcomes of interest with the intention of reducing experimenter demand and respondent confusion.

use three distinct methods to compare these two determinants, also touching on other potential redistributive determinants where possible (most notably taxation-related efficiency concerns). First, we compare treatment effects from comparable video information treatments. Second, we directly elicit respondents' beliefs about what drives their redistributive preferences. Third, we explore the predictive power of each determinant on redistributive preferences. All three methods indicate that inequality externality beliefs are a strong driving force behind individuals' preferences for redistribution. Inequality externality beliefs consistently approach broad economic fairness views in importance; they also clearly outperform taxation-related efficiency concerns. As far as we know, income maximization is the only other redistributive motive with similar efficacy. This represents our second main finding; inequality externality beliefs are a formidable causal determinant of redistributive preferences.

We then explore how inequality externality beliefs *structurally* differ from existing redistributive determinants. If individuals see inequality externalities as "just another way to talk about inequality", the wider implications of our findings may be limited. If instead these beliefs are structurally distinct with unique properties, they may have large consequences for the redistributive debate.

We find significant evidence for the latter. First, inequality externality beliefs are particularly impactful for high-income individuals. Individuals with an annual income above $100,000 are more likely than lower-income individuals to change their redistributive preferences from new inequality externality information. We find the opposite result for new information about economic fairness, where low-income individuals react more heavily. High-income individuals are also more likely to hold negative inequality externality beliefs than to believe that the economic system is unfair; a strong income gradient in beliefs about the fairness of the economic system [previously discovered by e.g. Hvidberg et al., 2022] is almost non-existent for inequality externality beliefs. We hypothesize that this is because inequality externalities represent a uniquely self-interested motive for individuals with high economic status to redistribute.

Second, the fairness-based information experiment is significantly more likely to make respondents feel *anger* than the externality-based treatments. This hints at how respondents react differently to these two concepts, potentially because fairness-adjacent information (about the evolution and distribution of incomes, for instance) is seen as more normatively based than information about externality-adjacent information (e.g. cross-country correlations of inequality and various outcomes). This indicates that a redistributive debate focused on inequality externalities could have less affective polarization. Third, historical evidence suggests differences in malleability. As evidenced by the GSS question, inequality externality beliefs in the U.S. have changed significantly since 1987. This differs to economic fairness views on whether the income distribution is unfair or whether hard work or luck is more important in becoming rich, which have stayed constant on average despite rising economic inequality. Fourth, fairness views are more polarized across political parties than externality beliefs, and fifth, inequality externality beliefs explain variation in redistributive preferences that we are unable to explain through other redistributive determinants. We summarize the above in our third main finding; inequality externality beliefs are structurally distinct from other redistributive determinants.

Put together, our results hint at potentially large implications for the overall debate on in-

equality reduction. Cross-country variation in redistributive equilibria has often been explained by varying philosophical ideals [Almås et al., 2020], differing racial heterogeneity or immigration [Alesina et al., 2001, 2023], or a potential lack of governmental trust [Kuziemko et al., 2015]. We suggest that the extent to which societies have been concerned about the societal consequences of economic inequality may also have a significant impact. Strong society-wide beliefs in the negative externalities of inequality could (eventually) lead to a redistributive consensus that is based on reducing shared costs across income brackets. Such a consensus could reduce conflicts of interest, be achieved without strong appeals to normative ideals, and, in the long term, lead to a more stable low-inequality equilibrium.

This paper is to the best of our knowledge among the first to explicitly study the idea of inequality externality beliefs, and thus also the first to directly empirically link stated externalities beliefs to individuals' preferences for redistribution. An extensive literature has examined various other determinants of redistributive preferences, in particular people's fairness ideals and their concerns about the efficiency costs of redistribution [e.g. Cappelen et al., 2007, Almås et al., 2020]. Of the two, fairness ideals are often found to be the stronger motivator [Durante et al., 2014], although there is some variation across various groups within the population [Fisman et al., 2015]. Papers have also explored the connection between redistributive preferences and beliefs about one's relative position [Cruces et al., 2013, Karadja et al., 2017], information about the level of inequality and the functioning of tax systems [Kuziemko et al., 2015, Stantcheva, 2021], and beliefs about social mobility [Alesina et al., 2018b, Gärtner et al., 2019], among many other topics. Citizens' concerns about the consequences of inequality are rarely discussed in this broad literature, despite having been proposed as a possible motive behind redistributive preferences [Alesina and Giuliano, 2011]. One exception is work by Rueda and Stegmueller [2016] who present correlational evidence of an association between the fear of crime and preferences for redistribution among high-income individuals in Western Europe, and explain the association through an externality-based theoretical argument. As a final link to the redistributive preference literature, we note that the consensus we find in inequality externality beliefs is reminiscent of the across-party consensus Norton and Ariely [2011] find for a reduced level of wealth inequality in the no-friction case.

Our work creates a survey-based background to the vast literature attempting to establish connections between economic inequality and various societal outcomes. In short, there exists correlational evidence indicating that inequality is an externality across various dimensions [Wilkinson and Pickett, 2011, Rufrancos et al., 2013, Bergh et al., 2016], but large-scale causal evidence is unlikely to be forthcoming due to the lack of exogenous variation of economic inequality.[6] In smaller settings, causal evidence can exist; economic inequality has been convincingly shown to affect subjective well-being [Card et al., 2012] and productivity [Breza et al., 2018] in the workplace through relative income concerns, and trust in laboratory and survey experiments [Gallego, 2016, Fehr et al., 2020b]. A full examination of this literature is beyond the scope of this paper.

We also connect to the theoretical literature on inequality as an externality. This relatively small literature [Thurow, 1971, Alesina and Giuliano, 2011, Støstad and Cowell, 2021] explores

[6] As well as other intrinsic concerns and insufficient data – see Støstad [2019] for a discussion.

optimal policy given that inequality's societal consequences are a concern for individuals or the social planner. In showing the widespread public beliefs in such effects, we give credibility to this assumption and thus the resulting (large) policy conclusions. In general, our results indicate that it might be prudent to more seriously consider the robustness – or fragility – of standard individualist frameworks to inequality externality effects.

The rest of the paper is organized as follows. Section 2 describes the theoretical framework behind the analysis. Section 3 presents the survey sampling methodology. Section 4 analyzes individuals' inequality externality beliefs, while Section 5 extends the analysis of the results to the redistributive preference dimension. Section 6 discusses structural differences between inequality externality beliefs and other redistributive determinants, and Section 7 concludes.

2. Theoretical Framework

The central idea behind this work is that economic inequality itself can affect society through various channels. We define an inequality externality as some factor that is potentially impacted by economic inequality, such as crime, social unrest, or economic growth. The externality framing is motivated by Støstad and Cowell [2021], who point out that economic inequality itself is an externality if it affects any other outcome that enters individuals' utility functions.[7]

Exploring every potential causal channel through which economic inequality could affect society is beyond the scope of this paper. Still, it is useful to note two points briefly. First, the majority of the externality channels we focus on could be caused by several different mechanisms. Second, inequality externalities are relatively simple to micro-found and can be *mechanical* in nature; in other words, they do not need to depend on perceived inequality. Assuming that incomes causally affect political opinions in a monotonic manner can be enough to micro-found an effect of income inequality on political polarization, for example. We discuss these two points further in Appendix II.B.1.

Preferences for redistribution Could beliefs about inequality externalities affect individuals' willingness to redistributive? To structure the discussion we create a stylized framework of individuals' redistributive preferences.

Suppose economic inequality θ affects various outcomes such as the general trust between people, the rate of innovation, or the quality of democratic institutions. Suppose further that the magnitude of the effect from outcome j can be denoted by α_j. For example, the quality of democratic institutions D could be directly affected by economic differences such that $D = D_0 + \alpha_D\theta$ where D_0 is anything else that affects the quality of democratic institutions, θ is economic inequality, and α_D is the magnitude of the inequality externality effect (potentially

[7]There are two ways to frame this. We will largely discuss the outcomes themselves as "inequality externalities", whereas Støstad and Cowell [2021] discusses inequality itself as an externality. Both descriptions are formally correct. Buchanan and Stubblebine [1962] defines an externality as present when $u^A = u^A(..., Y)$ and Y is under control by another individual. Inequality itself is an externality if it affects pertinent societal outcomes, i.e. the outcomes in our utility functions, as inequality itself is by definition determined by others. The outcomes themselves, e.g. crime, are also determined by others – this time *through* inequality – and are also in individuals' utility functions.

zero).[8] This implies that we have $\frac{\partial D}{\partial \theta} = \alpha_D$.[9] In other words, the sign of α_D denotes whether more economic inequality improves or worsens the quality of democratic institutions.

We write a simple model of individual i's redistributive preferences as:

$$\cup_i = x_i - \sum_j \gamma_{ij} \mathbb{E}_i(\alpha_j)\theta + \Upsilon_i. \tag{2.1}$$

Here x_i represents individual income, $\sum_j \gamma_{ij} \mathbb{E}_i(\alpha_j)\theta$ represents the net effect of any inequality externalities, and Υ_i represents the effect of any other redistributive determinants. The net inequality externality impact is the sum over all externality channels j of the inequality metric θ, individuals' expected belief of the true causal effect α_j of this type of inequality on outcome j, and a preference-term γ_{ij} which denotes the willingness to redistribute to affect this outcome.

Equation 2.1 shows a stylized way in which beliefs in inequality externalities could affect individuals' preferences for redistribution. It also illustrates the three key factors we will explore in the remainder of the work.

First, what are individuals' *inequality externality beliefs*? This is represented by $\mathbb{E}_i(\alpha_j)$ in (2.1). As the true effect of economic inequality on society is unknown,[10] individuals *beliefs* about α_j are crucial. These externality beliefs could be positive or negative; individuals may believe inequality increases or decreases the amount of economic growth, for example. We explore these beliefs in Section 4.

Second, do any such inequality externality beliefs affect redistributive preferences? This is represented by γ_{ij} in (2.1). This connection is simple in theoretical welfare frameworks, where adjusting for existing inequality externalities implies efficiency gains that the social planner takes into account (assuming there is some effect of the externalities on individuals' well-being). We show this theoretical connection through an optimal income taxation model in Appendix II.B.2. For individuals' preferences, however, the link is not necessarily as straightforward. Individuals may believe that inequality negatively affects society while also preferring non-redistributive solutions. Suppose that inequality affects crime; individuals may prefer to solve this through crime prevention rather than redistribution, for example.[11] Individuals' willingness to pay to affect externalities could also depend on incomes or simply be heterogeneous.[12] Even if some redistribution is preferred, then, the *magnitude* of this redistribution may be strongly heterogeneous across individuals which in turn affects redistributive preferences. γ_{ij} captures such potential heterogeneity. We explore the link between inequality externality beliefs and redistributive preferences in Section 5.

Third, how do inequality externality beliefs compare to other redistributive determinants? These other redistributive determinants are represented through Υ_i in (2.1). This term indicates

[8] We abstract away from interactions between D_0 and θ for simplicity. Although we also generally abstract away from interactions between θ and α_j, Section 4 shows that individuals' beliefs about α generally do not depend on θ.

[9] Although different outcomes are most likely affected by different types of economic inequality in practice, we will abstract from this for simplicity and consider θ as some combination of income and wealth inequality. We explore individuals' beliefs on what type of inequality matters in Section 4.

[10] The lack of exogeneous macroeconomic variation of inequality presents severe identification issues.

[11] In the social planner case, some additional redistribution is still preferred as the cost of the crime prevention program creates an incentive to reduce inequality.

[12] Some individuals may care much more about their own income than society-wide political polarization, for example. Altruistic individuals may also care about how inequality externalities affect others than themselves.

anything else the individual might care about when making redistributive decisions, for example philosophical principles, other-regarding preferences, or efficiency concerns. These other determinants may or may not function similarly to inequality externalities. One could imagine fairness concerns being modeled as $\gamma_{i,fair}\mathbb{E}_i(\alpha_{fair})\theta$, for example, where a natural question is whether the same individuals have $\gamma_{i,fair}\mathbb{E}_i(\alpha_{fair}) \neq 0$ and $\gamma_{i,ext}\mathbb{E}_i(\alpha_{ext}) \neq 0$. If Υ_i is very similar to the inequality externality term in both function and who is affected – say inequality externalities are interpreted as "just another way to talk about inequality", for example – the practical consequences of inequality's externality effects may be limited. If this is not true, and inequality externalities have unique properties in both who is affected and how they are affected, there may be large implications for the redistributive debate. We explore this topic in Section 6.

These three points illustrate the three main questions of our paper. What are U.S. citizens' inequality externality beliefs? Do these beliefs affect redistributive preferences? And are inequality externality beliefs structurally distinct from other redistributive determinants? The empirical portion of the paper will answer these questions in turn.

Before moving to the empirical analysis we make one additional observation. The redistributive preferences as set forth in (2.1) immediately imply a non-altruistic motive to have preferences for economic equality. This in turn implies a potential income-based heterogeneity in the importance of inequality externality beliefs. Suppose that a share of individuals only care about their own outcomes; a simple way to model this is to set $\Upsilon_i = 0$. If we also assume no inequality externality effects, these individuals simply maximize their income. A redistributive policy would thus be supported by individuals with low x_i and opposed by individuals with high x_i [reminiscent of Meltzer and Richard, 1981]. When individuals believe in negative inequality externality effects, $\sum_j \gamma_{ij}\mathbb{E}_i(\alpha_j)\theta$ becomes relevant and at least some individuals with high x_i may also support redistributive policy (to reduce the relevant inequality externalities). This is in contrast to the self-interested individuals with low x_i, who already supported redistribution and thus do not change their preferences. This heterogeneous reaction of self-interested individuals across the income spectrum leads us to hypothesize that high-income individuals are more likely to be affected by a shift in externality beliefs than low-income individuals. We return to this in Sections 5 and 6.

We now move to the empirical section of the paper.

3. Sampling Methodology

Our empirical results are based on two independent pre-specified surveys.[13] We will call these surveys Survey 1 ($N_1 = 4,371$) and Survey 2 ($N_2 = 2,360$). Survey 2 is a follow-up survey to Survey 1.

Survey 1 had two primary goals. First, to collect descriptive data on individuals' inequality externality beliefs. Second, to conduct an information experiment to connect these beliefs to individuals' redistributive preferences.

Survey 2, which was designed after receiving results from Survey 1, aimed solely to re-measure and improve the robustness of the descriptive findings from Survey 1. The survey populations

[13] See AsPredicted.org #82083 and #104271.

are distinct, and Survey 2 has no connection to the information experiment in Survey 1.

Where possible we merge the results from the two surveys. In practice, this means that we merge the control group of Survey 1 and all of Survey 2 for most of our descriptive results.[14] This merged descriptive sample has a total of 3,292 respondents.

Survey respondents were collected through the survey providers Lucid (Survey 1) and Dynata (Survey 2). Both Lucid and Dynata are commonly used by economic researchers [see e.g. Haaland and Roth, 2021, Andre et al., 2022].[15]

3.1. Survey 1 (Main survey)

Data for Survey 1 were collected between December 6th and 24th 2021 through the survey provider Lucid. 5,007 subjects completed the survey, which is reduced to 4,371 after routine data quality checks. The average survey duration for these respondents was 19 minutes and 11 seconds. Methodological details and the full questionnaire can be found in Appendix II.C.1.

3.2. Survey 2 (Follow-up)

Data for Survey 2 were collected between August 9th and October 8th 2022 through the survey provider Dynata. 2,479 subjects completed the survey, which is reduced to 2,360 after routine data quality checks. Survey 2 had the goal of further delving into the main *descriptive* results from Survey 1 with an independent sample from a distinct survey provider. The average survey duration for these respondents was 20 minutes and 31 seconds. Methodological details and the full questionnaire can be found in Appendix II.C.2.

3.3. Respondent characteristics

In both surveys we used quotas to aim for representativity along the dimensions of age, gender, geographical region and political affiliation (Democrat, Independent and Republican). These dimensions are therefore largely representative of the 2021 U.S. population by design. We also have a wide range of incomes in both surveys. As with other studies using online access panels, both surveys somewhat oversample white respondents and college-educated respondents.[16] Re-weighting respondents for full representativity on these dimensions does not change reported results significantly. We discuss sample representativity further in Appendix II.C.3.

4. Inequality Externality Beliefs

The first main objective of this paper is to explore U.S. citizens' beliefs in the consequences of economic inequality. Such beliefs, modeled by $\mathbb{E}_i(\alpha_j)$ in (2.1), are a prerequisite for any subsequent effect on redistributive preferences. They are also arguably intriguing in themselves. Beliefs about inequality are a central theme in behavioral economics; by creating the first database on individuals' inequality externality beliefs we contribute a novel perspective to this widely

[14] This does not significantly affect our results and is done to improve precision. We pre-specified that results from each survey would be shown side-by-side in the text where possible; as results are very similar across surveys, we instead merge the samples and show the side-by-side results in the Appendix.

[15] Lucid and Dynata both collect respondents from several distinct sources. These sources are partly chosen to collect a wide variety of respondents to ensure representativity in surveys like ours. They include brand loyalty programs, targeted online advertisements, and institutional partnerships.

[16] These disparities are typical for similar studies, see e.g. Stantcheva [2021].

explored topic. The subsequent results could improve our understanding of the complex social and economic dynamics surrounding economic inequality.

Prior to delving into the results it is important to acknowledge the limitations to our approach. We do not aim to find precise empirical estimates of $\mathbb{E}_i(\alpha_j)$, as this is likely too cognitively demanding for respondents. Instead we largely elicit the direction of $\mathbb{E}_i(\alpha_j)$ – whether more inequality leads to more or less economic growth, for example – which we couple with broad questions about how meaningful these channels are. We focus on *economic* inequality loosely defined as "differences in income and wealth", and thus abstract from other relevant dimensions such as gender inequality, racial inequality, and whether income or wealth inequality is more meaningful. We also abstract from the origin of the inequality, the difference between perceived and actual inequality, and differences between meritocratic and non-meritocratic inequality.

4.1. How does economic inequality change society?

Design Our main battery of inequality externality questions asks how respondents think inequality affects different aspects of society. We elicit opinions on whether inequality affects crime, corruption, political polarization, unemployment, innovation, economic growth, the quality of local public goods such as schools or libraries, people's overall quality of life (comparing people with the same income in more or less unequal societies), the quality of democratic institutions, and generalized trust.[17] The standard question asks: "*How does more economic inequality change the [amount of crime / overall level of trust / ...] in a country?*". In general, random noise will only bias results towards a zero net effect.[18] We use different phrasings for a large subset of respondents (avoiding using the word "inequality"); we discuss this and other extensions in Section 4.3.

Results Figure 1 characterizes the responses to these questions from the merged descriptive sample.[19] The immediate takeaway from this exercise is that respondents are most likely to believe inequality has a negative societal consequence for any outcome we survey – including economic outcomes such as economic growth and innovation. The data differences between the negative and positive outcomes come entirely from respondents' beliefs, as the division in Figure 1 is purely for visualization purposes.[20] As far as we know this represents the first systematic exploration of these beliefs in the American public. We will now discuss specific inequality externality beliefs.

First, there is a strong belief that economic inequality increases crime, which is a canonical inequality externality studied in previous research [Becker, 1968, Kelly, 2000, Fajnzylber et al., 2002]. 74% of respondents believe more economic inequality increases the amount of crime. This is the most agreement we find on inequality's effect on any specific variable. Similar but somewhat smaller figures are found for the percentage of respondents believing inequality increases the negative outcomes of social unrest (70%), corruption (66%), and political polar-

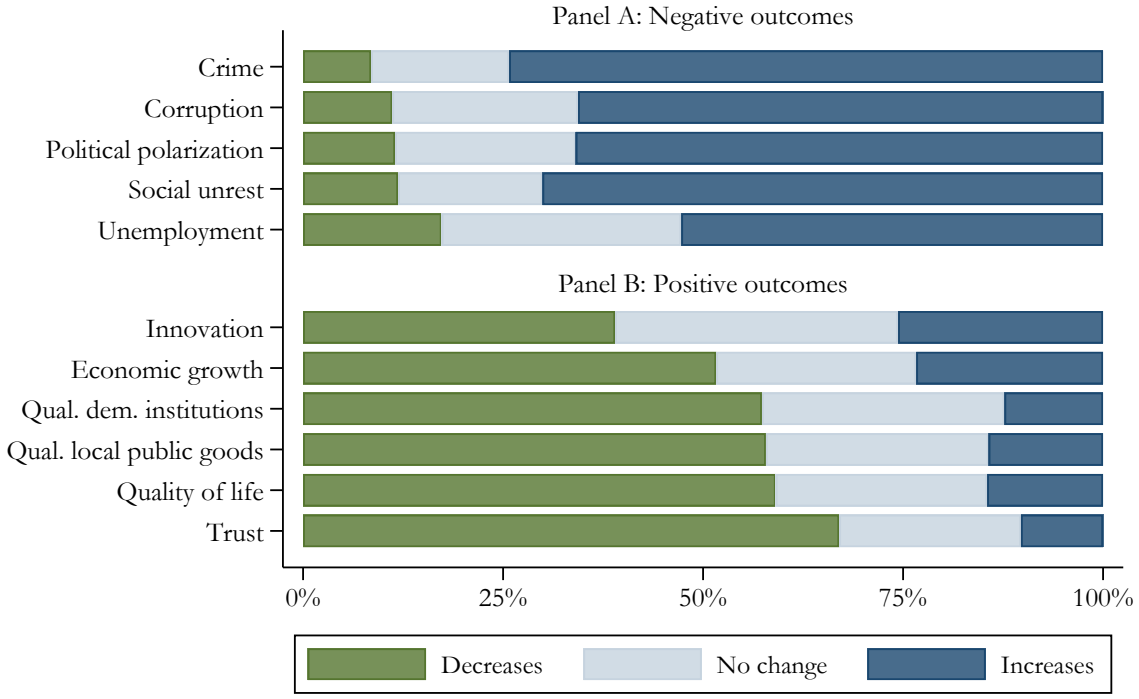
[17] We selected outcomes that have previously appeared in public and academic discussion about inequality's societal effects. All these outcomes were elicited in Survey 1; in Survey 2 we did not elicit unemployment, the quality of local public goods, and overall quality of life. We also generally added a short definition to the outcome in question; see Tables I11-I12 for these definitions.

[18] Questions were always symmetric around a neutral option and answer order was randomly flipped.

[19] The exact numbers are shown in Table I1. For data from Survey 1 or 2, see Figures and H2.

[20] Respondents were not given any indication on which outcomes were positive or negative in the survey.

Figure 1: Respondents' Beliefs About How More Economic Inequality Changes Society



Note. Answers to questions about how "more economic inequality" changes the designated outcomes. Full question example: "How does more economic inequality change the amount of crime in a country?" Answer option example: "More inequality → a lot more crime". The green (left) bars indicate the share of respondents that believe that inequality decreases the outcome in question, while the blue (right) bars indicate the opposite. Questions are ordered according to which portion of respondents believe that inequality decreases the variable. Answers are pooled from Survey 1 and Survey 2 ($N \in \{2990, 3292\}$), except for unemployment, quality of local public goods, and quality of life, which were only asked in Survey 1 ($N \in \{628, 643\}$). For only Survey 1 or 2, see Figures H1 and H2 respectively. The exact numbers are shown in Table I1.

ization (66%). A majority of respondents seem to believe that more unequal societies are less stable and law-abiding in general.

We also elicit individuals' beliefs on how inequality affects positive outcomes such as generalized trust or the quality of democratic institutions. Generalized trust presents the most agreement; 67% believe inequality decreases the overall level of trust in a country. Then follows quality of life, where we specifically ask respondents to compare between people with the same income in more equal or unequal societies. Under this definition, 59% believe inequality worsens quality of life generally speaking – more strong evidence that individuals believe inequality itself is a negative externality.

A clear majority believes inequality deteriorates the functioning of the collective parts of society, as observed through the number of respondents who believe inequality decreases the quality of local public goods (58%) and the quality of democratic institutions (57%). We note that it is simple to rationalize how inequality could theoretically improve these outcomes; more economic inequality could lead to more funding for local public goods, for example. These are not beliefs most U.S. citizens subscribe to. The percentage who believe inequality improves the quality of local public goods or the quality of democratic institutions is just 14% and 12% respectively.

The three last outcomes we elicit are on inequality's effect on economic growth, innovation, and unemployment. Inequality's effects on economic performance are the most ambiguous from

the variables we survey.[21] On one hand, one could argue that inequality promotes growth by strengthening incentives. On the other, one could argue that inequality reduces economic performance through aggregate demand, poverty traps, or the many potential negative effects we already discussed – on trust, criminal activity, democratic institutions, and so on.

Between these two arguments, U.S. citizens' beliefs clearly point towards the latter. A majority of respondents believe that inequality generally increases unemployment (53%) and reduces growth (51%). Somewhat less than a majority also believe that inequality decreases innovation (40%). The converse for these three data points – that inequality *decreases* unemployment and *increases* growth and innovation – is only believed by 17%, 23% and 26% respectively.[22]

We also examine these beliefs in a variety of other ways. When asked to choose which externality channels "matter the most", respondents indicate that crime and corruption are the most important negative externalities and economic growth and innovation are the most important positive externalities (although few respondents subscribe to the positive externality dimension in general). When asked whether a given externality channel is "meaningful", the quality of democratic institutions sees the highest consensus.[23] Answers also indicate that respondents believe these issues are important; a majority of respondents believe any given externality is "generally meaningful" (30%) or "very meaningful" (32%). We discuss these results on the size of the specific externality channels further in Appendix II.D.1. We also shortly move on to what respondents think about inequality's externality effects *overall*.

In Appendix II.D.2 we examine what type of economic inequality matters for these responses – so which θ is impactful in (2.1). Generally, most respondents believe that both bottom-based (the amount of relatively poor) and top-based economic inequality (the amount of relatively rich) is impactful. Still, bottom-based inequality appears more important than top-based inequality to respondents across outcomes (except in the case of corruption). Further, most respondents indicate that the same causal channel holds regardless of the initial level of inequality, which we discuss in Appendix II.D.3.

Finally, we note that percentage of respondents who believe that economic inequality does not affect society in any way is consistently low. Only 4.2% of respondents in Survey 1 and 3.7% of respondents in Survey 2 consistently chose "No change" to all questions they were asked. In other words, nearly every individual has *some* inequality externality belief.

4.2. What is the overall effect of inequality on society?

How do these specific beliefs translate into an overall view of inequality's consequences? In Figure H14 we show how respondents believe more economic inequality *generally* changes society.

[21] While one could conceivably argue that inequality has a positive effect on the other outcomes we study, e.g. crime or trust – say that inequality increases trust through increased segregation, for instance – the academic literatures on these outcomes have typically highlighted inequality's negative effects [see e.g. Wilkinson and Pickett, 2011] or argued that there is no such effect [see e.g. Hastings, 2018].

[22] Before moving on we note that some caution should be taken in interpreting these results, as respondents may confuse the effects of inequality itself on these factors with the effects coming from a redistributive tax system (classic incentive effects). Open-ended text questions indicate that such confusion is very rare, however. In a sample of 226 text responses on inequality's effect on economic growth, only one answer clearly confuses these issues. In general, answers discussing incentive issues generally focus on the incentive effects of inequality itself, e.g. *"Income inequality creates competition, which creates the incentive and motivation to improve oneself."*

[23] Results are otherwise similar. The four most meaningful channels are considered to be the quality of democratic institutions (70%), crime (67%), trust (66%), and corruption (65%).

Answer options range from an overall positive fashion to an overall negative fashion.[24]

62% of respondents believe that inequality changes society somewhat or a lot for the worse, or constitutes a negative externality overall. Only a small minority (12%) states that inequality has positive societal effects, or constitutes an overall positive externality. 26% of respondents believe that there is no net effect of inequality on society; roughly half of these respondents believe that inequality has no effect on society at all.[25] This indicates that $\sim 87\%$ of respondents believe in some sort of inequality externality in this setting; while a slightly more conservative estimate than from the specific externality questions,[26] the conclusion remains that the vast majority of U.S. citizens have some sort of inequality externality belief.[27]

We elicit three further general externality beliefs. Respondents are asked (i) whether unequal countries generally function worse,[28] (ii) whether they believe that inequality changing society for the worse through externality channels is a "very serious issue",[29] and (iii) whether "extremely high inequality levels would significantly increase the chances of a societal collapse". A majority of respondents answer that more unequal countries function worse (60%) and that inequality changing society through externality channels is a "serious" (29%) or "extremely serious" (22%) issue.[30] A large majority of respondents also answer "Yes, definitely" (25%) or "Yes, maybe" (47%) to whether extremely high inequality would significantly increase the chances of a societal collapse.

4.3. Robustness of externality beliefs

The specific inequality externality beliefs we show in Figure 1 are very robust to different specifications. This is illustrated in the Appendix Figures H3-H4, where we show that results stay very similar when we (i) weight respondents for full representativity to the 2021 U.S. population, (ii) restrict to only the distinct Survey 1 or Survey 2 samples, (iii) change the words "more inequality" to "larger differences in income and wealth", (iv) explain what "more inequality" and the initial reference point of inequality is through diagrams and words,[31] (v) inform respondents

[24]Note that this question was posed *before* the specific externality questions detailed previously. The design of these general questions, which varied across surveys in detail and accompanying explanation, are detailed in Appendix II.C.5. The main conclusions are similar across surveys despite different design choices. Survey 1 did not give respondents examples of potential channels to avoid bias. Survey 2 had a clearer definition of "changes society", including the specific examples of "*economic growth, crime, general trust, innovation, the quality of democratic institutions, and so on.*" In both surveys, a separate question specifies whether individuals believe inequality affects society *at all*. The accompanying data is shown in Table C2.

[25]The other half believes that good and bad effects cancel each other out. Note that this divide is the only aspect of our results that had significantly different responses across surveys, as we show in Table C2 and discuss in Appendix II.C.5. The difference is most likely from a question-specific ordering effect.

[26]We propose two potential reasons. First, the general question was posed before the specific questions in both surveys to avoid priming; individuals may have been reminded of an externality channel when the outcome was directly elicited. Second, measurement error would bias both values in opposite ways, essentially creating lower and upper bounds for the true value.

[27]The overall results do not change depending on whether we use different phrasing ("inequality", "equality" or "differences in income and wealth") – see Appendix II.D.5.

[28]Full question text: "How much do you agree with the following statement? Countries with more economic inequality usually function worse."

[29]Full question text: "Overall, do you think economic inequality changing society for the worse through one or more of the channels we discussed earlier - for example through increased crime / social unrest / corruption, or through decreased social cohesion - is a very serious issue?"

[30]42% answer that this is "not a very large" or "small" issue, and 7% do not believe it is an issue at all.

[31]Survey 1 does not explicitly explain what "more inequality" means (to keep question brevity). Most questions in Survey 2 had a reference point where "more inequality" was explained as the shift from a society with "*a large middle class and few with relatively small or large incomes, [where] the richest tenth of society earns 5 times*

that their answer is important and ask them to carefully consider their choice, while imposing a waiting period before an answer is possible, (vi) restrict the sample to respondents who succeeded on every attention check, and (vii) restrict the sample to respondents who answered a simple comprehension question on inequality correctly. Further, 98% of respondents confirm their choice when prompted and question ordering does not seem to have a noticeable effect on results. Open-ended text questions confirm that respondents understand the question topic and can rationalize their answers.[32]

The robustness check that has the largest effect is to change the phrasing from "more inequality" to "more equality" ($\sim$ 12% difference).[33] Respondents still largely believe in the negative externality dimension under this phrasing. We discuss how phrasing affects both specific and general externality beliefs in Appendix II.D.5.

Finally, we note that a placebo question ("*How do you think more economic inequality changes the number of daylight hours?*") posed in Survey 2 had a nearly perfectly symmetric answer distribution around the 89% of respondents who answered "No change", and that 93% and 94% of our final Survey 2 respondent sample succeeded on two simple attention checks designed to look similar to the specific externality questions.[34] This can be seen in Figure H2.

4.4. Heterogeneity in inequality externality beliefs

We next explore which parts of the population believe in inequality's negative consequences. In general, as we show in Figure 2 and Tables I13-I15, externality beliefs are similar across different demographic groups and political affiliations.[35]

The largest difference is for individuals who identify as or lean Republican. These respondents are significantly less likely to believe in negative externalities (that inequality reduces trust, increases crime...).[36] Still, these beliefs are relatively similar across party affiliation. Between the three potential answers for each channel – negative inequality externality, no externality, and positive inequality externality – self-reported Republican-leaning respondents are most likely to choose the negative externality option for every outcome we elicit except for innovation, where the no-externality option is most likely (the negative externality option being second).[37]

Notably, Republican-leaning respondents choose the negative externality option more frequently than the positive externality option for all three economic outcomes we elicit.[38] Republican-

as much as the poorest tenth of society" to a society with "*a small middle class and many with relatively small or large incomes, [where] the richest tenth of society earns 30 times as much as the poorest tenth of society*". Other Survey 2 questions kept the Survey 1 format. Results within Survey 2 do not significantly differ when this reference point is included or excluded.

[32] Between 65% and 85% of respondents write arguments that directly corresponds to a causal explanation of their answer to the multiple choice question. An example for the case of the quality of democratic institutions: "*When a small group of rich people occupy the majority of wealth in a society, the society tends to be more corrupt. When the society becomes more corrupt, the quality of democratic institutions would be worsened.*"

[33] The phrasing change was across all of Survey 2. Changing the phrasing to "larger differences in income and wealth" has a smaller effect, indicating that it is not the word "inequality" but rather the direction of the shift in inequality that impacts the results.

[34] For example, "*How do you think more economic inequality – could you please click the first answer option?*. Note that respondents were screened on failing too many attention checks, see Appendix II.C.

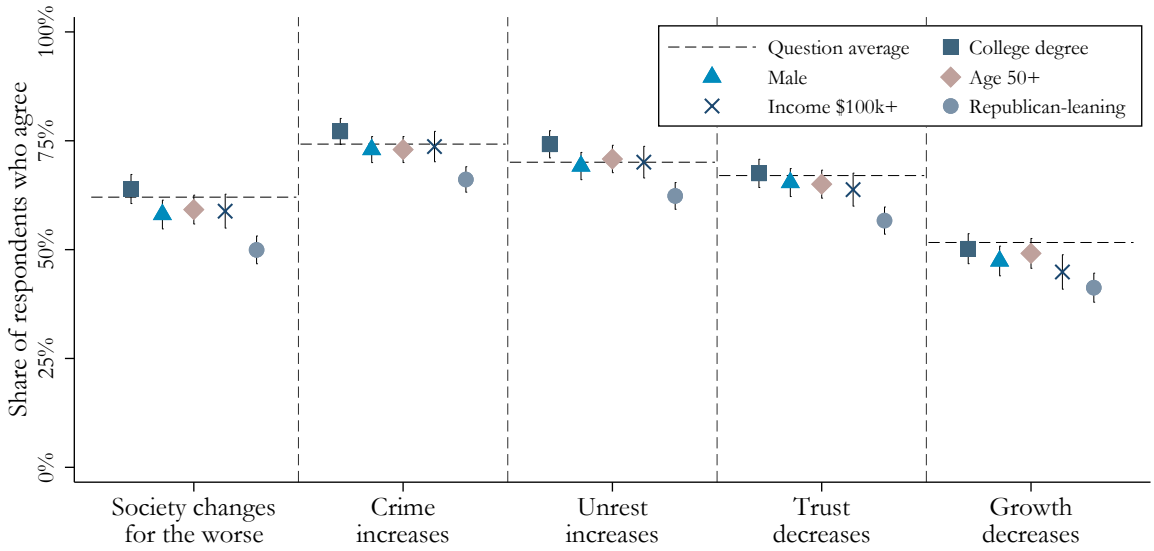
[35] The set of controls used in the tables was specified in our pre-analysis plan.

[36] We replicate Figure 1 for these two groups in Figure H5 and H6. We also replicate Figure 1 for very liberal or very conservative respondents in Figures H7-H8, and among respondents who feel closest to Bernie Sanders, Kamala Harris, Mitt Romney, or Donald Trump in Figures H9-H12.

[37] The same result holds for self-reported Republicans (excluding Republican-leaning Independents).

[38] This again holds for only Republicans.

Figure 2: Similar beliefs about inequality's consequences across groups



Note. Questions are from Figures 1 and H14. Shares are averages per group (no controls) with 95% error bars. See Tables I13-I15 for larger correlations with controls. $N = 3,290$.

leaning respondents are more likely to believe that more economic inequality *decreases* (41%) rather than *increases* (28%) economic growth, *increases* (41%) rather than *decreases* (20%) the level of unemployment, and *decreases* (31%) rather than *increases* (29%) the amount of innovation. These results are surprising, but are generally robust across our two surveys with distinct respondents and question methodology.[39] We show this for innovation and growth in Figure H13. We will further contextualize these party differences in Section 6.

The gender and age of the respondent do not generally predict their externality beliefs. College-educated individuals are consistently more likely to believe in negative externalities,[40] despite income and wealth generally not being significantly predictors (which we will return to in Section 6).[41] Finally, income inequality on the state level does not correlate with externality beliefs (not shown).[42] Overall, this analysis shows that beliefs in inequality externalities are widely held across every demographic group.

We summarize the above discussion in our first main result:

[39]The innovation result is the most fragile of the three, and changes under different question phrasing methods and robustness specifications. The growth results is robust to all specifications. The robustness of the unemployment externality beliefs were not explored (as they were not elicited in Survey 2). See Figure H13 for details.

[40]Note that this exaggerates the descriptive results by roughly 1 p.p., as our data has a larger share of college graduates than a fully representative national sample. See Appendix II.C.3 (for calculation) and Figures H3-H4 (for data weighted for full representativity).

[41]See Section 6.2.1 and Figure 8 for more. In the few exceptions, higher-income respondents generally believe somewhat less in inequality's negative consequences.

[42]Respondents who live in the *West* U.S. Census region have slightly stronger negative inequality externality beliefs than the remainder of our sample (a difference of ~ 5 p.p.).

Result #1

Regardless of their demographics or political associations, U.S. citizens tend to perceive inequality as having severe negative consequences and very few (if any) positive consequences.

We now move to the next main objective of the paper – exploring the effect of these inequality externality beliefs on individuals' redistributive preferences.

5. Redistributive Preferences and Inequality Externality Beliefs

In this section we explore whether inequality externality beliefs are a causal determinant of redistributive preferences. In the framework of Section 2, we have established widespread inequality externality beliefs $\mathbb{E}_i(\alpha_j)$ and will as such establish a causal connection for individual i if $\gamma_{ij} \neq 0$ for some outcome j where $\mathbb{E}_i(\alpha_j) \neq 0$. Robustly confirming such a link would broaden our comprehension of why redistribution occurs and differs across societies.

Our analysis centers on a video-based information provision experiment entirely contained in Survey 1 with $4,371$ respondents. Our main aim is to measure whether information about potential inequality externalities affect individuals' preferences for redistribution through shifted externality beliefs $\Delta\mathbb{E}_i(\alpha_j)$.

5.1. Experimental design

Survey 1 is divided into three parts. We show the survey structure in Figure 3. A video information intervention in Part 2 is our main treatment variation. Each respondent is randomly assigned to one of six groups; four treatment groups (20% chance, $\sim$ 875 respondents each) and two control groups (10% chance each). Respondents in each group sees a video, with the exception of the "passive" control group which we will return to. Three treatment groups are informed about the cross-country correlations of income inequality with various outcomes, which could affect respondents' inequality externality beliefs. One group is informed about the historical evolution of the income distribution. We will only discuss the three externality treatments and control groups in this section; the last "fairness" treatment will be further discussed in Section 6.

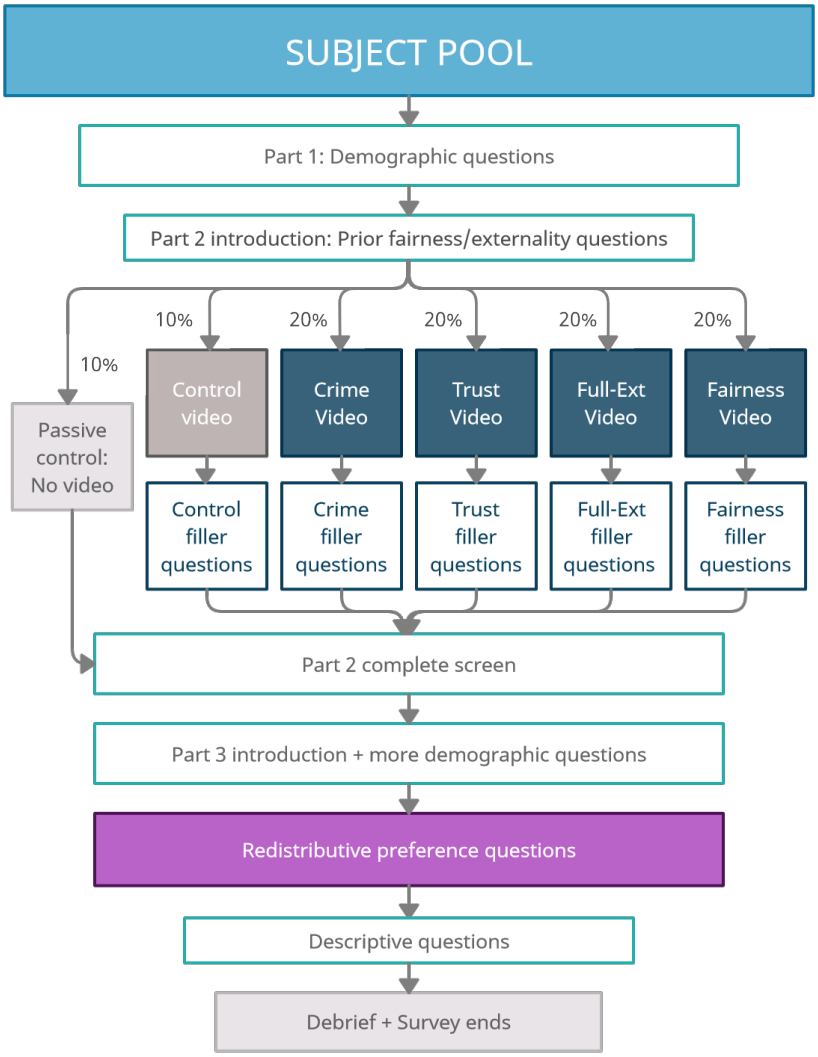
5.1.1. Video treatments

Each video is 1-2 minutes long. They are based on animated motion graphics that present information in an easily digestible way to prevent survey fatigue. Screenshots of two of these videos are shown in Figure 4.[43] Each externality video is designed to shift $\mathbb{E}_i(\alpha_j)$ in various ways; the below section briefly describes their contents. For more information see Appendix II.E.2.

Treatment group 1: Crime as an inequality externality Respondents mainly receive information on the positive cross-country correlation between intentional homicides and inequality with data from the World Bank and the World Inequality Database. As with the other two

[43] Links to the full videos are also in Figure 4. Screenshots from all five videos are shown in Figure E1.

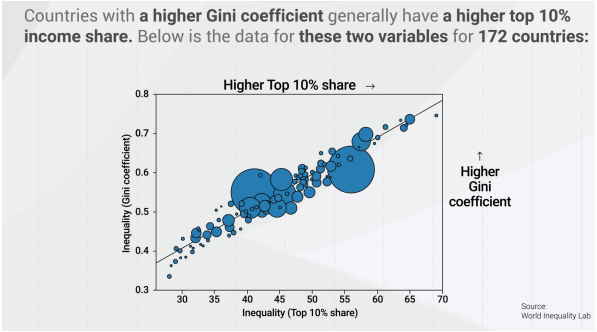
Figure 3: Survey Flow of Survey 1



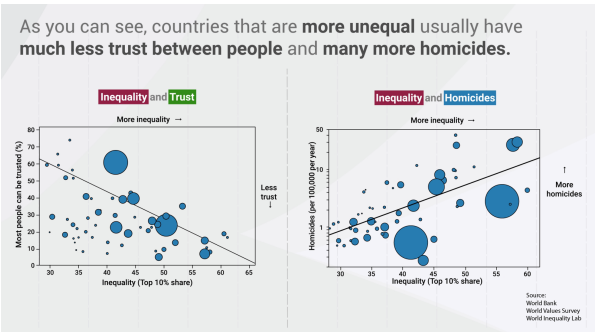
Note. The video information provision experiment is fully contained in Survey 1.

Figure 4: Treatment Videos, Example Screenshots

Active Control



Full Externality



Note. These are screenshots from the active control video (left) and the full externality video (right), two of the five videos used in the survey experiment. One video was shown to each respondent, except for the 10% of respondents in the passive control group. Click the following links for the full videos: Crime – Trust – Full externality – Fairness – Active control

externality videos, respondents are explicitly informed that the correlations do not necessarily imply causation.

Treatment group 2: Trust as an inequality externality Respondents mainly receive information on the negative cross-country correlation between generalized trust and inequality with survey data from the World Values Survey and the top 10% income share from the World Inequality Database. The video is structurally identical to Treatment 1.

Treatment group 3: Full externality treatment This treatment (see Figure 4) is designed as an all-encompassing externality treatment. It combines the information from Treatment 1 and 2 with additional correlational evidence which shows that inequality is not correlated to economic growth and innovation. It also includes a short discussion on how inequality could change societies in other ways.[44] By presenting broad evidence that highlights the negative effects of inequality and by showing that the evidence for positive externalities is rather limited, the treatment makes the strongest case for the negative consequences of inequality.

Summary of the variation induced through the treatments The three externality treatment groups provide information about inequality's externality properties[45] while attempting to avoid fairness-related topics. Theoretically, these treatments should affect $\Delta\mathbb{E}_i(\alpha_j)$ while ideally not affecting Υ_i. Comparing redistributive preferences to a control group thus gives us insights into whether $\gamma_{ij} \neq 0$ and whether inequality externality beliefs are a causal determinant of redistributive preferences. In Appendix II.E.5 we formalize this discussion.

The expected mechanism hinges on no other redistribution-related spillovers through Υ_i. This is naturally a large assumption which we explore in various ways. First, experimenter demand and priming present potential external validity issues for video information experiments such as ours. To minimize these issues we present two novel methodological approaches.

5.1.2. Dual control groups

The two control groups include one "passive" control group with no stimuli and one "active" control group where respondents see a neutral video on inequality metrics (see Figure 4). The aim of this method is to minimize the intrinsic issues with either a passive control (attention effects, lack of priming, attrition) or an active control (potential unintended and unmeasurable effects on outcomes). We merged these two groups on pre-specified outcome criteria. As far as we know this is a novel methodological choice within the information experiment literature. We discuss the approach further in Appendix II.E.3.

5.1.3. Secondary survey

We design the survey around a "secondary survey" to obfuscate the real purpose of our survey from respondents. This is a particular type of the obfuscated information treatments discussed

[44] The video notes that some researchers believe inequality can increase social unrest, corruption, and political polarization. It also contains a quotation by Amartya Sen ("*I believe virtually all the problems in the world come from inequality in one way or another*"), cited as a Nobel-prize winning economist.

[45] It should be noted that all our treatments are designed to have weakly positive effects on beliefs, in the sense that the induced variation in beliefs should always lead to a weakly larger demand for redistribution. This feature is implemented by design for two main reasons. First, it enables us to form clear hypotheses for potential treatment effects. Second, there are no clear existing correlations between inequality and societal outcomes that imply positive externalities. The correlational evidence we present on crime and trust is strong and robust to different specifications; this is not the case for any potential positive externalities we know of.

in Haaland et al. [Forthcoming].[46] We define a secondary survey as a logical flow of questions that explains the information treatment while disguising the true purpose of the survey. This relies on (i) separating the treatment and outcome variables as much as possible, and (ii) giving respondents a reason for having seen the provided information. The former is meant to reduce experimenter demand and priming; the latter is meant to avoid respondent confusion and suspicion.

In the present survey, after simple demographic questions (Part 1), the respondents are introduced to Part 2 of the survey. Respondents who are not in the passive control group are then showed a video and immediately afterwards a set of questions that are related to the video but nevertheless unrelated to our research questions. They are then informed that Part 2 of the survey is complete, implying that the video has fulfilled its purpose, and that Part 3 is beginning. Respondents then see a battery of unrelated demographic questions. After this structural break, all outcomes (e.g. redistributive preferences) are elicited. The net effect is a significant pause between the video and the questions of interest as well as a well-explained reason for *why* the respondent saw a video. The goal of the secondary survey is to reduce experimenter demand, priming, and respondent confusion. We discuss this approach further in Appendix II.E.4.

5.2. Experimental results

We compare observable characteristics across the treatment and control groups across in Appendix II.F.2. As expected from our research design, the groups are generally well-balanced.[47]

5.2.1. Redistributive preferences: Main treatment effects

We measure individuals' redistributive preferences with four survey questions and a combined index of these questions.[48] Note that small-scale redistributive games could not be used as micro-level preferences should not be affected by beliefs in macroeconomic inequality externalities.[49] The four questions are the following. First, a question on respondents' preferred level of redistribution on a scale from no redistribution to full redistribution. Second, a question from the European Social Survey which asks respondents whether the government should take measures to reduce inequality on a Likert scale.[50] Third, a question about whether the respondent believes that inequality is a very serious issue in the United States [used in Stantcheva, 2021, among others]. And fourth, a specific policy preference question which asks respondents about their preferred average tax rate for the "Top 10%" over seven different options.[51]

We pre-specified these four outcomes (general redistributive preferences, government should reduce inequalities, inequality is a serious issue, top tax rates) as well as a redistributive prefer-

[46] Obfuscated follow-up surveys are generally resource-heavy and assume the researcher has access to respondents over time. The main benefit of the secondary survey is that it functions as an obfuscated follow-up within a survey for cases when true obfuscated surveys are not possible.

[47] Though there are small differences in observables, these do not reveal any systematic differences across treatment groups. Note that our regressions control for observable characteristics; including or excluding these regressors does not change the results. We also discuss the (limited) differential attrition across treatments in Appendix II.C.

[48] All redistributive preference outcomes are shown in full in Appendix II.E.6.

[49] Survey questions regularly predict real-world outcomes in similar work [Haaland and Roth, 2021, Alesina et al., 2023].

[50] The main difference between the first two questions is the explicit presence of government.

[51] These options are 0%, 0-15%, 15-25%, 25-35%, 35-45%, 45-70%, and 70-100%. Each option contains a short explanation (e.g. "*35-45%: I want to tax them at a higher rate than now, but not very high*").

ences index ("RP Index") combining these four outcomes.[52] The index was pre-specified to be the primary outcome.

Table 1 shows how each treatment affects individuals' redistributive preferences. Most importantly, the full externality treatment has a significant and, for this kind of study, reasonably large effect on respondents' redistributive preferences. Three of the four pre-specified measures of redistributive preferences are significant and the effect on the aggregate redistributive preference index is significant at the 1% level.[53] The index increases by 11 percent of a standard deviation in response to the treatment, or about $\frac{1}{8}$ of the difference between Republican- and Democrat-leaning subjects.

Under the caveat that the mechanism is as expected – which we return to shortly – this represents the first part of our second main finding:

Result #2a

Inequality externality beliefs are a causal determinant of redistributive preferences.

The sizable treatment effect also indicates that these beliefs could be a substantial determinant of these preferences; we explore this further in Section 6.1.

Table 1: Main Treatment Effects of Video Information Experiment

	(1) RP Index b/se	(2) Wants redistribution b/se	(3) Gov. reduce ineq. b/se	(4) Ineq. is serious issue b/se	(5) Increase top taxes b/se
Crime Ext. Tr.	0.037	0.031	0.007	0.020	-0.005
	(0.036)	(0.020)	(0.020)	(0.019)	(0.021)
Trust Ext. Tr.	0.043	0.006	0.036*	0.017	0.004
	(0.037)	(0.021)	(0.020)	(0.020)	(0.022)
Full Ext. Tr.	0.107***	0.050**	0.048**	0.069***	-0.012
	(0.037)	(0.021)	(0.020)	(0.020)	(0.022)
Controls	Yes	Yes	Yes	Yes	Yes
R2	0.391	0.169	0.293	0.313	0.170
Observations	4371	4371	4371	4371	4371

Note. This table reports results from a regression of different redistributive preference outcomes and the treatment dummies, as well as socio-economic control variables. The RP index is normalized on the sample and has units of the number of standard deviations. The remaining variables are binary (0-1). Controls not listed in the table include trust in government, race, income-group, age-group, education, employment status, geographic region. Standard errors are in parentheses. *Significance levels*: *10%, **5%, ***1%.

The crime and trust externality treatments have only weak and mostly insignificant effects on redistributive preferences (though largely in the expected direction). In general, it appears that information about inequality externalities is more convincing when given in a comprehensive

[52] This main outcome index was pre-specified as the standardized sum of dummy versions of all the four outcomes. Where to split the four questions into dummy variables – e.g. 35-45% and above for the tax question – was also pre-specified, intending to split each question into roughly equal fractions.

[53] We do not find any effect of any externality treatments on preferences for top-income taxation. As the other treatment effects from the full externality treatment are strongly robust, this is somewhat surprising. This can be due to the respondents not fully internalizing the connection between higher top tax rates and lower inequality, or because respondents believe that the effect of inequality on trust and crime is primarily affected by inequalities near the bottom as corroborated by Table D1. We also note that the active control showed a surprisingly high treatment effect for this variable (see Appendix II.F.1) – the non-result from the externality treatments could also be driven by this anomaly.

fashion. In other words, discussing the widespread effects of inequality is more impactful than focusing on any single type of externality.[54]

5.2.2. Mechanism

Why do redistributive preferences change? Here we show that the treatment effect mechanism appears to come through inequality externality beliefs with limited spillovers.

First-stage outcomes After the redistributive preference outcomes we elicited respondents' externality beliefs (representing $\Delta\mathbb{E}_i(\alpha_j)$) and broad economic fairness views (a proxy for $\Delta\Upsilon_i$). We define fairness views as denoting respondents' answers to questions related to the fairness of the economic system or the origins of income differences. These questions represent our first-stage results.

The four externality questions we use are described in Section 4.[55] We also elicit respondents' broad economic fairness views with two questions.[56] The primary goal of these questions is to explore (and potentially exclude) spillover effects of the externality treatments on economic fairness views.

The first-stage outcomes of the experiment are shown in Figure 5. As can be seen in the first four columns, each of the three externality treatments significantly change respondents' general and specific externality views. The targeted specific externality concern is most affected; as an example, the crime externality belief is the most affected by the crime and full externality treatments.[57] Overall, the first stage treatment effects are strong as each video increases beliefs in the intended direction by roughly 10 percentage points.[58] These results are corroborated in an open-ended text question about externality opinions, where respondents are significantly more likely to mention the topics from their corresponding treatment videos without explicitly discussing the video itself.[59] None of the externality treatments significantly affect broad economic fairness views (last two columns), showing that spillovers to other determinants of redistributive preferences are limited. Indeed, this provides some evidence that inequality externality beliefs are relatively independent from economic fairness beliefs.

Mechanism: Other evidence Additional results also indicate that redistributive preferences were shifted through respondents' inequality externality beliefs.[60]

[54] Theoretically this is sensible. If only the externality belief of crime is affected, redistributive preferences are only affected proportionally to $\gamma_{i,crime}$ instead of being affected by all $\gamma_{i,j}$. This could be a small number; from the incomplete subset presented in Figure H15, we can estimate that crime represents at most 12% of individuals' externality concerns. Although we do observe spillovers to other externality beliefs – which would increase this number – we consider this the most likely reason for why the Crime and Trust videos have non-significant treatment effects.

[55] Specifically the questions on general externality beliefs, crime, trust, and economic growth. All first-stage questions are shown in full in Appendix II.E.6.

[56] The first asks whether the current distribution of income and wealth in the US is fair (or unfair) because people get what they are entitled to (or not). The second asks whether hard work or luck "has more to do with" why a person is rich. We note that these questions explore "broad" fairness ideas in the sense that they can include (i) beliefs about the origin of economic inequalities, and/or (ii) views on economic fairness given a known economic process and distribution. In this case the first question includes both (i) + (ii), whereas the second question only includes (i).

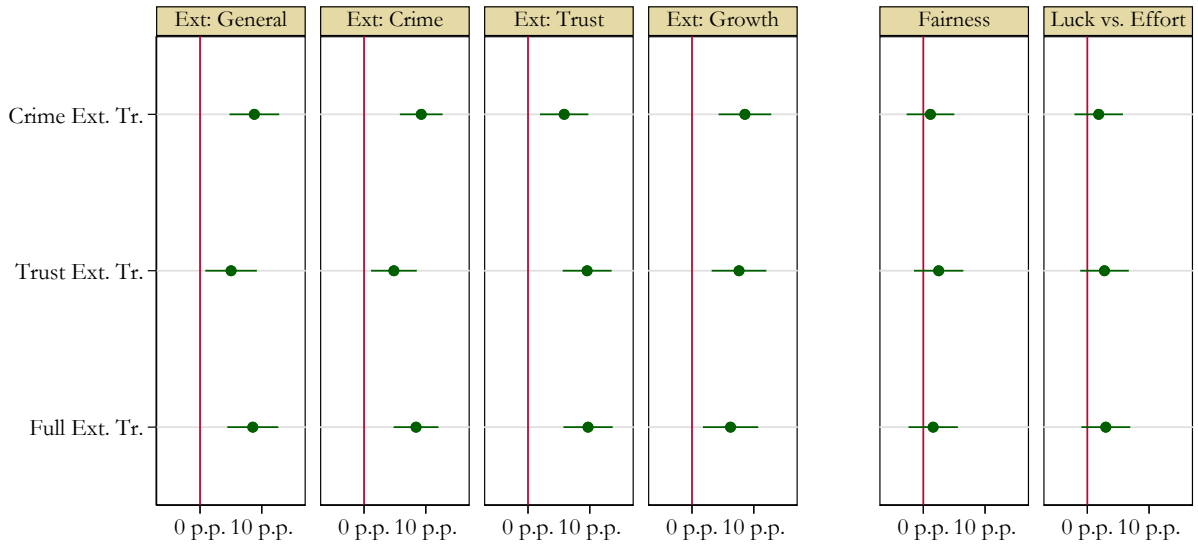
[57] We do note that there seems to be externality-based spillovers; the crime treatment changes their trust externality belief, for example.

[58] These are particularly sizable given that the control means for negative externality beliefs are already high – the crime- and trust-externality beliefs in the control group are at 76% and 68% respectively, for instance.

[59] See Appendix II.F.3.

[60] Note that we did not include this analysis in our pre-analysis plan.

Figure 5: First-stage Effects of Treatments



Note. This figure reports results from a pre-specified regression of different externality beliefs and fairness views on the treatment dummies as compared to the control group. The general and specific externality belief variables (left) are discussed in Section 4. The fairness variables (right) indicate whether the respondents believe the distribution of income in the U.S. is generally unfair (*Fairness*) and whether the respondents believes high-income individuals became rich mainly due to luck or effort (*Luck vs. effort*). Controls include political leaning, gender, trust in government, race, income-group, age-group, education, employment status, and geographic region. Error bars characterize 95% confidence intervals. Appendix Table I20 presents the point estimates and standard errors. *Significance levels*: *10%, **5%, ***1%.

First, we conduct a mediation analysis and show that the magnitude of the treatment effects are reduced after controlling for the matching first-stage beliefs. This is as expected if redistributive preferences are shifted through these beliefs, and is shown in Table I24 and detailed in Appendix II.F.4. Second, the externality treatment effects are strongest for individuals who did not believe in negative inequality externalities at the beginning of the survey (Table I25). Third, respondents that self-reported that they learned something new are more likely to have changed their redistributive preferences (Table I26).

Overall, our results strongly suggest that redistributive preferences are shifted through the expected mechanism of $\Delta\mathbb{E}_i(\alpha_j)$ with limited spillovers through Υ_i.

5.2.3. Heterogeneous treatments

We now explore heterogeneous treatment effects across incomes and political affiliations.[61]

Individuals above $\$100,000$ in yearly income are affected about twice as much by the full externality treatment as lower-income individuals (Table I29). The result is not driven by unequal shifts in first-stage beliefs, which are similar across income groups. This is in-line with the theoretical discussion in Section 2. Inequality externality beliefs are a reason for self-interested individuals to prefer more redistribution, and the self-interested individuals who are originally against redistribution are more likely to be at the top of the income distribution (as self-interested individuals near the bottom benefit economically from redistribution). It follows that high-income individuals should shift their redistributive preferences more for an income-independent shift in externality beliefs.[62] This has intriguing consequences for the broader

[61] We will discuss how these compare with the fairness treatment in Section 6. Note that we did not pre-specify the interaction in income, but include it as it is both robust and of particular interest.

[62] Assuming similar shifts across the income distribution for altruistic respondents.

redistributive conversation, as inequality externality beliefs appear to have the potential to reduce the existing heterogeneity in redistributive preferences across incomes. Such a reduction in heterogeneity could reduce conflicts of interest and incentivize a redistributive consensus across the income distribution.

Across political affiliations, the treatment effect of the full externality treatment is largely driven by Democratic-leaning respondents (Table I30). This is despite similar shifts in first-stage beliefs $\Delta\mathbb{E}_i(\alpha_j)$ across all political affiliations. As such, Republican-leaning respondents seem to learn about inequality externalities but not change their preferences for redistribution. This indicates that Republican-leaning respondents have low γ_{ij}.[63]

5.2.4. Robustness of treatment effects

The conclusions from the information experiment are generally very robust to various specifications. In Appendix II.F.5 we discuss the robustness of the treatment effects to (i) fully representative population weights, (ii) keeping respondents with very fast/slow survey completion times or unusual text answers, (iii) excluding all respondents who failed at least one attention check, (iv) using only one control group, (v) not controlling for observable characteristics, (vi) using different sets of control variables, (vii) using non-dichotomized outcome variables, and (viii) multiple hypothesis testing. Point estimates do not change in a noteworthy fashion to any of these checks.[64] We discuss this further in Appendix II.F.5.

6. Comparing inequality externality beliefs to other redistributive determinants

So far we have established that inequality externality beliefs are widely held in the United States and that they causally affect redistributive preferences in at least a marginal way. This, in turn, raises further questions. Are these beliefs a *sizable* determinant for redistributive preferences? And do they have distinct properties as compared to other known redistributive determinants? This section discusses these two questions.

6.1. Impact on redistributive preferences

This section characterizes the relative importance of inequality externality concerns as determinants of redistributive preferences. We particularly compare the externality beliefs to *broad economic fairness views*. We define these broad fairness views as the combination of people's beliefs about the origin of the economic distribution (e.g. whether hard work or luck is more important) and whether they believe any perceived unfairness is problematic. We use these views as a point of comparison because fairness views have been identified as a crucial motive behind individuals' preferences for redistribution [e.g. Cappelen et al., 2007, Durante et al., 2014, Almås et al., 2020], thus serving as a useful benchmark. To that end, we pre-specified three different approaches which we now go through in turn.

[63] In other words, Republican-leaning respondents seem to have a low willingness to redistribute to reduce known negative inequality externalities.

[64] Standard errors, however, increase significantly under some of these procedures – specifically reweighting for representativity, dropping all attention check failures, and using only one control group. As a result, certain treatment effects that are statistically significant in the original data no longer reach this threshold under these specifications.

6.1.1. Comparing information treatments

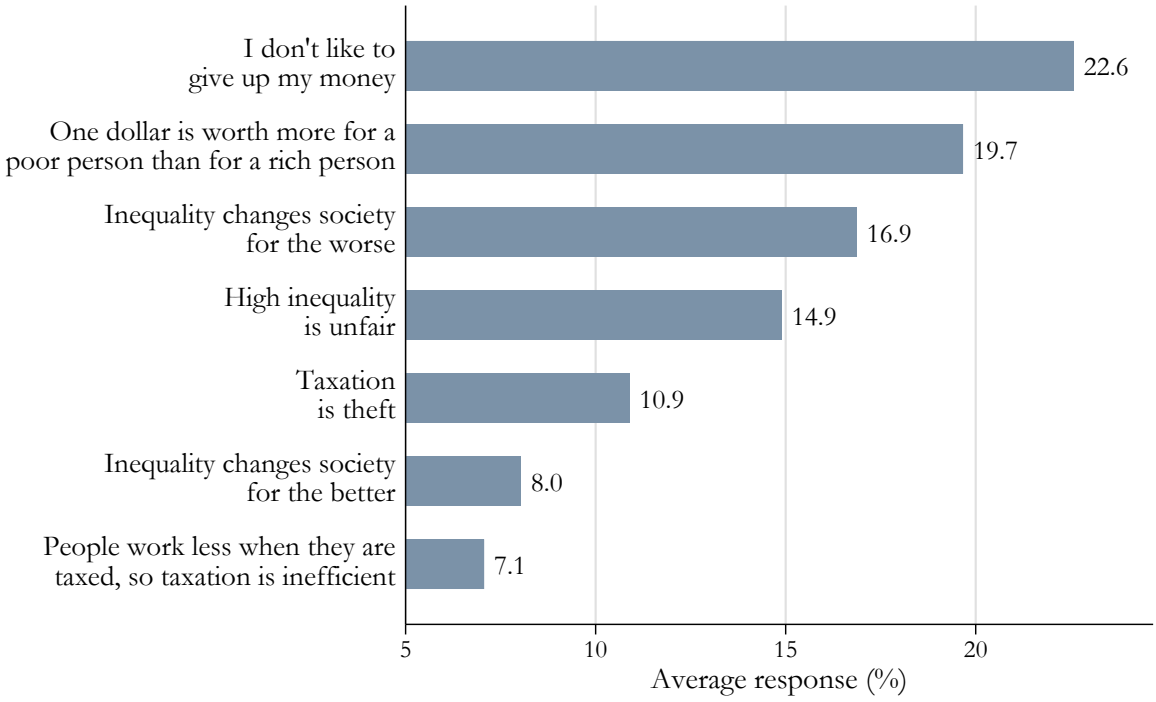
The information experiment also included a final "fairness" treatment. This treatment group received information on how U.S. manufacturing workers' wages stagnated while productivity increased between 1980-2019, using data from the Economic Policy Institute. This is contrasted to the growth of the top 1% income share in the U.S. from the World Inequality Database.

This video has about twice the effect on our pre-specified index of redistributive preferences as the full externality video, which we show in Table I18.[65] As these are marginal effects that are also dependent on the efficacy of the treatment video, this is only indicative evidence for the relative strength of these arguments as a whole. We discuss the design and mechanism of the fairness video further in Appendix II.F.6.

6.1.2. Directly ranking redistributive motives

In a more direct approach, subjects were asked to allocate 100 points across different motives behind preferences for redistribution. This survey-item provides direct evidence of respondent's redistributive motives under the assumption that they are able to discern and report these motives.

Figure 6: Directly Elicited Motives of Preferences for Redistribution



Note. Question text: *When thinking about your preferred level of redistribution, what matters most to you? Please indicate what dimensions matter by giving scores below that add up to 100.* Answer option texts are identical to graph labels. Standard errors are approximately 0.6%. Sample is the merged descriptive sample (N=3,292). Results are very similar across surveys.

The negative inequality externality motive receives broad support in this very direct ap-

[65] The coefficients are 0.107 and 0.208 respectively. We can reject equality of the two coefficients at the 5% significance level (p = 0.012, t-test). As there are some spillovers from the fairness video to externality beliefs, a strict magnitude comparison could overestimate difference between the underlying determinants. See more in Appendix II.F.6.

proach, coming in third among the seven options (Figure 6). This is below income maximization and concerns about diminishing marginal utilities of income. It is slightly above a strict fairness option, and relatively high above the last remaining options – a libertarian motive, positive externality considerations, and concerns about the efficiency losses from taxation.

This notable result indicates that if directly asked, a substantial portion of U.S. citizens indicate that inequality externalities are a large driver of their redistributive preferences. The results also indicate that inequality externality beliefs are on the same order of magnitude as economic fairness views – if broken down into natural groupings, inequality externalities are roughly 70% as important as economic fairness ideas.[66] Inequality externalities are also seen as much more significant than traditional efficiency concerns, which underscores the commonly found result that efficiency concerns are relatively unimportant in this debate [Durante et al., 2014, Stantcheva, 2021]. While this approach naturally has caveats, we believe the direct elicitation approach is in many ways the clearest way to understand respondents' preferences on such topics. We discuss this approach further in Appendix II.G.1.

6.1.3. Predictive power on redistributive preferences

We also pre-specified an analysis of the predictive power of externality beliefs and three other sets of variables on redistributive preferences. This analysis consists of regressions on the redistributive preference index that include regressors of respondents' answers to two questions on, respectively: (i) fairness views, (ii) externality beliefs, (iii) political preferences, or (iv) respondents' trust in government and belief that higher taxes lead to efficiency losses. We then compare the explanatory power of these models using the adjusted R^2.

We show the results in Table 2.[67] The two inequality externality beliefs we include explain roughly 20% of variation in redistributive preferences.[68] This is somewhat below that of fairness views (28%) and equal to that of political preferences (20%). The predictive power of determinants often found in the academic literature – governmental trust and a belief that taxation leads to less work – is very small (5%).

When combining all determinants into one regression, externality beliefs remain highly significant. This is also true when including more non-externality variables into the regression.[69] This indicates (but does not ascertain) that they capture some variation that is not captured by the other determinants we include. This is particularly notable as it pertains to fairness views, which are often used as a proxy for redistributive preferences in academic work. If inequality externality beliefs capture independent and causal variation in individuals' redistributive preferences, future academic work focused on redistributive preferences may do well to measure these beliefs directly.

Finally, we note that respondents' opinions on whether taxation reduces work effort is no longer significant in the combined regression, mirroring the previous section. We discuss this method further in Appendix II.G.2.[70]

[66] The calculation combines positive and negative externality beliefs and compares this to the diminishing marginal utility and strict fairness motives.

[67] All question answers are in pre-specified binary form designed to for 50/50 splits. The binary nature of the regressors constrain their predictive power.

[68] This does not include the 10% explanatory power of the demographic controls.

[69] A pre-specified version with three questions per group reaches the same conclusions, for example.

[70] Our conclusions remain identical when analyzing other versions of this regression, notably with more questions

Table 2: Horse-Race: Predictive Power of Beliefs on Redistributive Preferences

	(1) RP Index b/se	(2) RP Index b/se	(3) RP Index b/se	(4) RP Index b/se	(5) RP Index b/se	(6) RP Index b/se
Rich because of luck		0.624***				0.401***
		(0.060)				(0.057)
Society is unfair		0.620***				0.416***
		(0.059)				(0.056)
Belief uneq. countr. worse			0.434***			0.269***
			(0.058)			(0.050)
Neg. externality belief			0.640***			0.272***
			(0.058)			(0.054)
Leans Republican				-0.429***		-0.245***
				(0.084)		(0.072)
Sanders/Harris supporter				0.533***		0.260***
				(0.085)		(0.075)
Trusts the government					0.436***	0.131**
					(0.066)	(0.054)
Taxation reduces work					-0.115*	-0.004
					(0.061)	(0.048)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R2	0.104	0.382	0.297	0.296	0.148	0.494
Observations	932	932	932	932	932	932

Note. This table reports results from a regression of different redistributive preference outcomes on fairness views, political views, externality beliefs and attitudes towards the government, as well as socio-economic control variables. Controls not listed include gender, race, income-group, age-group, education, employment status, geographic region. Standard errors are in parentheses. *Significance levels*: *10%, **5%, ***1%.

The three preceding methods have all attempted to answer the question of whether inequality externality beliefs are a *sizable* determinant for redistributive preferences. Each (imperfect) method reaches qualitatively similar conclusions. Inequality externality beliefs are a large determinant of redistributive preferences; on the same order of magnitude but somewhat less powerful than broad economic fairness views. To the extent that the relative strengths of these determinants can be numerically explored, the three methods find the consistent results of inequality externality beliefs being roughly 50-70% as impactful as broad economic fairness views.[71] Meanwhile, inequality externality beliefs are consistently more impactful than taxation-based efficiency concerns.[72] Similar conclusions are also found when using a Gelbach decomposition to explore which survey questions explain the redistributive preference differences across Democrats and Republicans (Appendix II.G.2).

The above can be summarized in our second main result, building on Result #2a:

included.

[71] Naturally this numerical exercise must be taken with extreme caution, as even clearly defining what such a number means is challenging.

[72] To the extent that we have data on the topic this also holds for trust in government.

Result #2

Inequality externality beliefs are a *sizable* causal determinant of redistributive preferences, substantially outperforming taxation-related efficiency concerns and approaching broad economic fairness views.

6.2. Unique properties of inequality externality beliefs

Inequality externality beliefs are thus widely held and a strong determinant of redistributive preferences. But are they simply another way of discussing inequality in the vein of economic fairness ideas, or do they also have other unique properties? It is easy to imagine either; in the form of (2.1), $\sum_j \gamma_{ij}\mathbb{E}_i(\alpha_j)\theta$ may or may not be similar to Υ_i. In this section we discuss various evidence to indicate structural differences between inequality externality beliefs and other redistributive determinants.

6.2.1. Descriptive consensus across demographic groups

As described in Section 4.4, there is a widespread consensus across various groups that inequality is a negative externality. Here we show that this consensus appears unique to externality beliefs and does not extend to economic fairness views,[73] which are consistently more polarized across both economic status and political affiliations.

We first show how inequality externality beliefs and fairness views differ across economic status. As briefly mentioned in Section 4, externality beliefs are relatively independent of income and wealth. This contrasts to economic fairness views, where respondents with high economic status are generally more likely to think the economic distribution is fair [a result we find in our data which is consistent in the previous literature, e.g. Valero, 2021, Hvidberg et al., 2022]. We show an example of this pattern in Figure 7.[74] That fairness questions are more polarized across economic status is a general result for nearly all our questions. We show this by plotting every externality and fairness question we have already discussed from Survey 1 in Figure 8.[75] In sum, while economic status strongly predicts fairness views, both rich and poor are overall likely to think inequality has negative consequences.

The same pattern also holds across political affiliations, where fairness views are even more consistently polarized. This is illustrated in Figure 9 for the same two example questions and in Figure 10 for the same broader set of questions.[76] Generally speaking, Republican-leaning respondents appear much more likely to have negative inequality externality views than to believe

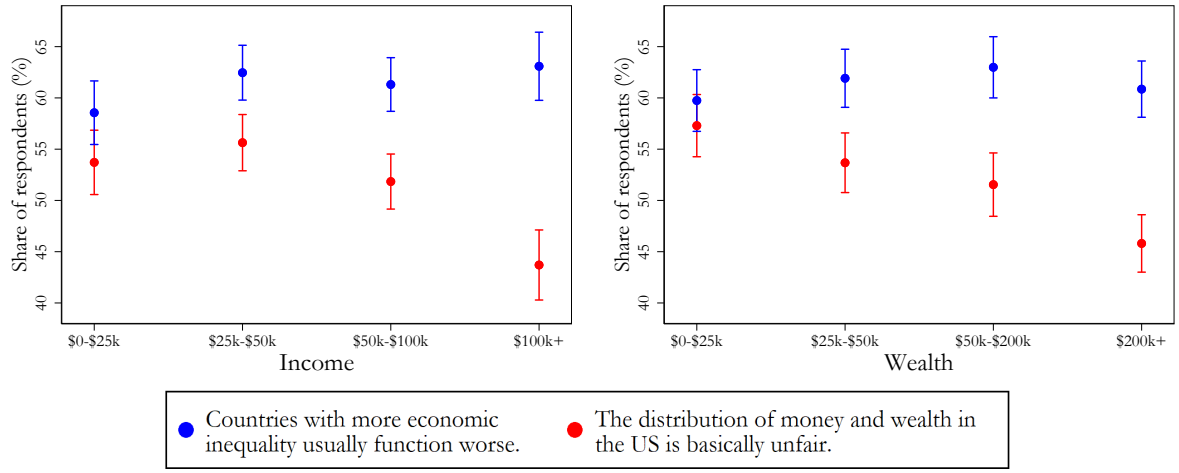
[73] As previously mentioned, these views denote respondents' answers to questions related to the fairness of the economic system or the origins of income differences.

[74] These two questions ask respondents to agree or disagree with the statements that (i) "*The distribution of money and wealth in the U.S. is basically fair, because everybody has an equal opportunity to succeed*" and (ii) "*Countries with more economic inequality generally function worse*". These questions were chosen as an example as they were posed *before* the treatment intervention, allowing us to use the full Survey 1 sample ($N = 4,371$).

[75] The externality questions are those shown in previous figures (Figure 1 and Figure H14). All fairness questions in the survey are shown. The selection of these questions is for simplicity; the same pattern holds for every externality- and fairness-related question in Survey 1. In Survey 2 (Figure H23) a similar pattern holds, although there is some overlap between questions. This strong result is robust to adding demographic controls within Survey 1. In Survey 2, adding demographic controls makes three more externality questions overlap.

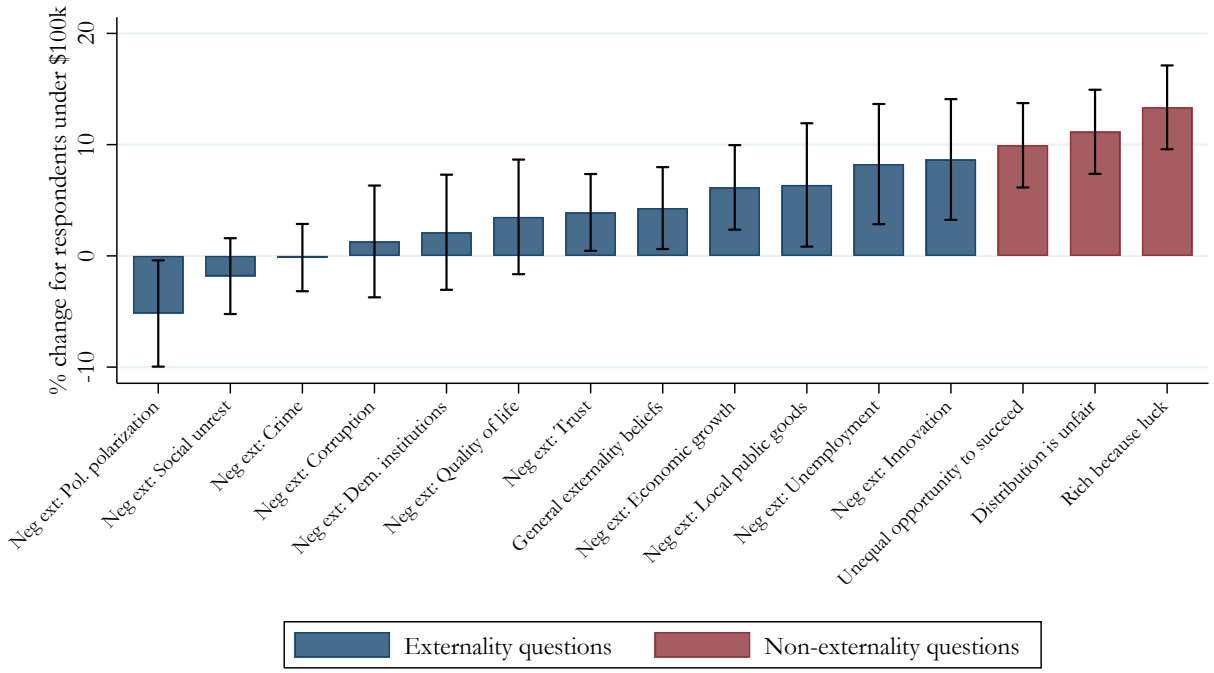
[76] The same result holds for every externality- and fairness-question in Survey 1, and nearly every set of questions in Survey 2 (see Figure H21). The result is also robust to adding a standard set of controls.

Figure 7: An Example of Externality Beliefs and Fairness Views over Income and Wealth



Note. These graphs use the pre-treatment externality and fairness questions with the full Survey 1 sample ($N = 4,371$). Respondents are asked whether they agree with the following statements: *"The distribution of money and wealth in the US is basically fair, because everybody has an equal opportunity to succeed."* and *"Countries with more economic inequality usually function worse."*. For the equivalent graph from Survey 2, see Figure H22.

Figure 8: The Effect of Income Level on Fairness Views and Inequality Externality Beliefs



Note. Difference in pro-inequality sentiment (e.g. inequality does not increase crime, does not decrease trust, is fair) for respondents with incomes above $100,000 across selected externality and non-externality ("fairness") questions in Survey 1 where the only controls are for treatment groups. The same relation (every externality question has less polarization than any "fairness" question) holds for every question in Survey 1, and if we restrict the sample to only the control group. With a standard set of controls the three most income-dependent externality questions (those on innovation, local public goods, and unemployment) are roughly as income-dependent as the fairness questions. Questions are largely split on pre-specified criteria or natural binary points (e.g. agree/disagree), keeping total shares close to 50% where possible. Error bars are 95% confidence intervals. $N = 4,391$.

that the economic system is unfair. We note that measurement error presents a potential caveat for these results.[77]

Generally speaking, that fairness views are more income- and party-polarized than externality beliefs is an extremely consistent result across both surveys. These results indicate that the large agreement across demographic groups about inequality's externality effects is a unique feature of these types of arguments. Still, as is clear from the heterogeneous treatment effects we discussed in Section 5.2.1, consensus on *descriptive* statements (e.g. $\mathbb{E}_i(\alpha_j)$) does not necessarily imply a subsequent consensus on redistribution. We now return to these heterogeneous treatment effects to compare the externality treatments to the fairness treatment.

6.2.2. Heterogeneous treatment effects

Income As previously discussed, the externality treatments are generally stronger for top-income individuals than lower-income individuals. The opposite is true for the fairness treatment, which is much stronger among low-income individuals (Table I29).[78]

This is both intuitive and in accordance with the theoretical arguments discussed in Section 2. While nothing precludes other-regarding preferences to have a similar form as externality beliefs – for example $\Upsilon_{fair,i} = \gamma_{i,fair}\mathbb{E}_i(\alpha_{fair})\theta$ – our evidence indicates that there is significant individual-level heterogeneity in both (i) descriptive externality or fairness beliefs, $\mathbb{E}_i(\alpha_{ext})$ as compared to $\mathbb{E}_i(\alpha_{fair})$, and (ii) who allows these beliefs to affect their redistributive preferences, or $\gamma_{i,ext}$ as compared to $\gamma_{i,fair}$.

In other words, fairness-based and inequality externality-based arguments for redistribution likely affect different subsets of the population. Self-interested individuals is one intuitive example; such individuals have no other-regarding preferences ($\gamma_{i,fair} = 0$) but may very well be affected by inequality externalities ($\gamma_{i,ext} \neq 0$). As discussed in Section 2, this presents a natural explanation to the heterogeneous treatment effects across income and suggests intrinsic differences between equity-based and externality-based arguments about redistribution.

Political Party Affiliation As previously discussed, the externality treatment is largely driven by Democratic-leaning respondents. This stands in contrast to the fairness treatment effect, which is roughly equal across political affiliations (Table I30).[79] This is the opposite of the descriptive beliefs we have just discussed. Overall, this is puzzling; we hypothesize that the libertarian fairness principles of Republican-leaning respondents could be difficult to overcome with consequentalist arguments.

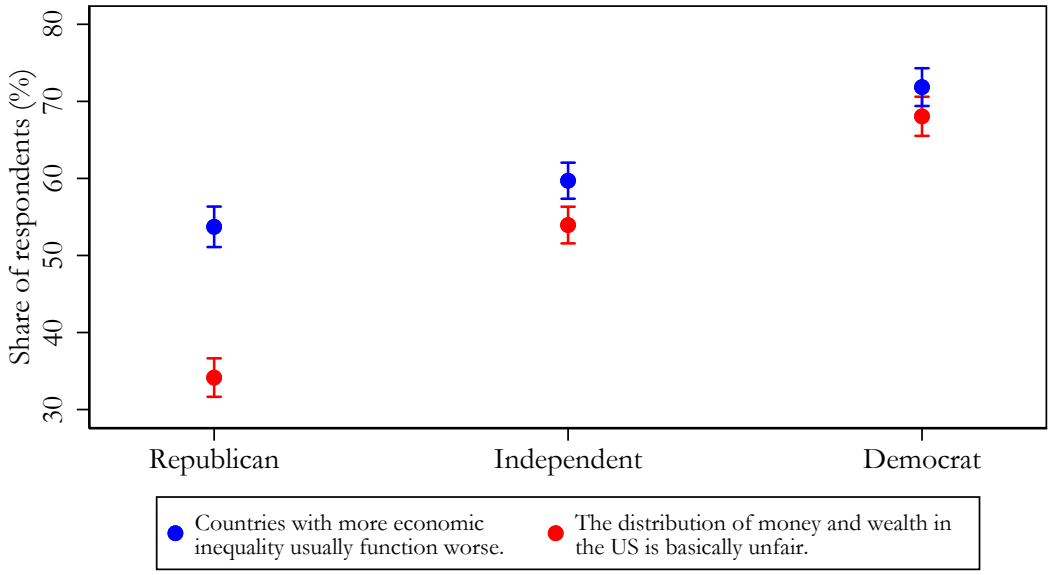
6.2.3. Emotional reactions: Anger

[77] Suppose there are systematic differences in the amount of measurement error in individuals' responses to these questions across question type. For example, individuals could be more familiar with fairness questions and thus answer these questions more accurately based on their genuine beliefs. This could explain the low polarization of externality questions; in the extreme case where there is no signal in respondents' answers, one would expect no aggregate differences across groups. We find some evidence of larger measurement error in externality questions, as respondents' answers are less strongly correlated across externality questions than fairness questions. However, we also find that the vast majority of respondents confirm their externality beliefs when asked in a follow-up open-ended text question in Survey 2. The relatively small party and income differences for externality views persist across these open-ended text answers.

[78] This is not driven by differential take-up of the first-stage belief, which is equal across both groups.

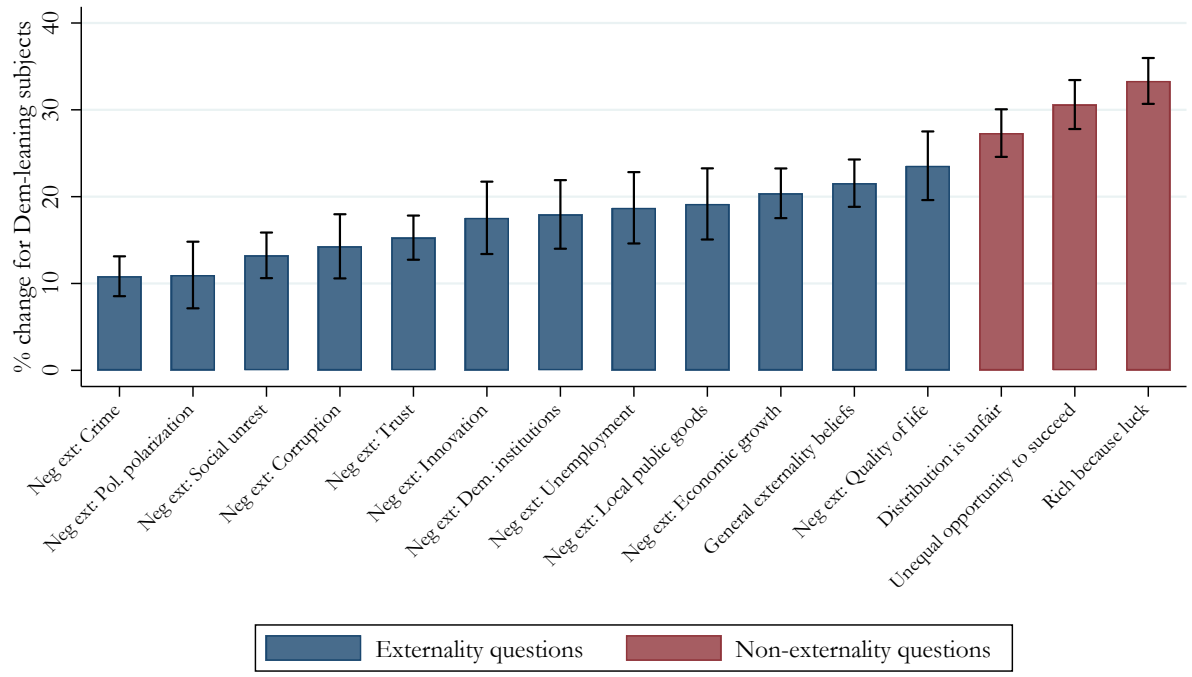
[79] This is also not driven by differential take-up of the first-stage belief, which is again equal across both groups.

Figure 9: An Example of Externality Beliefs and Fairness Views over Party Affiliation



Note. This graph uses the pre-treatment externality and fairness questions with the full Survey 1 sample ($N = 4,371$). Respondents are asked to agree or disagree with the following two statements: "*The distribution of money and wealth in the US is basically fair, because everybody has an equal opportunity to succeed*" and "*Countries with more economic inequality usually function worse*". The equivalent graph for Survey 2 respondents is Figure H20

Figure 10: The Effect of Party Affiliation on Fairness Views and Inequality Externality Beliefs

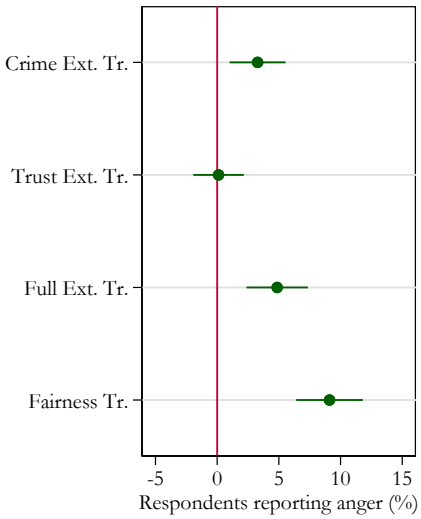


Note. Difference in pro-inequality sentiment (e.g. inequality does not increase crime, does not decrease trust, is fair) for Democrat-leaning respondents across selected externality and non-externality ("fairness") questions in Survey 1 where the only controls are for treatment groups. The same relation (every externality question has less polarization than any "fairness" question) holds for every question in Survey 1, and if we restrict the sample to only the control group. The relation also holds with a standard set of controls. Questions are largely split on pre-specified criteria or natural binary points (e.g. agree/disagree), keeping total shares close to 50% where possible. The equivalent graph for Survey 2 respondents is Figure H22. Error bars are 95% confidence intervals.

The psychological channel through which these arguments operate may also be different. Indeed, respondents who were shown the fairness video were significantly more likely to report their emotional reaction as anger than those who saw any other video.[80] While the absolute percentage of such respondents is relatively small (11.7%), the increase from the control video is highly statistically significant ($p < 0.0001$, t-test) and nearly twice as large as for any other video.[81] This asymmetry is not carried over for other emotions; the equivalent differences between the fairness and full externality videos are not statistically significant for *concern*, *surprise*, *indifference* and *confusion*.

This leads us to hypothesize that part of the difference in efficacy between these two videos, and thus potentially the two type of arguments, come from the extent to which they invoke anger in respondents.

Figure 11: Treatment Effects on Anger



Note. Error bars depict 95% confident intervals. Reference is control group. The full distribution of emotional reactions by treatment is found in Table I22.

6.2.4. Historical data

Finally we briefly explore the existing historical data to establish whether these respective beliefs could change over time. The only panel question we know of on U.S. citizens' inequality externality beliefs comes from the General Social Survey, where respondents were asked whether "large income differences are necessary for America's prosperity" in five waves between 1987 and 2021.[82] We compare the time trend of this question to people's beliefs about whether luck or hard work is more important for success (another GSS question) in Figure 12, overlaid with the bottom 50% income share.

While positive inequality externality beliefs have decreased from 34% in 1987 to 12% in 2021, mirroring the increase in U.S. income inequality in the same period [Saez and Zucman, 2020], beliefs about what determines success have not changed significantly. Other questions related to economic fairness principles also show very little movement in the period.[83] Although caution is suggested in interpreting this data, the implications for belief malleability and changes in the redistributive debate over time are intriguing.[84]

Taken together, the above discussion draws a strong contrast between inequality externality beliefs and other equity-based arguments for redistribution. We find circumstantial evidence for (i) inequality externality beliefs being particularly impactful for top-income individuals,

[80] Respondents were asked to self-report their emotional reaction to the video at the end of Survey 1.

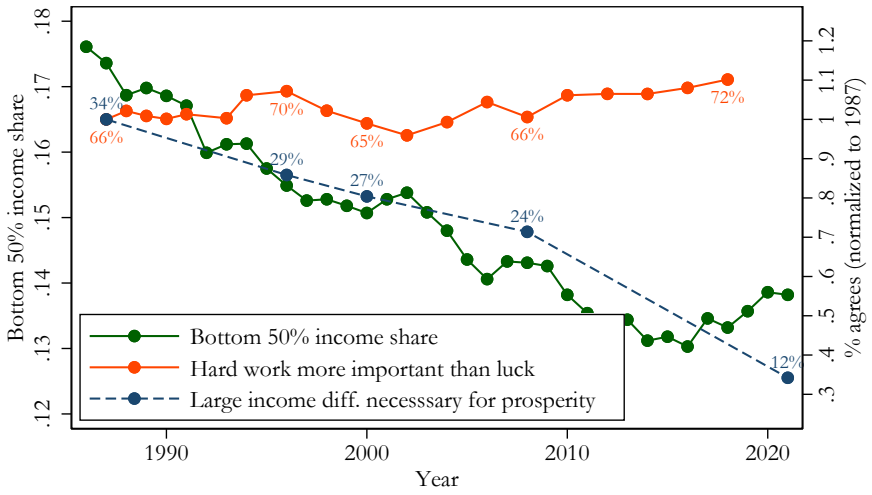
[81] Only 2.8% of respondents report anger from the control video. The second-highest video is the full externality video (7.8%); third-highest is the crime video (6.1%); fourth-highest, roughly equal to the control, is the trust video (2.9%). Respondents in the fairness treatment group are significantly more likely to report anger when compared to those in the full externality group ($p < 0.001$, t-test).

[82] We interpret this as a question about positive inequality externality beliefs through the channel of economic growth.

[83] See, for example, this question from Gallup on on whether "*the distribution of money and wealth [...] is fair*": https://news.gallup.com/poll/182987/americans-continue-say-wealth-distribution-unfair.aspx

[84] We invoke caution for two reasons. First, the questions are imperfect for our purposes. Second, our information experiment shows that broad economic fairness views *can* change with new information.

Figure 12: Historical externality data



Note. Data from the General Social Survey and the World Inequality Database (wid.world)

who are also more likely to hold negative inequality externality beliefs than to believe that the economic system is unfair, (ii) there being a larger consensus around inequality externality beliefs across political parties than comparable economic fairness views, (iii) fairness-based arguments leading to more anger in respondents than externality-based arguments, and (iv) time trends of inequality externality- and equity-based beliefs being significantly different over time. This leads to our third and final result:

Result #3

Inequality externality beliefs are structurally distinct from traditional equity-based reasons to redistribute.

This can be attributed to various theoretical rationales. Most of these rationales hinge on the fact that inequality externalities are consequentalist by nature. This differs from the philosophical ways in which we often think about inequality. When Amartya Sen asks *inequality of what?* [Sen, 1979], the point is that nearly all of us want equality in some dimension but the dimension for which we want equality differs. Some may want equality in liberty, whereas others want equality in incomes. These are intrinsic philosophical preferences that lead to diverging policy preferences. There is a different dimension to this question, however, which is one of *costs*. Inequality in various dimensions is likely to lead to societal changes, or externalities; these changes impose costs or benefits that may be orthogonal to our philosophical preferences. As we are at least somewhat pragmatic beings, this also changes the resulting policy calculus. An individual who wishes for equality in liberty may accept some taxation if the alternative is social unrest, for example.

It follows that inequality externalities could affect even completely self-interested individuals who receive no monetary or "philosophical" benefits from redistribution. Further, as redistributive decisions could be based on how to avoid these shared costs, externality-based reasoning

could lead to a broader consensus on what to do about economic inequality. In sum, inequality externalities present a clear efficiency dimension to the redistributive problem that is often absent in equity-based arguments.

The potential implications are large. Most importantly, the extent to which different countries have focused on the positive or negative consequences of inequality could shift redistributive equilibria. Suppose that inequality's negative consequences have historically played a larger role in Western European countries, for instance, and only recently became widespread in the United States. This presents a natural explanation for why Western European countries resisted the rise in income inequality seen in the United States since 1980. Indeed, beliefs in the *positive* consequences may have lead some Americans to accept and even hasten a rise in economic inequality. Now that such beliefs appear to have changed, opinions on redistributive questions may also have changed as a result.

Our results also lead us to speculate that part of the relatively polarized climate around redistribution in the United States could have been reduced had the political discussion focused more on inequality's externalities. As we have shown, such arguments appear likely to lead to less division across incomes and less anger.

These are naturally speculative hypotheses around which there is considerable uncertainty. Our main purpose in this article is to propose inequality externality beliefs as a meaningful determinant for redistributive preferences with unique properties from existing determinants. While this has intriguing ramifications for a wide variety of societal questions – as we discuss above – we expect future work to more precisely examine these ideas.

7. Conclusion

This paper marks the first positive analysis of both individuals' inequality externality beliefs and these beliefs' role as a determinant for redistributive preferences. Using two representative surveys of a total of 6,731 U.S. citizens we find that individuals believe inequality affects society through various ways, and that individuals largely believe that inequality has *negative* rather than *positive* effects on society. A large majority of individuals believe economic inequality increases crime (74%), decreases trust (67%), and reduces economic growth (51%), for example. In collecting these and other data points, this paper has thus created the first extensive data set of inequality externality beliefs in any country.

We have shown that these inequality externality beliefs are a causal determinant for redistributive preferences by using an exogenously provided information treatment. Three separate methods indicate that the magnitude of this determinant is large; externality beliefs have an effect on the same order of magnitude as broad fairness views in determining redistributive preferences. We also find indicative evidence that inequality externality beliefs are stronger determinants of redistributive preferences than traditional efficiency concerns about the distortive effects of taxation. As such, this paper presents the first strong evidence that individuals' beliefs about the *consequences of inequality* are impactful for their redistributive preferences.

The work further discussed how inequality externality beliefs have unique properties that are rarely seen in other redistributive determinants, particularly comparing to equity-based ideas. Inequality externalities appear to represent a distinct conversation about redistribution due to their consequentialist nature, focusing on positive arguments and affecting even self-interested

individuals. Ideas centered on these topics incite less anger than comparable economic fairness arguments; descriptive statements on the topic are relatively similar across political groups and appear to have changed significantly over time; and inequality externality beliefs have a particularly large impact on top-income individuals, indicating intriguing political economy effects. Overall, our conclusions could have broad implications for the redistributive equilibria in different countries.

Chapter 3

A Universe of Arguments

A Universe of Arguments*

Max Lobeck† and Morten Nyborg Støstad‡

Abstract

We present a novel survey-based methodology to evaluate the efficacy of classes of statements which mitigates the influence of researcher bias. We apply this methodology to redistributive arguments, where we elicit an unbiased sample of arguments based on either *fairness ideas* or *inequality's societal consequences* and evaluate their efficacy and emotional content across three surveys in the United States ($N_1 = 298$, $N_2 = 215$, $N_3 = 4010$). Our final "Universe of Arguments" has 160 redistributive arguments and a total of 32,300 argument evaluations. Respondents self-report significantly more anger ($p < 0.002$) in reaction to fairness arguments than to arguments based on inequality's consequences. This is partly driven by the average fairness argument having more emotional content than the average argument on inequality's consequences. While both types of arguments are broadly convincing, we find indications that individuals near the top of the income distribution are relatively more swayed by arguments on inequality's consequences.

*We are grateful to Emmanuel Saez, Marc Fleurbaey, Stéphane Gauthier, Thomas Piketty, Nicolas Jacquemet, Laurence Jacquet, Olof Johansson-Stenman, and seminar participants at the Paris School of Economics for helpful comments and suggestions. This work has been funded by the U.C. Berkeley James M. and Cathleen D. Stone Center on Wealth and Income Inequality. This study was pre-registered under AsPredicted #113265 and #115794. Version: November 30, 2023.

†University of Konstanz, Universitätsstrasse 10, 78464 Konstanz, Germany, Cluster of Excellence "The Politics of Inequality", Thurgau Institute of Economics, e-mail: max.lobeck@uni-konstanz.de

‡Paris School of Economics, 48 Boulevard Jourdan, 75014 Paris, France. Phone: +33766142152. e-mail: morten.stostad@psemail.eu.

1. Introduction

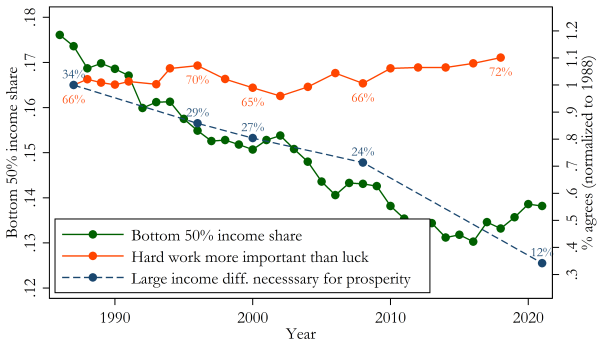
Many economic and societal problems hinge on the evaluation of various types of statements. An investor may choose whether to invest based on specific narratives, for example, or a voter may strengthen their political preference from a persuasive speech. It follows that researchers often want to understand the efficacy and emotional content of various types of statements. Evaluating the nature of a *type* of statement is challenging, however. The key problem is choosing which statements to evaluate within the broader class of statements. This choice is usually at best arbitrary, and at worse presents room for researcher bias. In either case, the generalizability of any resulting findings is compromised.

In this paper we present a novel method to avoid this problem. We use this method to compare the efficacy and emotional content of two types of redistributive arguments against each other. The method is centered around three steps. First, *eliciting* statements from individuals through a carefully-worded prompt, attempting to gather an unbiased sample of statements from the distribution of interest. Second, using an independent set of survey respondents to *quality check* the sample for statements that do not fit the prompt or are otherwise unwanted. Third, using another independent set of survey respondents to *evaluate* the resulting "Universe of Arguments". The aim of the method is to reduce concerns of researcher bias into well-known sources (the phrasing of questions and sample selection).

We apply this method to the question of redistribution, where we evaluate the efficacy and emotional content of two types of redistributive arguments in the United States. These two types of arguments are arguments based on either *fairness ideas* or *inequality's societal consequences*. We will describe inequality's consequences as "inequality externalities", following Støstad and Cowell [2021] and Lobeck and Støstad [2023].[1] In short, "fairness arguments" are equity-based redistributive arguments and focus on whether individuals *deserve* their income or wealth. "Externality arguments" are efficiency-based and focus on economic inequality's consequences. Examples include increasing crime, changes to the economic growth rate, or deteriorating social cohesion.

Our interest in these arguments is partially motivated by Figure 1. As shown in the graph, the income share of the bottom 50% has sharply decreased in the United States since 1987. Following this trend, the share of U.S. citizens that believe that "large income differences are necessary for prosperity" – a measure of positive inequality externality beliefs – has significantly decreased during the same years. The share of citizens who believe "hard work is more important than luck" in becoming rich, however – a measure of economic fairness views – has

Figure 1: Fairness and externality beliefs over time



Note: Data from wid.world and the General Social Survey.

[1] Economic inequality is affected by individuals' market decisions. If economic inequality affects relevant societal outcomes, it follows that economic inequality is an externality.

stayed relatively constant (or even increased). This indicates that fairness-based and inequality externality-based statements potentially have different malleability and responsiveness to actual economic inequality. If redistributive arguments based on these ideas are also functionally distinct in other ways – leading to different amounts of anger, for example – it could explain both historical and cross-country differences in redistributive debates.

Our approach combines three surveys. In a first survey respondents were asked to write arguments for or against redistribution. Each respondent received two prompts in randomized order, which were identical except for a request that the argument was based on either fairness ideas or inequality's societal consequences. We gather a total of 596 arguments from 298 respondents. We then used a second survey to ensure the final sample of arguments were on-topic and sensible. This left 190 arguments in our "universe of arguments", of which 160 were pro-redistribution and 30 were anti-redistribution. We intentionally over-sampled Democrats and Independents to target pro-redistributive arguments, which were our main focus. We then showed this "universe of arguments" to a separate sample of 4010 respondents to evaluate the arguments for convincingness and emotional content, specifically anger.

We have three main findings. First, fairness arguments make respondents self-report significantly more anger. This is driven by respondents that *agree* with the arguments, and is partly due to fairness arguments appealing to emotions more regularly than externality arguments. Second, we find that both types of arguments are generally convincing – reinforcing the finding from Lobeck and Støstad [2023] that externality beliefs are comparable to fairness views as determinants for preferences for redistribution. Third, we find indications that high-income respondents ($> \$75,000$) disproportionately find externality arguments convincing – although this is not statistically significant for the pre-specified income split of $\$50,000$, or the higher $\$100,000$ (where sample size becomes low).

In addition, we find no evidence that Republicans find either type of redistributive argument (fairness or externality) disproportionately convincing as compared to Democrats. This differs from the descriptive beliefs described in Lobeck and Støstad [2023], where fairness views are more polarized than externality beliefs across political groups (a finding we replicate).[2] We hypothesize that this is because individuals who are against redistribution on principle are not swayed by nearly any argument used to motivate this redistribution. In other words, although most people would like a more equal economic distribution [Norton and Ariely, 2011], the *method* with which this is achieved is often crucial. To put it differently, it could be that many individuals are generally against *any* argument for state-lead redistribution, even when they share the overarching goal of reducing economic inequality.

In sum, these findings indicate that cross-country and historical differences in redistributive debates could be at least partly explained by what *type* of redistributive arguments are common in the country or time period. The affective polarization around redistributive debates in the current United States could be due to a focus on fairness over externality motivations, for example. Likewise, support for redistributive policies in the upper class could be explained

[2]That is to say that the party difference across various questions is consistently higher for fairness questions than externality questions. For example, Democrats and Republicans are both likely to believe that inequality increases the amount of crime. However, Republicans are much more likely to believe that hard work leads to success, or that the income distribution is overall fair.

by whether the most common arguments for redistribution are based on fairness or externality ideas.

This paper primarily relates to the large literatures on preferences for redistribution and survey-based methodology. The literature examining fairness views is large; Cappelen et al. [2007], Almås et al. [2020] and Stantcheva [2021] are among the many papers establishing a connection between individuals' fairness views and their preferences for redistribution. The externality angle is examined theoretically by Alesina and Giuliano [2011] and Støstad and Cowell [2021], and empirically by Lobeck and Støstad [2023].

Most related to this paper is Lobeck and Støstad [2023], which shows that both fairness and externality-based arguments causally change preferences for redistribution. This paper also shows that a fairness-focused video has respondents self-reporting more anger than three externality-focused videos. These results are only evaluated for those specific video treatments, however; a multitude of specific design choices could have impacted results. In contrast, this paper evaluates a larger set of arguments sampled in an unbiased fashion from the distribution of such arguments in the population.

Several other papers have similar research designs to Lobeck and Støstad [2023] [e.g. Kuziemko et al., 2015, Stantcheva, 2021], where the main research goal has been to establish the existence of a causal connection between some belief (e.g. fairness beliefs or policy knowledge) and redistributive preferences. This paper instead tries to elucidate other distinguishing features of these connections – such as their strength and emotional content – by using a larger sample size of arguments that are not affected by research design choices. In effect, we evaluate 160 unbiased arguments instead of a handful of specifically-designed video treatments. This allows us to extrapolate our findings to a larger degree than is usually possible. Finally, we also relate to a literature around narratives and their relation to economic outcomes, among them Alesina et al. [2018a], Roth et al. [2020] and Andre et al. [2022].

The paper is structured as follows. Section 2 introduces the method, survey design, and sample collection. Section 3 discusses the results. Section 4 concludes.

2. Methodology

2.1. Theoretical framework

Suppose a statement F can be represented by its observable and unobservable properties X in a multi-dimensional space such that we can write $F(X)$ to fully classify the statement. These properties can represent anything that changes the nature of the statement, for example complexity, factual accuracy, or emotional content. An individual j evaluates the statement to decide on an individual-specific outcome Y_j, for example whether the statement convinces the individual or evokes an emotional reaction. This evaluation is done through an individual-specific evaluation function ϕ_j, such that $Y_j = \phi_j(F(X))$.

The researcher is interested in the distribution $\mathcal{G}$ of these Y_j for some population of individuals j, conditional on certain properties being of a specific pre-determined type which we designate as $X_{type} = X_W$. This could designate that the argument is a redistributive argument based on fairness ideas, for example. Formally, the researcher wants to find an estimate of $\mathcal{G}(Y|X_{type} = X_W)$. Importantly, the remaining X_{-type} are not specified and can vary freely. Obtaining this

distribution would allow the researcher to analyze the properties of this *type* of statements for the outcome in question.

There are many potential reasons for why this could be of interest. It is particularly simple to create examples when the researcher has such estimates for several different types of statements. For example, the researcher could wish to know whether the truthfulness of an political argument drives polarization in voting decisions V. The researcher would define X_T as related arguments that are truthful and X_F as related arguments that are false. By comparing the variance $Var\left(\mathcal{G}(V|X_{type} = X_T)\right)$ and $Var\left(\mathcal{G}(V|X_{type} = X_F)\right)$, the research question could be answered. Another researcher could explore which type of investment advice spurs more investment, what type of monetary policy statement is seen as more convincing, and so on.

To do so, however, the researcher needs to find an unbiased estimate of $\mathcal{G}(Y|X_{type} = X_W)$. It is not enough to pick specific statements and modify them slightly to satisfy the requirements imposed by X_W, as there are two related problems. First, it is often difficult to appropriately modify only the characteristics in X_{type} without also affecting other X (for example the length, complexity, or priming of the argument). Still, in the context of survey experiments, this can at times be corrected for by only modifying a very small part of the statement/video/argument. The second issue is more intrinsic, however. Such an approach introduces potential bias by specifying the other X. This is problematic because there could be significant interactions between the type of the argument and other argument properties, and it is not possible for the researchers to randomly allocate all the unobserved and potentially unknown X due to the infinite and unknowable nature of these multi-dimensional properties. To illustrate we return to the above example of truthfulness and polarization in voting behavior. Suppose that, unbeknownst to the researcher, truthfulness only drives polarization in voting behavior if the argument is particularly sophisticated. The researcher, being an academic, writes sophisticated arguments without considering this as a specific choice, varies the truthfulness, and presents the arguments to a representative sample. They then (erroneously) conclude that the hypothesis is generally true. This example illustrates the potential problem; the researcher must always make a choice of *which* statements to show respondents. The conclusions drawn from the test are only valid under the specific X_{-type} the researcher has specified. In information experiments, this indicates that specific design choices make it difficult to establish external validity.

We propose instead to draw a random sample of statements F from the statement distribution $\mathcal{F}(X|X_{type} = X_W)$ where the only restriction is that $X_{type} = X_W$, letting the remaining X_{-type} vary freely. The goal of this step is to find an unbiased estimate of the distribution of statements when $X_{type} = X_W$, which can be evaluated and finally used to test hypotheses for $\mathcal{G}(Y|X_{type} = X_W)$. Although this increases the amount of noise significantly, it also allows much broader conclusions from the resulting $\mathcal{G}(Y|X_{type} = X_W)$. Hypothesis testing on this distribution would not suffer from the same issues of external validity discussed above.

Assuming random sampling from the statement distribution and unbiased estimation of the resulting $\mathcal{G}$, the researcher could test hypotheses on these types of statements without further restrictions. There are many potential examples of the usefulness of this approach, some of which we have already mentioned. In the remainder of the paper we will explore outcome distributions from two types of redistributive arguments in order to understand resulting social dynamics and

economic policy-making.

2.2. Survey methodology

To find and evaluate the distributions discussed above we conducted three surveys between October 24th and December 23rd 2023 in the United States. Survey 1 and Survey 2 were conducted on the survey platform *Prolific* ($N_1 = 298$, $N_2 = 215$), and Survey 3 was conducted by the survey company *Dynata* ($N_3 = 4010$).

The method we describe follows a three-step process. Each step corresponds to one survey. The first step is to *elicit* the type of statements or arguments the researcher wishes to evaluate, constructing an unbiased estimate of $\mathcal{F}(X|X_{type} = X_W)$. Due to imperfect responses, this sample will have potential contamination from arguments where $X_{type} \neq X_W$. This necessitates a second step to *quality check* the statements, removing any off-topic or nonsensical statements from the sample and ensuring that $X_{type} = X_W$. The third step is to let a distinct representative sample *evaluate* the arguments, constructing $\mathcal{G}(Y|X_{type} = X_W)$. All three steps are conducted by survey respondents who are only given minimal prompts and are unaware of the testable hypotheses.

Representativity It should be noted that strict representativity of the U.S. population was not enforced in Surveys 1 and 2. This was an intentional choice to reduce survey costs. In Survey 1, individuals were allowed to choose whether they wrote pro-redistributive or anti-redistributive arguments. As right-wing respondents were unlikely to write pro-redistributive arguments, which were the main focus of the study, we intentionally over-sampled Democrats and to a lesser degree Independents. This means that the resulting distribution of arguments has been largely written by left-wing respondents. We also did not specifically quota on other demographic dimensions beyond gender, again to reduce survey costs. We consider this a potential limitation of our survey. In Survey 2, the primary purpose was simply to quality check the data; as this is a simple task where the results are likely similar across different demographic groups, we again opted to not enforce strict representativity.

Survey 3, which evaluates the arguments gathered from Surveys 1 and 2, is broadly representative of the U.S. population. We pre-specified quotas across gender, age, political affiliation, income groups, region, and race. These quotas were kept, which ensures a diverse and representative group of respondents for the argument evaluations.

2.3. Survey 1 (Elicitation)

Survey 1 was conducted between November 16 and November 27 2022 with a total sample of $N_1 = 298$.[3]

Respondents were informed before agreeing to the survey that they would be asked to *"write arguments for or against economic redistribution"*, and that these arguments would be shown to other survey respondents. After being asked about their political leaning, respondents were asked two sets of questions in random order. Each set of questions had the goal of eliciting a redistributive argument from respondents. One set of questions focused on fairness ideas, and the other focused on inequality externality ideas.

[3]Two-thirds of the data collection ($N = 199$) was done on November 16th and 17th. The last third ($N = 99$) was done on November 27th, after the first round of quality checks resulted in too few final arguments.

In each set, respondents were first asked which type of argument they would prefer to write; a pro-redistribution or anti-redistribution argument based on the idea in question (fairness or inequality's consequences).[4] Then they were prompted to write a brief argument for such ideas, not exceeding three sentences.[5] Respondents were incentivized to write high-quality arguments. Specifically, respondents were informed that the survey payout was doubled (from 0.5 to 1.0) if the respondent's arguments was "chosen" (passed the quality check) and was found convincing by a majority of respondents. The median survey time was 6 minutes and 43 seconds.

The two pro-redistributive prompts are shown below:

Question text: Fairness elicitation

> Imagine you want to convince a friend to support **more** economic redistribution with an argument about how this would **be fair**. Please write a **brief** (3 sentences maximum) argument below.
> You can make any argument you want as long as it relates to economic fairness issues (high incomes, low incomes, which people deserve income increases, and so on). You don't need to explicitly use the word "fair" unless you want to, but the argument should be about fairness.
> **Remember that convincing arguments will be rewarded – if your arguments are found to be convincing, your survey payout will be doubled.**
> **So, why should we redistribute more?**

Question text: Inequality externality elicitation

> Imagine you want to convince a friend to support **more** economic redistribution with an argument about how economic inequality has **negative consequences** for society. Please write a **brief** (3 sentences maximum) argument below.
> Please do not discuss economic fairness issues, but instead focus your argument on how inequality affects societies in other ways. You can for example make arguments for redistribution about how economic inequality affects the amount of [*two of crime, economic growth, corruption, innovation, social unrest, trust, political polarization*], or society overall – but please use your own words and ideas.
> **Remember that convincing arguments will be rewarded – if your arguments are found to be convincing, your survey payout will be doubled.**
> **So, why should we redistribute more?**

See Appendix III.A.1 for the anti-redistributive versions.[6] Respondents were required to answer both prompts. As such, we had a total of 596 arguments. Roughly 80% of these

[4] Question text: *Which of these do you prefer to make an argument for?*
Fairness version: "Why we should redistribute **[more/less]**, as the market distribution of resources is **[fair/unfair]**.
Externality version: "Why we should redistribute **[more/less]**, as inequality changes society for the **[worse/better]** (more inequality → a [worse/better] society in various ways)".

[5] Any argument above three sentences was automatically shortened to fit this restriction.

[6] 92% of respondents choose to write two pro- or anti-redistributive arguments (and not one of each)

arguments were pro-redistribution. As mentioned previously, this was by design; we intentionally focused on pro-redistributive arguments because the majority of individuals believe in *negative* but not *positive* consequences of inequality [Lobeck and Støstad, 2023]. Indeed, as we show in Section 2.4, respondents generally struggled with writing coherent anti-redistributive arguments based on inequality externality ideas.

2.4. Survey 2 (Quality check)

Survey 2 was conducted between November 26th and November 27th 2022 with a total sample of $N_2 = 215$. Respondents were informed before agreeing to the survey that they would be asked to "*evaluate 16 arguments to make sure that they are sensible and on-topic*", and that the arguments would be about economic redistribution.

In recent literature [e.g. Andre et al., 2022] this type of task is often performed by research assistants who are not informed of the hypotheses to be tested. This is due to concerns of researcher bias; if the researchers themselves conducted the task, it is easy to imagine how they could consciously or subconsciously bias the results.

We posit that research assistants, in the absence of rigorous obfuscation mechanisms, are also likely to be affected by similar biases. There are several reasons for this. First, research assistants have personal incentives for the research to be "successful", as this is likely to lead to stronger reference letters, a potentially better publication resume, or a better relationship with the employer. Second, any one research assistant will see hundreds or thousands of responses to be coded, making it easier to infer research hypotheses. Third, research assistants are usually aware of the research portfolio of their employer. Put together, research assistants have both a motive and often means to bias research outcomes.

We thus propose to crowd-source the quality check to a large number of outside individuals who (i) have no incentive for the research to be "successful", (ii) will only see a small number of statements each, (iii) are not aware of the research portfolio of the responsible academics.

We enlisted *Prolific* respondents for this task. Respondents were shown an argument and told it was written by another survey respondent. They were then first asked whether the argument was overall sensible and on the correct general topic (for either *more* or *less* redistribution depending on how the author of the question classified it). They were then asked to classify the argument as being about either fairness ideas, how economic inequality changes something in society, or neither. We show full question texts in Appendix III.A.2. Each of the 596 arguments was evaluated by between 4 and 8 respondents, on average 5.5, and each respondent in Survey 2 (215 in total) evaluated 16 arguments.

We pre-specified that we would find 200 arguments, 160 of which would be pro-redistribution and 40 of which would be anti-redistribution (half of each type being on each idea). We also pre-specified two criteria for arguments to be included in this final "Universe of Arguments". First, arguments needed to be evaluated by 75% or more of respondents as making sense and being on the correct overall topic. Second, arguments needed to be evaluated as on the correct idea (fairness or externalities) by 75% or more of respondents. We pre-specified that these criteria would be lowered if our initial sample failed to reach these goals. A slight modification of the criteria for the first data quality check (that the question makes sense) was done for the externality arguments, where we decreased the pass threshold to 71%. A larger modification

was done for the second data quality check (classifying the correct topic), where the threshold for both types of arguments was lowered to 60%.

Overall, however, descriptive data on both data quality check questions were similar across fairness and externality arguments.[7]

After Survey 1 and Survey 2 we were then left with 200 arguments. 160 of these arguments are part of the main "Universe of Arguments". Our designated "Universe of Arguments" thus consists of pro-redistributive arguments; 80 of which focus on fairness ideas and 80 of which focus on inequality externality ideas.

We note that the methodological choices to keep this "Universe of Arguments" relatively unbiased comes at the cost of significant noise. We only offer minimal guidance on what kind of argument to write in Survey 1, and respondents in Survey 2 might also have differing notions of what "makes sense", or what exactly is meant by fairness or inequality's consequences. Arguments are also relatively long at up to three sentences; they have a median character length of 279 characters. While the methodology should allow this noise to be relatively unbiased, this leads to potential power issues, and is the key limitation to our method. However, the same methodological choices allows us to be more confident when results *are* statistically significant. In sum, the method presents a relatively unbiased sample of arguments that may have a high variance.

We now move to Survey 3, which *evaluates* all 200 arguments.

2.5. Survey 3 (Evaluation)

Survey 3 was conducted between December 8th and December 30th.[8]. Before the survey respondents were told that the survey had been authored by a non-partisan group of economists and that they would be asked about their "attitudes on several topical issues". The survey began with demographic questions before eliciting pre-treatment fairness views, externality beliefs, and redistributive preferences. Then respondents were then told that they would be asked to evaluate ten different arguments on redistribution (one shown at a time).

Each respondent was asked three questions per argument. First, whether they were "*personally convinced*" by the argument or statement.[9] Second, whether they would "*be willing to have a longer conversation with this person about these ideas*".[10] Third, whether "*a discussion about this argument could provoke an emotional reaction like anger or agitation in you*". For the third question respondents could indicate whether the anger was due to agreeing with the argument

[7]The one exception was anti-redistributive arguments about inequality externalities, which were often classified as fairness arguments by Survey 2 respondents. As such, we lowered the quota to 10 such arguments (from 20). We instead added 10 pro-redistributive "expert" arguments from 5 pairs of "experts" (one of each type per "expert"). These "experts" were Barack Obama, Nicholas Kristof, Bernie Sanders, Tucker Carlson, and ChatGPT (which was fed the same question as Survey 1 respondents). We will not discuss the results from this further in this paper, but note it here as it was a change from the pre-analysis plan. Respondents who received these arguments were not told who had written/spoken the argument and, unlike the other arguments, were not told that it was written by another survey respondent.

[8]89% of responses were collected before December 15th.

[9]Options were [Very convinced/Convinced/Neither convinced nor unconvinced/Unconvinced/Very unconvinced].

[10]Options were [Yes/No].

or thinking it was nonsense.[11] On average, each argument was viewed by 204.4 respondents.[12]

Respondents' answers to these questions are our main outcomes. Afterwards we also elicited post-treatment externality beliefs, fairness views, and redistributive preferences. The average survey time was 17 minutes and 56 seconds. In sum there are 80 pro-redistributive arguments of each type (160 in total), and a total of 32,300 evaluations.

3. Results

We show brief descriptive statistics from each type of argument in the Universe in Table 1, noting the shares of each type of statement that include a positive statement and/or an emotional appeal.[13] As arguments can be relatively long, many arguments include both.

Table 1: Universe of Arguments: Descriptive Results

	(1) Fairness arguments	(2) Externality arguments	Overall
Positive statement	69.41%	95.00%	82.21%
Emotional appeal	94.94%	18.92%	56.90%
# Arguments	80	80	160
# Evaluations	16,140	16,160	32,300

The argument types show significant differences, which is notable due to the relatively unbiased method of elicitation. The largest difference is for the emotional content of the argument; while 95% of fairness arguments are emotionally charged, only 19% of externality arguments can be similarly classified. Externality arguments are also more likely to be based on positive statements; 95% of such arguments were based on phenomena that are observable in principle. To illustrate we show two examples below;

Positive non-emotional argument: *"Redistribution of economic resources promotes equality and societal balance. This redistribution would reduce crime and social unrest while improving societal innovation and prosperity."*

Normative emotional argument: *"For most supporters of economic redistribution, it really just comes down to a moral judgement that it is wrong for one person to have more than they need while others do not have enough to survive or thrive."*

We will now compare fairness-based and externality-based arguments in the Universe over various outcome dimensions from evaluations (convincingness, anger, and so on). To do so we will first find the average outcome for each argument after adding a pre-specified set of controls.[14]

[11] Options were [Yes, because I think the argument is nonsense/Yes, because I agree with the argument/Partly, because I think the argument is nonsense/Partly, because I agree with the argument/No, not really/No, not at all].

[12] Of the arguments used in the "Universe", the average was 201.9 respondents.

[13] In the current version this is classified by the authors.

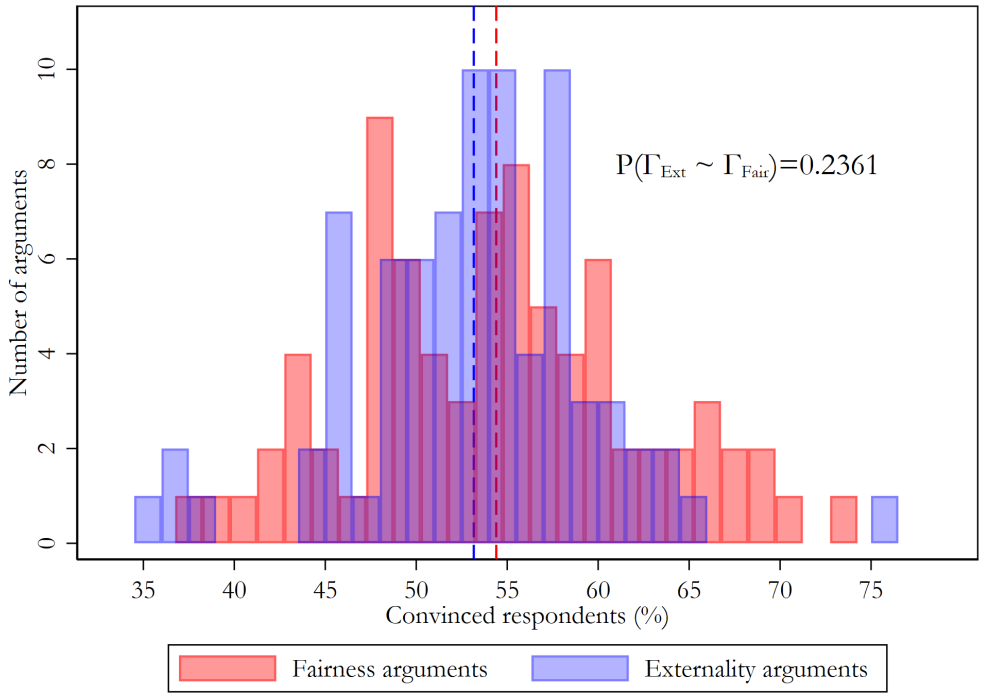
[14] Our standard set of controls are binary variables for leaning Republican over Democrat, gender, self-identifying as black, self-identifying as non-white, four income groups ($0-$25,000, $25,000-$50,000, $50,000-$100,000, $100,000+), six age groups (20-29, 30-39, 40-49, 50-59, 60-69, 70+), having a college education, being unemployed, not being in the work force (e.g. students or seniors), and region (South, West, Northeast, Midwest).

We will then perform a Kolmogorov-Smirnov test of the resulting shares to find the likelihood that both types of arguments were drawn from the same distribution.

3.1. Convincingness

The average percent of respondents convinced by each argument is shown in Figure 2. Although fairness arguments are on average slightly more convincing, the difference is not statistically significant. The percentage of respondents who are convinced by the average fairness-based argument is 54.4%. The corresponding percentage for the average externality-based argument is 53.1%. The standard deviation of the fairness arguments (7.8%) is somewhat higher than that of the externality arguments (6.7%), indicating more dispersion in the quality of fairness arguments.

Figure 2: Convinced respondents across argument type



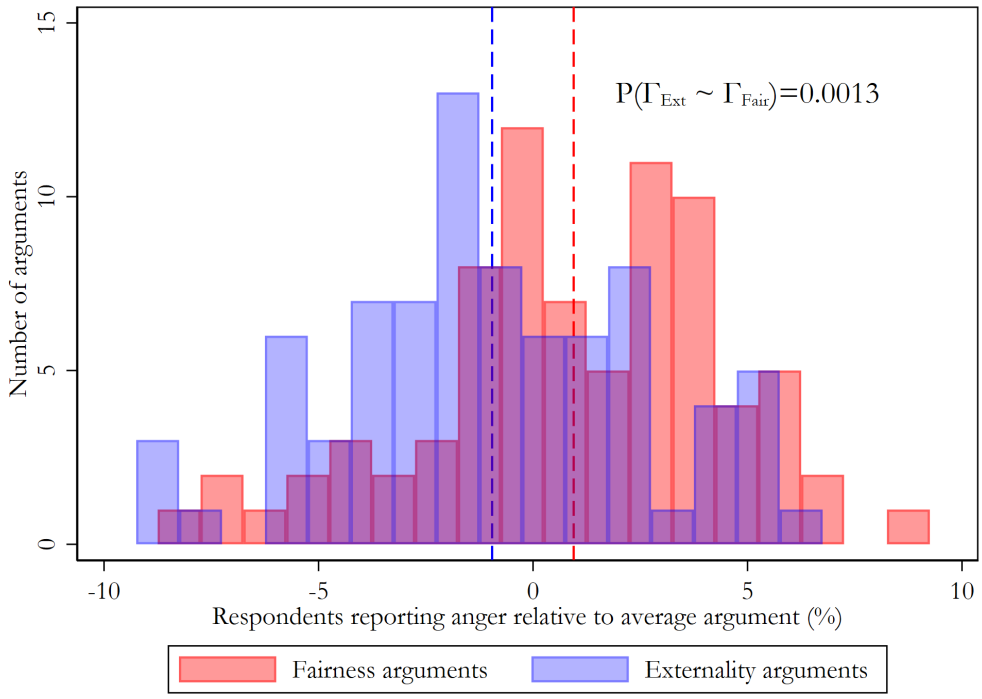
Note. The percentage of respondents reporting to be "Convinced" or "Very convinced" by each argument in the Universe. There are 160 arguments, and each argument was viewed by an average of 202 respondents. In total there are 32,300 observations.

3.2. Anger

The average percent of respondents reporting anger relative to the average argument is shown in Figure 3. Respondents are significantly more likely to report anger in response to a fairness argument than to an externality argument ($p < 0.002$). This difference is visible in the two histograms in Figure 3 and in the pre-specified regression in Table 2. This is largely driven by respondents who report anger because they *agree* with the argument, as shown in Figures 4 and B1.

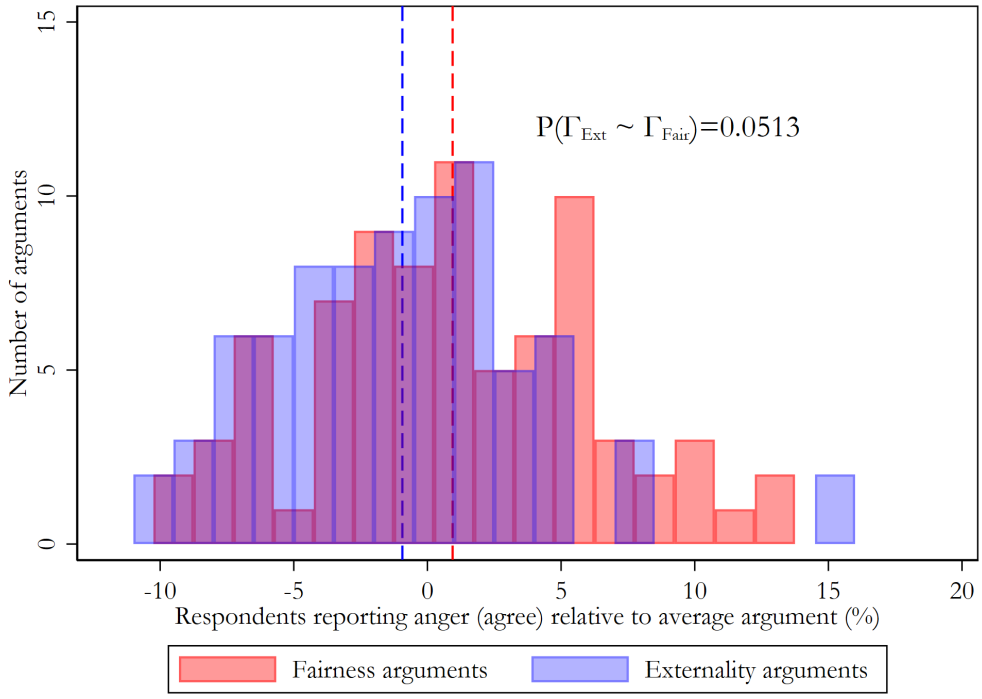
This difference is largely driven by fairness arguments being more *emotionally charged*. This can be shown through adding controls for which arguments appeal to *emotions* and/or *positive statements* (Table 2) to the pre-specified regression on evaluation outcomes with standard

Figure 3: Self-reported "anger or agitation" across argument type



Note. The percentage of individuals responding "Yes" or "Partly" to a question about whether "*a discussion about this argument could provoke an emotional reaction like anger or agitation in you*". There are 160 arguments, and each argument was viewed by an average of 202 respondents. In total there are 32,300 observations.

Figure 4: Self-reported "anger or agitation" due to agreement across argument type



Note. The percentage of individuals responding "Yes, because I agree with the argument" or "Partly, because I agree with the argument" to a question about whether "*a discussion about this argument could provoke an emotional reaction like anger or agitation in you*". There are 160 arguments, and each argument was viewed by an average of 202 respondents. In total there are 32,300 observations.

errors clustered on the individual-level. While both emotionally charged content and positive statements increases the fraction of respondents who self-report anger,[15] the difference between fairness-based and externality-based arguments is entirely driven by whether the argument is emotionally charged.

Table 2: Anger: Regression results

	(1) dAnger b/se	(2) dAnger b/se	(3) dAnger b/se	(4) dAnger b/se	(5) dAnger b/se
ExtArg	-0.019***	-0.019***	-0.001	-0.021***	-0.003
	(0.006)	(0.005)	(0.009)	(0.006)	(0.009)
Emotions			0.023***		0.026***
			(0.009)		(0.009)
Factual				0.010	0.014*
				(0.007)	(0.008)
Controls	No	Yes	Yes	Yes	Yes
R2	0.00	0.02	0.02	0.02	0.02
Observations	32300	32300	32300	32300	32300

Note. This table represents the regression coefficients for the pre-specified anger regression, with additional regressions including dummies for whether the argument was positive or emotional. Controls are binary variales for leaning Republican over Democrat, gender, self-identifying as black, self-identifying as non-white, four income groups ($0-$25,000, $25,000-$50,000, $50,000-$100,000, $100,000+) six age groups (20-29, 30-39, 40-49, 50-59, 60-69, 70+), having a college education, being unemployed, not being in the work force (e.g. students or seniors), and region (South, West, Northeast, Midwest). *Significance levels*: *10%, **5%, ***1%.

3.3. Convincingness: Heterogeneity

Lobeck and Støstad [2023] finds that there is more consensus for externality beliefs than fairness views across both political groups (Democrats and Republicans) and economic status (wealth, income groups).

To evaluate whether this holds true for redistributive arguments we check whether the differences in finding arguments convincing is smaller across different income groups and political affiliations for externality arguments than fairness arguments.

3.3.1. Heterogeneity in incomes

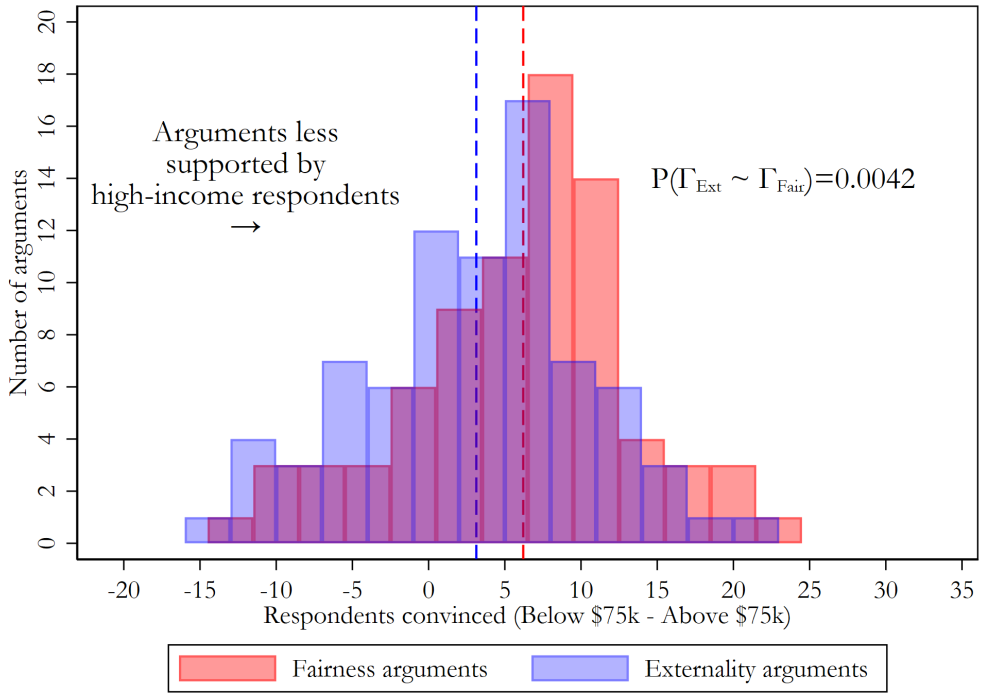
Lower-income respondents are more likely to be convinced by arguments in our sample. Respondents with yearly incomes above $\$75,000$ are convinced by 49% of arguments, while those below are convinced by 56% of arguments. This is strongly significant in a t-test ($t = 20.63$).

In the academic literature, fairness views are regularly found to be polarized across incomes – higher-income respondents believe the economic system is more fair. Lobeck and Støstad [2023] noted that this was in contrast to inequality externality beliefs, which were found to be relatively constant across income groups. Our initial hypothesis was thus that this descriptive belief difference would translate into differential support for redistributive arguments on these topics, where higher-income respondents would be relatively more likely to support redistributive arguments based on inequality externalities than those based on fairness ideas.

We find some evidence of this. In Figure 5 we show the difference in convincingness across these two income groups (the 56% vs. 49% difference from above), separating fairness and

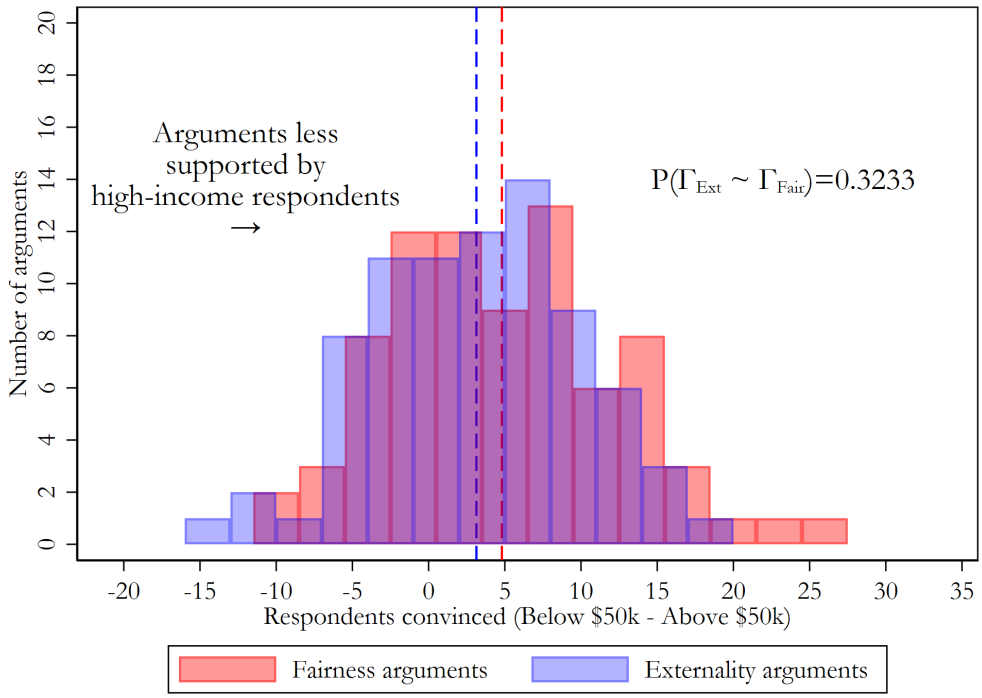
[15] This is only due to respondents who agree with the argument, as we show in Tables C1-C2.

Figure 5: Heterogeneity in convinced respondents for incomes above/below $75,000 across argument type



Note. The percentage point difference for respondents with incomes above and below $75,000 reporting to be "Convinced" or "Very convinced" by each argument in the Universe.

Figure 6: Heterogeneity in convinced respondents for incomes above/below $50,000 across argument type



Note. The percentage point difference for respondents with incomes above and below $50,000 reporting to be "Convinced" or "Very convinced" by each argument in the Universe.

externality arguments and adding standard controls.[16] The externality arguments have a smaller gap between high-income and low-income respondents than fairness arguments ($p < 0.005$). However, this is not the case for the specification we pre-specified, which used yearly incomes above and below $\$50,000$ ($p = 0.3233$, see Figure 6). It is also not true for yearly incomes above and below $\$100,000$ (where sample size is smaller, $p = 0.1160$, not shown). As such, our evidence is only indicative.

3.3.2. Heterogeneity in political affiliation

Democratic voters are unsurprisingly more likely to be convinced by pro-redistributive arguments in our sample. Respondents who identify as Democrats are convinced by 68% of the pro-redistributive arguments, while respondents who identify as Republicans are convinced by 41% of the pro-redistributive arguments.

In Lobeck and Støstad [2023], fairness views were found to be more polarized across political affiliations than externality beliefs. Our initial hypothesis was thus that this difference would translate into differential support for redistributive arguments – where Republicans respondents would be relatively more likely to support redistributive arguments based on inequality externalities than those based on fairness ideas.

We find no evidence of this, which we show in Figure 7. The average difference in percentage of convinced respondents across political affiliation is essentially identical for fairness and externality arguments. This stands in contrast to the descriptive evidence from both Lobeck and Støstad [2023] and our own survey – which, while weaker than Lobeck and Støstad [2023], shows the same descriptive pattern of party polarization being higher for fairness questions than externality questions (not shown).

Having a longer conversation We also asked respondents whether they would be willing to have a longer conversation with the individual who wrote the argument to discuss these ideas. We did not find any significant differences across argument types, which we show in Appendix III.B.2.

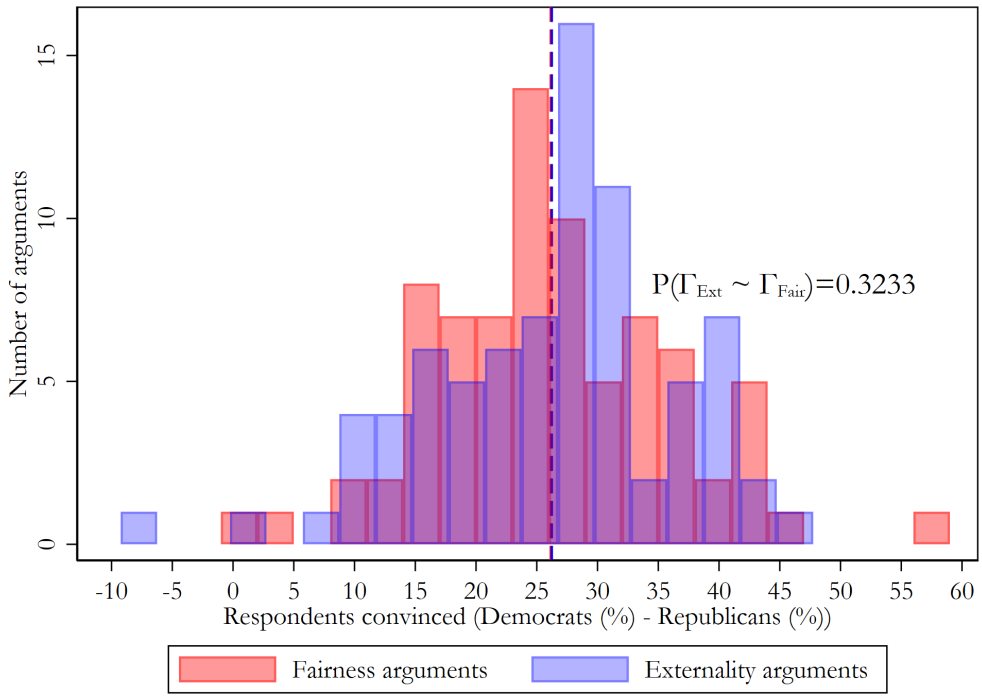
4. Conclusion

Evaluating the efficacy and emotional content of various types of arguments is crucial to understanding many economic and societal problems. However, the evaluation process is often fraught with researcher bias, compromising the generalizability of any resulting findings. This paper presented a novel survey-based methodology designed to gather a so-called "Universe of Arguments" and intended to minimize researcher bias in the evaluation of different classes of statements or arguments.

We applied this methodology to the question of redistribution and evaluated the efficacy and emotional content of redistributive arguments based on either fairness ideas or inequality's externality effects. Our primary result is that fairness arguments incites significantly more anger in respondents, driven by respondents who agreed with the argument. This is driven by fairness arguments being more likely to appeal to the emotions of the reader. We also find that both types of arguments are broadly convincing.

[16] For the income-specific analysis presented here we do not include the income groups.

Figure 7: Heterogeneity in convinced respondents for political affiliation across argument type



Note. The percentage point difference for Democrats and Republicans reporting to be "Convinced" or "Very convinced" by each argument in the Universe. There are 160 arguments, and each argument was viewed by an average of 202 respondents. In total there are 32,300 observations.

We also find indicative evidence that externality arguments are relatively more convincing than fairness arguments at the top of the income distribution. Meanwhile, in contrast to the descriptive evidence from Lobeck and Støstad [2023], we found no evidence that Republicans find either type of redistributive argument disproportionately convincing as compared to Democrats.

Overall, this study's findings indicates that differences in redistributive debates could be influenced by the *type* of argument the redistributive debate focuses on. In particular, cross-country and historical differences in affective polarization could potentially be partly explained by the extent to which the redistributive debate has focused on fairness-based or externality-based issues.

Résumé détaillé en français

Introduction générale

"*Dans un État qui veut être préservé du plus grand de tous les fléaux, [...] il ne doit y avoir parmi les citoyens ni pauvreté extrême, ni excès de richesse, car l'une et l'autre produisent ces maux.*". – Platon (360 av. J.-C.)

Cette thèse se concentre sur les conséquences de l'inégalité économique et la redistribution à travers des explorations théoriques et empiriques. Le débat controversé sur l'inégalité économique et la redistribution occupe depuis longtemps une place importante dans la sphère publique. Au cœur de ce débat se trouve la question de savoir *pourquoi* nous devrions nous préoccuper de l'inégalité économique. Le dilemme traditionnel est celui entre l'équité et l'efficacité ; la réduction des inégalités est bénéfique pour redistribuer des riches vers les pauvres, ce qui favorise l'équité. Toutefois, cela a un coût en termes d'efficacité, car la réduction des inégalités par le biais de la fiscalité entraîne une perte sèche potentiellement importante. Cet arbitrage entre équité et efficacité est au cœur des politiques publiques et de la recherche économique axée sur l'inégalité depuis des décennies.

Dans cette thèse, j'envisage une approche différente. Il est souvent suggéré dans le débat public que l'inégalité économique affecte la société d'une manière ou d'une autre. Ces *conséquences de l'inégalité économique* pourraient inclure des problèmes qui nous affectent tous, par exemple une augmentation des troubles sociaux, des systèmes politiques dysfonctionnels ou une détérioration de la confiance générale. Elles suggèrent une raison potentiellement importante de redistribuer, qui est également présente en l'absence de toute préoccupation d'équité. L'arbitrage traditionnel entre l'équité et l'efficacité se transforme ainsi en un exercice d'équilibre plus complexe, dont la forme n'est pas triviale. Tant la politique optimale du gouvernement que le niveau de redistribution favorisé par l'individu pourraient être modifiés à la lumière de ces conséquences des inégalités. L'objectif principal de cette thèse est de discuter ces changements du problème de la redistribution, d'abord du point de vue de la politique optimale dans le premier chapitre, puis sous l'angle des changements comportementaux dans les préférences des individus en matière de redistribution dans les deuxième et troisième chapitres.

Chapitre 1 : L'Externalité de l'Inégalité : Conséquences pour la Conception des Impôts

Au cours des soixante dernières années, la modélisation économique a régulièrement utilisé des fonctions d'utilité individualistes et des fonctions de bien-être social pour évaluer les options politiques. L'influence de ces modèles sur le monde réel a été considérable ; c'est pourquoi la manière dont ils traitent les inégalités économiques est également très importante. Il existe plusieurs raisons bien formulées de prévenir les différences économiques dans le cadre standard, sur lesquelles nous reviendrons prochainement, mais un facteur crucial est resté négligé : les conséquences de l'inégalité économique sur la société et, partant, sur le bien-être des individus. Supposons, par exemple, qu'une plus grande inégalité des revenus ou des richesses modifie de manière causale le taux de criminalité, l'ampleur des troubles sociaux ou la polarisation politique d'une société. Dans ce cas, même les individus purement égoïstes sont affectés par les différences économiques entre les personnes, que leurs revenus individuels changent ou non. Étant donné que pratiquement toutes les activités du marché affectent l'ampleur de l'inégalité économique, il

s'ensuit que l'inégalité économique elle-même pourrait être une externalité. Ce chapitre explore les conséquences de cette idée.

L'analyse que nous présentons peut être divisée en deux composantes principales. Le premier volet porte sur le thème principal du chapitre, à savoir le concept d'inégalité économique en tant qu'externalité. Ce concept est exploré de manière générale. Nous expliquons pourquoi un terme d'inégalité économique dans la fonction d'utilité est la manière la plus appropriée de modéliser les effets de l'inégalité sur la société, à la suite de Thurow [1971] et de Alesina and Giuliano [2011], et pourquoi un tel terme ne peut pas être approximé mathématiquement par des fonctions de bien-être social (SWF) appropriées ou des fonctions d'utilité concaves. En tant que tels, la plupart des modèles qui excluent les externalités supposent également que l'inégalité économique ne modifie pas la société de manière significative. Comme il s'agit d'une hypothèse potentiellement importante, nous examinons comment le fait de l'affaiblir et de permettre diverses externalités d'inégalité affecte à la fois l'intuition économique générale et les barèmes d'imposition optimaux. Nous autorisons des externalités d'inégalité positives ou négatives et établissons des micro-fondations potentielles, qui sont souvent simples. L'externalité peut exister en présence d'individus rationnels et totalement égoïstes.[17]

La deuxième partie de l'article se concentre sur le modèle non linéaire optimal d'imposition des revenus de Mirrlees [1971], dans lequel nous calculons les taux marginaux d'imposition optimaux de manière analytique et numérique en présence de différents types d'externalités d'inégalité. Bien que nous nous concentrions sur une externalité d'inégalité de revenu après impôt, nous introduisons également d'autres types d'externalités d'inégalité dans le modèle (revenu avant impôt, utilité) et nous faisons varier la métrique d'inégalité elle-même. Pour déterminer l'ampleur plausible d'une externalité d'inégalité de revenu dans le monde réel, nous utilisons trois méthodes distinctes, dont la principale utilise les données d'enquête de Carlsson et al. [2005] et qui impliquent toutes des fourchettes d'ampleur similaires. Enfin, nous effectuons un exercice d'optimisation inverse pour examiner comment les pondérations implicites du bien-être social dans le système fiscal américain changent si la conception du barème fiscal est influencée par une externalité d'inégalité de revenu.

L'idée principale de notre article est que la grande majorité des modèles économiques basés sur le bien-être supposent implicitement que l'inégalité économique n'a pas d'effets significatifs sur d'autres variables socio-économiques, et que l'assouplissement de cette hypothèse modifie radicalement les conclusions du modèle. Nous montrons explicitement ces changements dans l'imposition optimale des revenus (OIT), où les conclusions théoriques comme celles basées sur des simulations sont affectées. Nous discutons comment d'autres types de modèles pourraient être affectés de la même manière. Dans le contexte de l'OIT, nous trouvons deux résultats principaux. Premièrement, la présence d'une externalité d'inégalité a un impact particulièrement prononcé sur les taux d'imposition marginaux optimaux les plus élevés. Il s'agit d'un résultat théorique qui se confirme dans nos simulations numériques ; dans notre spécification principale, le taux marginal d'imposition supérieur optimal passe de 63% à 81% lorsque l'on introduit une externalité d'inégalité de revenu médian après impôt à partir des données d'enquête de Carlsson

[17]Le fait que ces individus intéressés soient affectés par l'externalité est la principale différence entre notre concept et les préférences de type "other-regarding". Ces préférences posent un problème philosophique pour l'élaboration des politiques car elles sont fondées sur les émotions des individus [Harsanyi, 1977, Goodin, 1986].

et al. [2005]. Deuxièmement, notre analyse révèle que l'aversion totale pour l'inégalité dans le système fiscal américain actuel est insuffisante pour tenir compte à la fois des pondérations du bien-être social qui diminuent avec le revenu et d'une préoccupation importante pour les effets externes de l'inégalité. Alors que le système fiscal actuel pourrait être rationalisé comme donnant la priorité aux transferts de revenus vers les personnes à faible revenu [Hendren, 2020], il ne peut pas non plus contenir une préoccupation réaliste pour les effets externes de l'inégalité étant donné la capacité globale du barème fiscal à atténuer l'inégalité.

Avant d'examiner en détails nos résultats, nous examinerons brièvement ce que nous savons sur la manière dont l'inégalité économique affecte les différentes facettes de la société et de la vie des individus. Il est difficile d'établir une causalité sur le sujet pour plusieurs raisons, la première étant l'absence de variation exogène de l'inégalité macroéconomique.[18] Cependant, les analyses empiriques sur le sujet ne manquent pas, et il y a dans l'ensemble de fortes indications que l'inégalité économique agit comme une externalité de diverses manières. Premièrement, de nombreuses données expérimentales et microéconomiques ont indiqué ces dernières années que l'inégalité économique entre les travailleurs ou les sujets d'expérience avait un impact sur la satisfaction de la vie [Card et al., 2012], la productivité [Breza et al., 2018], la confiance [Fehr et al., 2020b] et la coopération [Xu and Marandola, 2022]. Deuxièmement, comme l'ont popularisé Wilkinson and Pickett [2011], il existe de solides corrélations au sein des différents pays entre l'inégalité des revenus et divers résultats négatifs pour la société.[19] Nous présentons deux corrélations de ce type pour la confiance générale et les homicides dans la figure REF. Troisièmement, tant les profanes que les experts expriment souvent la conviction que l'inégalité modifie la société ; aux États-Unis, la grande majorité des citoyens estiment que l'inégalité économique a une incidence négative sur un large éventail de résultats sociétaux [Lobeck and Støstad, 2023]. Des préoccupations similaires ont été exprimées par d'éminents politiciens, philosophes et économistes.[20] Des expériences en laboratoire indiquent également qu'une majorité d'individus renonceraient à une partie de leurs revenus pour vivre dans des sociétés plus égalitaires sur le plan macroéconomique [Carlsson et al., 2005, Bergolo et al., 2022]. Quatrièmement, il est facile de créer des fondements microéconomiques réalistes de diverses externalités de l'inégalité, comme nous le montrons dans la section [sec:Further-Theoretical-Discussion]. D'autres articles ont accordé plus d'attention à des canaux potentiels spécifiques ; Benabou [1996] Auclert and Rognlie [2018], et Mian et al. [2020] n'en sont que quelques exemples.[21]

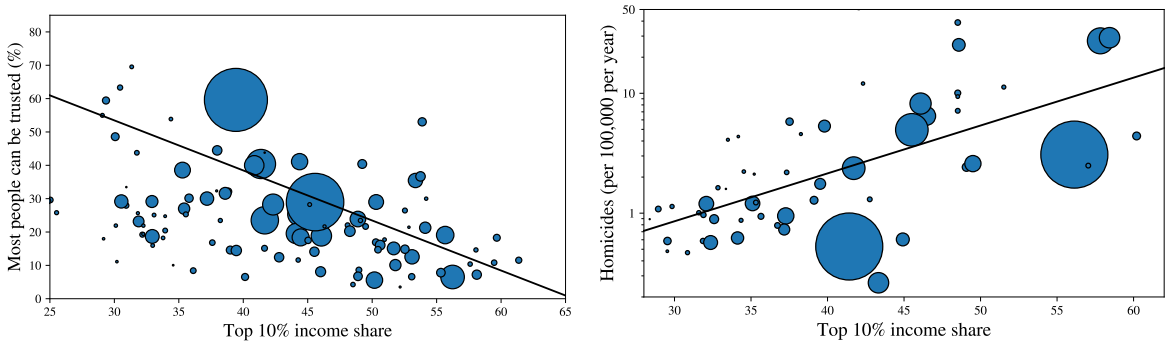
[18]Parmi les autres préoccupations figurent les erreurs de mesure et les données manquantes sur l'inégalité économique, une variabilité généralement faible de l'inégalité dans le temps, la causalité inverse (lorsque les résultats affectent également l'inégalité), les effets non linéaires de l'inégalité sur les résultats, la nature imbriquée des effets de l'inégalité et des effets de la pauvreté ou du revenu individuel, la question de savoir quel type d'inégalité est importante, etc.

[19]La littérature dans ce domaine est trop abondante pour être résumée ici. Des exemples d'analyses pertinentes peuvent être trouvés dans [Rufrancos et al., 2013] pour la criminalité et Bergh et al. [2016] pour la santé individuelle.

[20]Comme Plato [2016] : "Dans un État qui souhaite être sauvé du plus grand de tous les fléaux [...], il ne devrait y avoir parmi les citoyens ni pauvreté extrême, ni, encore une fois, excès de richesse, car les deux sont productifs de ces deux maux", [Greenspan, 2014] : "Vous pouvez voir l'impact détériorant de [l'inégalité] sur notre système politique actuel", ou [Obama, 2011] : "Ce type d'inégalité - un niveau que nous n'avons pas vu depuis la Grande Dépression - nous blesse tous."

[21]Bien que la plupart des éléments présentés dans ce paragraphe indiquent que l'inégalité économique est une externalité négative, nous ne partons pas de cette hypothèse en général. Jones [2022] explique par exemple com-

Figure 8: Les corrélations de l'inégalité



Note: A gauche : Corrélation entre les pays de la confiance généralisée (World Values Survey) et la part des 10 % de revenus les plus élevés (World Inequality Database). À droite : Corrélation entre les homicides (World Bank) et les 10 % de revenus nationaux avant impôts les plus élevés (World Inequality Database). Les deux corrélations corrélations sont relativement peu affectées par les contrôles standards. La surface des points de données est proportionnelle à la population. La surface des points de données est proportionnelle à la population. Noter l'échelle logarithmique du taux d'homicide.

Supposons donc que ces effets existent et qu'ils sont pertinents pour le bien-être. Comment envisager leurs conséquences globales dans un cadre welfariste ? L'approche la plus intuitive consiste simplement à modéliser chaque externalité individuellement. Il est toutefois difficile d'imaginer qu'un tel modèle puisse être traitable, et cette stratégie nécessiterait également une évaluation empirique problématique de l'importance réelle de chaque canal d'externalité. Pour ces raisons, cette voie semble donc globalement irréalisable. Entre-temps, un planificateur social standard averse à l'inégalité ou rawlsien (maximin) est généralement insuffisant pour modéliser les externalités d'inégalité basées sur le revenu ou la richesse, car l'externalité introduit un fossé entre ce qui est optimal sur le plan individuel et sur le plan sociétal.[22]

Une solution plus appropriée consiste à inclure une mesure de l'inégalité dans la fonction d'utilité de l'individu. Une telle externalité peut exister même lorsque les individus sont parfaitement rationnels et égoïstes, contrairement aux modèles avec des préférences traditionnelles pour d'autres considérations (ORP). À titre d'exemple, imaginons un individu parfaitement égoïste dans une société où l'inégalité des revenus augmente la criminalité dans un cadre de coût d'opportunité de type Becker [1968]. Supposons que l'inégalité des revenus et donc la criminalité augmentent et que la personne se fasse voler son vélo en conséquence. L'individu subit un choc négatif et préférerait sans aucun doute, en l'absence de tout autre changement, l'état antérieur (plus égalitaire) du monde. Par conséquent, si l'inégalité entraîne une augmentation de la criminalité, l'inégalité devrait entrer dans sa fonction d'utilité. Un argument similaire pourrait être avancé pour toute autre variable affectée par l'inégalité.

ment les revenus les plus élevés peuvent stimuler l'innovation. On pourrait imaginer que de telles préoccupations soient plus répandues dans des sociétés plus égalitaires sur le plan économique que celles que nous connaissons aujourd'hui.

[22] Les individus non altruistes choisiront leur propre effort de travail sans tenir compte de l'impact de cet effort sur le niveau global d'inégalité des revenus, par exemple ; si l'inégalité des revenus affecte des facteurs sociétaux, ce choix comporte une dimension d'externalité qui n'est pas bien modélisée par la simple actualisation de l'utilité individuelle.

Nous ne sommes pas les premiers à remarquer que les effets de l'inégalité économique sur la société peuvent impliquer un terme d'inégalité dans la fonction d'utilité. L'idée a été développée pour la première fois par Thurow [1971], qui montre que le premier théorème du bien-être échoue si la distribution des revenus est un bien public pur. Depuis lors, l'idée a périodiquement refait surface. Kaplow [2010], par exemple, mentionne que la distribution économique pourrait affecter des variables telles que la criminalité, ce qui pourrait impliquer des effets fiscaux optimaux. Alesina and Giuliano [2011] examinent brièvement comment les effets de l'inégalité sur la société pourraient affecter la consommation et donc l'utilité, et Rueda and Stegmueller [2016] examinent comment l'inégalité peut agir comme une externalité dans le cas de la criminalité. Nous ajoutons à cette littérature en détaillant l'effet d'une externalité d'inégalité dans un modèle économique spécifique et bien connu - le modèle de Mirrlees - et en approfondissant l'analyse présentée dans ces travaux. Pour mieux comprendre l'idée générale, nous (i) clarifions la structure mathématique de l'externalité, classons ses composantes clés et développons une statistique suffisante pour l'ampleur de l'externalité compte tenu d'une mesure de l'inégalité, (ii) faisons une première approche de l'estimation de l'ampleur d'une externalité d'inégalité de revenu après impôt sur la base du coefficient de Gini, (iii) formulons une série de statistiques suffisantes pour l'ampleur de l'externalité d'inégalité de revenu après impôt sur la base d'une mesure de l'inégalité de revenu, (iv) formuler un ensemble de conséquences théoriques des externalités d'inégalité basées sur le revenu et la richesse qui sont pertinentes pour la littérature économique au sens large, et (v) créer des micro-fondations pour diverses façons dont l'inégalité économique peut modifier des facteurs sociétaux pertinents.

Notre étude de cas utilise le modèle de Mirrlees [1971], dans lequel nous introduisons différents types de termes d'inégalité dans la fonction d'utilité de l'individu. En tant que modèle largement utilisé pour décrire l'OIT, le modèle de Mirrlees représente un pilier important de l'économie publique [Diamond and Mirrlees, 1971, Diamond, 1998, Saez, 2001]. Le modèle original de Mirrlees suppose qu'il n'y a pas d'externalités, une hypothèse qui a été examinée dans un grand nombre d'articles. Nous reviendrons sur la manière dont notre travail contraste avec la littérature existante, mais la manière la plus brève de décrire notre contribution technique est que nous sommes le premier article à explorer l'effet d'un terme d'inégalité des revenus dans la fonction d'utilité individuelle dans le modèle continu de Mirrlees. Le planificateur social est confronté à un arbitrage entre la maximisation des recettes fiscales et la fixation du niveau préféré d'inégalité des revenus après impôt, ce que développent les modèles de Oswald [1983], Kanbur and Tuomala [2013], et Aronsson and Johansson-Stenman [2020]. Nous explorons plusieurs types d'externalités d'inégalité dans ce cadre, en nous concentrant principalement sur une externalité d'inégalité de revenu après impôt, et nous résolvons le problème à la fois analytiquement et numériquement. Plusieurs hypothèses sont faites pour simplifier : il n'y a pas d'effets de revenu, il n'y a pas de marge extensive de l'offre de travail, il y a séparabilité du revenu, de l'effort de travail et de l'externalité d'inégalité, et le seul instrument disponible pour le planificateur social est celui de l'impôt sur le revenu.

Comme nous l'avons mentionné, notre exercice OIT produit deux résultats principaux. Premièrement, les taux marginaux d'imposition les plus élevés sont particulièrement sensibles à l'externalité de l'inégalité. L'intuition découle de la manière dont une faible augmentation de

l'impôt marginal à une tranche d'imposition donnée affecte respectivement les recettes fiscales et l'inégalité des revenus après impôt. En général, les effets sur l'inégalité des revenus après impôt - qui sont pertinents pour le bien-être dans notre cas - sont plus fortement influencés par la répartition du barème d'imposition que les effets standards sur les revenus. Ceci peut être vu à travers le cadre détaillé dans Saez [2001], qui discute les conséquences d'une petite augmentation d'impôt comme (i) l'effet mécanique sur les recettes fiscales, et (ii) les réponses comportementales des agents. Dans le cas standard de la maximisation des recettes, les deux effets s'opposent toujours. Une augmentation d'impôt entraîne mécaniquement une hausse des recettes fiscales, ce qui est positif pour le bien-être. Dans le même temps, les changements de comportement des agents, qui se détournent de l'offre de travail, entraînent des distorsions et diminuent les recettes fiscales, ce qui est négatif sur le plan du bien-être. Il en résulte l'arbitrage classique entre l'équité et l'efficacité. En revanche, les deux effets peuvent également s'harmoniser dans leur impact sur l'inégalité des revenus après impôt. L'effet mécanique est similaire au cas des recettes, puisque le fait de collecter des recettes fiscales auprès des personnes se situant au-dessus d'une tranche d'imposition donnée pour les redistribuer diminue toujours l'inégalité des revenus après impôt (sauf pour la tranche d'imposition la plus basse). Les réactions comportementales des agents augmentent ou diminuent toutefois l'inégalité des revenus après impôt en fonction de l'endroit où se situe la hausse d'impôt. En bas de l'échelle, un changement de comportement consistant à renoncer à l'effort de travail accroît l'inégalité des revenus. En haut de l'échelle, un changement de comportement consistant à renoncer à l'effort de travail diminue l'inégalité des revenus. Cela crée une asymétrie de distribution, où l'effet des réponses comportementales sur l'inégalité des revenus (et donc sur le taux d'imposition optimal) est opposé au sommet et à la base de la distribution. Cela signifie que, en ce qui concerne l'inégalité des revenus après impôt, les effets fiscaux optimaux de l'effet mécanique et des réponses comportementales des agents s'opposent toujours au bas de la distribution et s'harmonisent au sommet. Les taux d'imposition marginaux supérieurs sont donc particulièrement sensibles à l'externalité de l'inégalité.

Dans nos simulations numériques, l'application de l'estimation médiane de l'externalité entraîne une augmentation du taux d'imposition marginal supérieur optimal de 63% à 81%. Compte tenu des valeurs standard des paramètres et de l'ampleur raisonnable de l'externalité, nous constatons un très large éventail de taux marginaux d'imposition supérieurs optimaux possibles, allant de taux marginaux d'imposition supérieurs négatifs (¡0%) si l'inégalité est une externalité positive à des taux marginaux d'imposition supérieurs extrêmement élevés (¿90%) si l'inégalité est une externalité négative. Cette gamme de taux d'imposition maximaux optimaux est plus large que celle basée sur les valeurs standard des paramètres dans le cas de l'absence d'externalité, où les taux d'imposition marginaux optimaux se situent généralement entre 50% et 80%. Cette fourchette étroite dans le cas classique s'explique en partie par le fait que chaque fonds souverain standard converge vers le même taux d'imposition maximal optimal dans le modèle de Mirrlees.[23] Cela a sans doute réduit l'importance accordée à la "dimension de l'égalité" dans l'analyse des taux d'imposition maximaux optimaux - dont nous

[23] Cela s'explique par le fait que l'avantage d'un revenu supplémentaire au sommet de l'échelle se rapproche de zéro dans la plupart des modèles standard, soit en raison des pondérations de bien-être social qui diminuent le revenu, soit en raison de la diminution de l'utilité marginale du revenu.

montrons qu'elle peut être très pertinente tant que l'inégalité elle-même affecte l'individu.[24] Les préoccupations en matière d'inégalité individuelle qui découlent d'une externalité d'inégalité diffèrent donc des préoccupations en matière d'inégalité sociale modélisées par un planificateur social ayant une aversion pour l'inégalité. Nous trouvons naturellement des taux d'imposition maximaux optimaux supérieurs au taux de Laffer qui maximise les recettes, car les effets directs de l'égalité impliquent que le planificateur social pourrait échanger certaines recettes contre des niveaux d'égalité modifiés. Nos résultats, si l'inégalité est une externalité négative, fournissent également une base théorique pour des arguments politiques qui n'étaient pas étayés auparavant - tels que les taux d'imposition marginaux supérieurs élevés d'après-guerre aux États-Unis et au Royaume-Uni.

Notre deuxième résultat principal est lié à ce dernier point et provient de l'exercice d'optimum inversé popularisé par Bourguignon and Spadaro [2012]. Cette méthode calcule les poids de bien-être social (SWW) implicites des systèmes fiscaux du monde réel sous l'hypothèse que le barème fiscal a été fixé de manière optimale. Comme le montrent Lockwood and Weinzierl [2016] et Hendren [2020], les SWW du barème fiscal américain sont généralement décroissants en fonction du revenu. Nous introduisons une externalité d'inégalité dans ce cadre, ce qui nous permet de comprendre les ramifications sur les SWW implicites du système fiscal si le planificateur social considérait l'inégalité comme une externalité lors de la conception du barème fiscal. Grâce à cet exercice, nous constatons que le système fiscal américain de 2019 n'est pas suffisamment réticent à l'inégalité pour tenir compte à la fois d'un motif de transfert socialement progressif et d'une préoccupation réaliste concernant les effets sociétaux de l'inégalité économique. L'intuition est que tout barème fiscal contient une certaine quantité d'aversion "totale" à l'inégalité qui peut expliquer soit des SWW décroissants en fonction du revenu, comme dans Lockwood and Weinzierl [2016] et Hendren [2020], soit une préoccupation non négligeable pour les effets d'externalité de l'inégalité. L'aversion totale actuelle pour l'inégalité dans le barème fiscal américain est toutefois trop faible pour expliquer ces deux phénomènes. Si le planificateur social américain considérait l'inégalité comme une externalité par rapport à notre valeur médiane, les SWW implicites augmenteraient fortement avec le revenu - indiquant qu'un dollar au bas de la distribution vaut cinq dollars au sommet. Nous en concluons que le barème fiscal américain actuel n'est pas progressif dans les transferts ou qu'il se préoccupe des effets externes de l'inégalité d'une manière nettement inférieure à nos estimations empiriques.

Nous présentons également une poignée de résultats plus modestes. Le barème fiscal optimal est sans ambiguïté plus progressif (régressif) en cas d'externalité négative (positive) de l'inégalité des revenus après impôt, une plus grande progressivité étant définie comme un niveau plus faible d'inégalité des revenus après impôt. La forme classique en U que l'on trouve dans la littérature sur l'imposition optimale des revenus est fragile à l'inclusion d'une externalité négative (mais non positive) d'inégalité de revenu après impôt. Une externalité d'inégalité de revenu avant impôt de taille modérée rend les taux marginaux d'imposition utilitaristes optimaux plus proches des systèmes fiscaux du monde réel, où les taux marginaux augmentent largement avec le revenu. Plus généralement, les résultats numériques et théoriques du modèle OIT dépendent fortement du type et de l'ampleur de l'externalité de l'inégalité.

[24] Cela ne veut pas dire que l'externalité de l'inégalité est un problème d'équité dans le cadre standard de l'équité et de l'efficacité, où il s'agit clairement d'un problème d'efficacité.

Nous allons maintenant décrire brièvement en quoi notre travail diffère de la littérature existante sur l'OIT. Comme notre approche suppose la séparabilité entre l'externalité et le reste de la fonction d'utilité pour des raisons de simplicité,[25] notre cadre est une extension des modèles présentés dans Oswald [1983] et en particulier Kanbur and Tuomala [2013], qui examinent tous deux une externalité basée sur le revenu moyen. Nous notons trois nouveautés techniques principales par rapport à la littérature existante. Premièrement, nous introduisons une manière nouvelle et simple de prendre en compte les termes d'inégalité dans les fonctions d'utilité individuelle dans les cadres de taxation optimale. Ceci est possible grâce à la famille de mesures d'inégalité que nous utilisons,[26] qui simplifie une externalité analytiquement difficile à résoudre en une combinaison linéaire d'externalités de consommation avec des effets marginaux variables qui dépendent du rang de revenu de l'individu.[27] En tant que tel, nous pouvons utiliser une grande partie du cadre d'externalité existant, y compris Oswald [1983] et Kanbur and Tuomala [2013], pour évaluer ce qui serait autrement un problème analytique difficile. La deuxième contribution à la littérature consiste donc à explorer les ramifications d'une extension de Kanbur et Tuomala (2013) qui permet à l'externalité marginale de dépendre de la position de l'individu dans la distribution (et donc aussi de son revenu). Bien qu'Oswald [1983] et Kanbur and Tuomala [2013] mentionnent cette possibilité, aucun des deux documents n'explore explicitement la question. Nous mobilisons une approche par perturbation pour clarifier l'intuition derrière nos résultats, contrairement à ces deux articles qui utilisent l'approche par mécanismes d'incitations (dont nous résolvons également la version modifiée). Notre analyse conduit à de nouvelles perspectives concernant l'effet des externalités distributives sur l'imposition optimale des revenus, et en particulier sur les taux d'imposition maximaux optimaux. Troisièmement, nous résolvons le problème d'optimum inversé de [Bourguignon and Spadaro, 2012] en présence d'une externalité globale et illustrons les conséquences pour les SWW implicites du système fiscal américain de 2019. Les externalités globales sont rarement abordées dans cette littérature - nous ne connaissons que Tsyvinski and Werquin [2017], qui aborde le principe de compensation dans un cadre basé sur l'équilibre général et est donc à la fois conceptuellement et mathématiquement différent de notre travail. Étant donné l'importance accordée aux effets de l'inégalité sur la société dans la rhétorique politique, nous pensons qu'il s'agit d'un exercice particulièrement intéressant dans notre cadre.

D'une manière générale, notre travail vient s'ajouter à la littérature déjà abondante sur les externalités dans le cadre de la fiscalité optimale. Cette littérature a été particulièrement développée pour les externalités environnementales [Sandmo, 1975, Bovenberg and van der Ploeg, 1994, Cremer et al., 1998, par example] et les préoccupations relatives au revenu relatif/ORP [Boskin and Sheshinski, 1978, Persson, 1995, Aronsson and Johansson-Stenman, 2008, 2015,

[25] Il s'agit d'une hypothèse importante car elle limite la manière dont l'ampleur de l'externalité est liée aux utilités marginales du revenu et du travail. L'hypothèse de séparabilité des externalités est affaiblie, entre autres, par Pirttilä and Tuomala [1997] et Jacobs and De Mooij [2015]. Nous supposons également la séparabilité entre l'utilité du revenu et celle du travail, toujours par souci de simplicité. Pour plus d'informations sur l'hypothèse de séparabilité, voir Gauthier and Laroque [2009].

[26] Les mesures d'inégalité typiques utilisent souvent des valeurs absolues et des intégrales multiples qui dépendent des variables endogènes du modèle.

[27] Il s'agit de la même famille que celle utilisée dans Simula and Trannoy [2022], développée en même temps que le présent document. La famille elle-même est générale et permet différents types de mesures de l'inégalité des revenus.

2018c, 2020]. Notre analyse est particulièrement liée à celle d'Aronsson and Johansson-Stenman [2020], qui examine différents types d'ORP, y compris l'aversion classique pour l'inégalité de type Fehr and Schmidt [1999], dans un modèle OIT à trois agents. Nous approfondissons cette analyse en utilisant un ensemble plus large de spécifications liées à l'inégalité dans un modèle continu complet de type Mirrlees. Kanbur et al. [1994] montrent également qu'il est possible de mettre directement l'accent sur les problèmes de répartition dans le modèle OIT, en termes de pauvreté dans la fonction de bien-être social, ce qui contraste avec notre métrique de répartition continue dans la fonction d'utilité de l'individu.

En résumé, nous examinons dans ce chapitre la manière de modéliser les conséquences de l'inégalité dans les cadres welfaristes. Notre principale suggestion est de traiter l'inégalité économique elle-même comme une externalité, ce qui a des implications importantes pour la théorie classique. Nous montrons cela à travers le modèle classique d'imposition optimale non linéaire des revenus, où nous nous concentrons sur une externalité d'inégalité des revenus après impôt. Les taux d'imposition supérieurs sont particulièrement affectés par l'externalité ; dans notre spécification principale, le taux d'imposition marginal supérieur optimal passe de 63 % à 81 %. Notre modèle fournit également une base théorique pour les choix fiscaux des gouvernements du monde réel qui sont irrationnels dans le cadre des méthodes standard de taxation optimale. Enfin, nous constatons que l'aversion totale pour l'inégalité induite par le système fiscal américain actuel est insuffisante pour tenir compte à la fois des pondérations du bien-être social qui diminuent avec le revenu et d'une préoccupation significative pour les effets d'externalité de l'inégalité.

Chapitre Deux : Les Conséquences de l'Inégalité : Croyances et Préférences Redistributives

Pourquoi devrions-nous nous préoccuper des inégalités économiques ? Cette question a fait l'objet d'innombrables controverses, tant dans le débat public que dans la littérature académique. La discussion porte souvent sur l'égalité par opposition à l'efficacité [Okun, 1975]. D'une part, la redistribution est nécessaire pour corriger les résultats jugés injustes. D'autre part, la redistribution elle-même impose des coûts d'efficacité. Ce dilemme de longue date a façonné l'intuition économique et le débat sur la redistribution pendant des décennies.

Cependant, tous les arguments en faveur de l'égalité ne sont pas liés à l'équité, et certains sont même liés à l'efficacité. Dans cet article, nous menons des enquêtes à grande échelle pour quantifier les croyances des individus sur les *conséquences de l'inégalité*, qui offrent des arguments d'efficacité pour soutenir ou s'opposer aux politiques de redistribution. De telles conséquences se produisent lorsque l'inégalité économique affecte quelque chose qui nous tient à cœur, par exemple le niveau de troubles sociaux, le taux de croissance économique ou la confiance générale entre les gens. Nous appelons ces conséquences *externalités de l'inégalité*, en partant de l'idée qu'elles sont un effet secondaire de l'inégalité économique à laquelle nous contribuons tous par des actions sur le marché [Støstad and Cowell, 2021].[28] Ce cadre montre clairement que l'inégalité elle-même peut entraîner des coûts d'efficacité. Ce qui est important pour le présent document, c'est que les croyances des gens sur ces conséquences peuvent varier, ce qui peut à son tour

[28] Ces actions sur le marché peuvent aller de l'effort de travail à des décisions d'investissement.

affecter les demandes globales de redistribution. Par exemple, s'il existe un consensus sur le fait que les grandes différences économiques conduisent à des révolutions violentes, un consensus en faveur d'une politique de redistribution peut être obtenu simplement en soulignant ces risques. En général, ce que les gens pensent de ces externalités de l'inégalité - ce qu'elles sont et leur impact - pourrait influencer les préférences en matière de redistribution et même les paysages politiques. La connaissance de ces croyances pourrait, à son tour, clarifier notre compréhension des raisons pour lesquelles les gens se préoccupent de l'inégalité.

À la suite de ces observations, le présent document pose deux questions principales. Premièrement, les citoyens s'attendent-ils à ce que l'inégalité économique change la société - et si oui, comment ? Deuxièmement, dans quelle mesure ces croyances ont-elles un impact causal sur les préférences des citoyens en matière de redistribution ? Pour répondre à ces questions, nous menons deux nouvelles enquêtes représentatives de la population américaine, en échantillonnant un total de 4 371 et 2 360 citoyens américains distincts avec les fournisseurs d'enquêtes professionnels Lucid et Dynata. Ces deux enquêtes nous permettent de créer les premiers ensembles de données complets sur les croyances des citoyens américains en matière d'externalité de l'inégalité. Nous explorons le lien entre ces croyances et les préférences en matière de redistribution à l'aide de plusieurs méthodes, dont la plus importante est une expérience d'information basée sur une vidéo. Cette expérience d'information est conçue pour isoler l'effet causal des croyances en l'externalité de l'inégalité sur les préférences en matière de redistribution ; elle nous permet également de comparer la façon dont les croyances en l'externalité de l'inégalité affectent les préférences en matière de redistribution par rapport aux opinions générales en matière d'équité. Enfin, nous discutons des différents degrés de polarisation dans les arguments redistributifs fondés sur l'équité et ceux fondés sur l'externalité des inégalités.

En dépit de leur impact politique et économique potentiel, les croyances dans les externalités de l'inégalité ont, à notre connaissance, rarement fait l'objet d'une étude formelle. Alors que les idées générales des individus sur l'inégalité, la redistribution et l'équité économique sont largement étudiées dans les enquêtes internationales - d'innombrables questions explorent ces sujets dans le World Values Survey, le Gallup World Poll, et ainsi de suite - les questions sur les conséquences de l'inégalité sont extrêmement rares. Les quelques questions qui ont été posées à des échantillons représentatifs portent généralement sur les externalités positives de l'inégalité, et plus particulièrement sur l'idée que l'inégalité économique augmente le niveau de croissance économique ou d'innovation. La seule question relative aux États-Unis dont nous ayons connaissance provient du General Social Survey (GSS), qui a demandé aux personnes interrogées lors de cinq vagues entre 1987 et 2021 si elles étaient d'accord avec l'idée que "les grandes différences de revenus sont nécessaires à la prospérité de l'Amérique". La question montre une tendance à la baisse constante : alors que 34 % des personnes interrogées estimaient que l'inégalité était nécessaire à la prospérité en 1987, elles ne seront plus que 12 % à le penser en 2021.

Alors que la question ci-dessus est un exemple d'externalité *positive*, nos résultats montrent que la plupart des citoyens américains croient aussi en l'existence d'externalité *négative*. Presque tous les individus ($\sim$ 90%) pensent que l'inégalité affecte la société d'une manière ou d'une autre, et une majorité cohérente ($\sim$ 60%) pense que l'inégalité économique a des conséquences

sociétales globalement néfastes.[29] Nous nous penchons sur les raisons potentielles et trouvons des convictions fortes dans des domaines spécifiques ; 76 % des personnes interrogées pensent qu'une plus grande inégalité économique augmente le nombre de crimes, par exemple, et 68 % pensent qu'elle détériore le niveau général de confiance dans la société. Alors que seulement 23 % des personnes interrogées pensent qu'une plus grande inégalité économique augmente la croissance économique (ce qui rappelle la question de l'GSS), 51 % pensent l'inverse, à savoir qu'une plus grande inégalité économique diminue la croissance économique. Les personnes interrogées ont des opinions tout aussi négatives quant à l'effet de l'inégalité économique sur la prévalence de la corruption ou des troubles sociaux, sur la qualité des institutions démocratiques, etc. Les résultats sont robustes aux différentes méthodologies et formulations des questions, et sont presque identiques dans les différents échantillons représentatifs.[30].

Qui croit donc à ces conséquences néfastes de l'inégalité ? Si les démocrates sont 15 à 20 points de pourcentage plus susceptibles que les républicains de croire aux externalités négatives,[31]. les principales conclusions sont similaires d'un camp politique à l'autre. Quelle que soit leur affiliation politique, les personnes interrogées sont plus susceptibles de penser que l'inégalité économique nuit à la société plutôt qu'elle ne l'aide, et ce pour tous les résultats que nous obtenons. Les démocrates, les indépendants et les républicains sont tous plus enclins à penser qu'une plus grande inégalité économique *diminue* plutôt qu'*augmente* la croissance économique et l'innovation, par exemple. Le fait que ces convictions soient relativement similaires d'un parti à l'autre est particulièrement évident lorsque nous les comparons aux opinions générales en matière d'équité économique, que nous utilisons comme référence tout au long de l'étude. Ces préoccupations fondées sur l'équité - concernant par exemple la question de savoir si la répartition actuelle des revenus est *équitable* - sont nettement plus divisées entre les partis (∼30-35 p.p.). En effet, deux répondants types de chaque parti politique sont plus susceptibles d'être d'accord sur n'importe quelle question relative aux externalités que sur n'importe quelle question relative à l'équité dans l'un ou l'autre de nos deux échantillons. Cette non-polarisation des convictions en matière d'externalité est frappante et pourrait refléter une différence intrinsèque entre les convictions en matière d'externalité et les préoccupations en matière d'équité. Nous constatons également une non-polarisation similaire des croyances en matière d'externalité en fonction du statut économique ; alors que les moins bien lotis sont beaucoup plus susceptibles de dire que la répartition économique est injuste, les riches et les pauvres sont à peu près aussi susceptibles de croire aux conséquences négatives de l'inégalité.

Après avoir établi l'existence de croyances généralisées en l'externalité de l'inégalité au sein de la population américaine, nous nous penchons sur les implications de ces croyances. Plus précisément, nous testons si ces croyances constituent un déterminant causal des préférences en matière de redistribution, ce qui vient s'ajouter à la vaste littérature explorant les déterminants

[29] $\sim 10-15\%$ pensent que l'effet net est positif, $\sim 15-20\%$ pensent que les effets positifs et négatifs "s'annulent", et $\sim 5-15\%$ ne croient pas que l'inégalité affecte la société.

[30] Les personnes interrogées répondent de la même manière aux questions très simples et très détaillées, et aux questions posées au début ou à la fin de l'enquête. Remplacer le mot "inégalité" par "égalité" ou "différences de revenu et de richesse" ne modifie pas les résultats globaux (lorsque la formulation "égalité" implique que plus d'égalité entraîne, par exemple, moins de criminalité). Dans une question placebo où la réponse raisonnable est "pas de changement", presque toutes les personnes interrogées choisissent cette option

[31] 78 % des personnes interrogées d'orientation démocrate pensent que les inégalités augmentent les troubles sociaux, contre 62 % des personnes interrogées d'orientation républicaine, par exemple

des préférences des individus en matière de redistribution. Notre principale méthode pour explorer cette question est une expérience d'information exogène. Nous utilisons quatre vidéos faciles à comprendre pour informer les personnes interrogées sur quatre ensembles différents de relations empiriques : (i) la corrélation au niveau des pays entre l'inégalité des revenus et la criminalité (traitement de l'externalité de la criminalité) ; (ii) la corrélation entre les pays de l'inégalité des revenus et la confiance (traitement de l'externalité de la confiance) ; (iii) les informations combinées de ces deux vidéos, couplées à des preuves empiriques plus larges (nous montrons aux personnes interrogées qu'il n'y a pas de corrélations significatives entre l'inégalité des revenus et l'innovation ou la croissance économique). Nous montrons aux personnes interrogées qu'il n'y a pas de corrélation significative entre l'inégalité des revenus et l'innovation ou la croissance économique. Les trois premiers traitements sont conçus pour modifier de manière exogène diverses croyances en matière d'externalité de l'inégalité, tandis que le dernier est conçu pour modifier de manière exogène les opinions en matière d'équité, qui constituent point de référence dans notre étude. Nous formalisons plusieurs nouvelles méthodologies de conception d'enquêtes afin d'éviter les effets de demande et d'amorçage, y compris ce que nous appelons une *enquête secondaire* (un écart structurel bien expliqué entre le traitement et les résultats d'intérêt) et *groupes de contrôle doubles* (en utilisant à la fois un groupe de contrôle actif et un groupe de contrôle passif et en les fusionnant sur la base de critères préétablis).

Nous constatons que le traitement de l'externalité totale et le traitement de l'équité ont tous deux un pouvoir prédictif significatif sur des préférences redistributives plus élevées après le traitement par rapport au groupe de contrôle ($p < 0,01$). Ces résultats sont robustes à toute une série de spécifications différentes. Les effets du traitement de l'externalité de la criminalité et de la confiance ne sont pas statistiquement significatifs, mais vont dans le sens attendu d'une augmentation des préférences redistributives. L'ampleur de l'effet du traitement de l'externalité complète est environ la moitié de celle du traitement de l'équité. D'autres tests montrent que chaque vidéo a affecté les préférences redistributives par le biais du mécanisme attendu, avec des retombées et un amorçage limités, ce qui signifie que l'expérience peut être interprétée de manière causale. En résumé, nous établissons que les croyances en matière d'externalité de l'inégalité sont des déterminants causaux des préférences en matière de redistribution, avec une ampleur significative.

Nous explorons également l'importance relative des croyances en l'externalité de l'inégalité et des préoccupations en matière d'équité à l'aide de deux autres méthodes. Premièrement, nous comparons le pouvoir prédictif des croyances d'externalité, des opinions d'équité, des préférences politiques et des préoccupations d'efficacité "classiques" dans l'estimation des préférences en matière de redistribution. Deuxièmement, nous demandons simplement aux personnes interrogées ce qu'elles prennent en compte lorsqu'elles réfléchissent au niveau de redistribution qu'elles préfèrent. Ces deux approches montrent que les croyances en matière d'externalité sont environ deux tiers aussi importantes que les opinions en matière d'équité dans la détermination des préférences en matière de redistribution, un ordre de grandeur qui est généralement cohérent avec les effets de traitement de la vidéo. Nos résultats montrent donc que les croyances en l'externalité de l'inégalité sont un déterminant causal des préférences en matière de redistribution d'une ampleur légèrement inférieure, mais sur la même échelle que les opinions en matière

d'équité - ce qui est remarquable compte tenu de l'attention comparative accordée à ces deux déterminants dans la littérature académique.

Les comparaisons strictes de l'ordre de grandeur négligent toutefois des distinctions importantes, car nos résultats indiquent également des différences structurelles dans la manière dont ces deux types d'arguments fonctionnent. Premièrement, les personnes qui ont vu la vidéo sur l'équité sont nettement plus susceptibles de déclarer que leur réaction à la vidéo était de la colère. Deuxièmement, l'effet de traitement de la vidéo sur l'équité est réparti entre les revenus ; les personnes à faible revenu sont nettement plus influencées par le traitement de l'équité que les personnes à revenu élevé. Cela contraste avec le traitement de l'externalité, qui est largement convaincant dans l'ensemble de la distribution. Ces deux points corroborent l'histoire de la polarisation que nous trouvons dans les données descriptives et qui est relativement intuitive. Les arguments en faveur de l'équité exigent, presque par définition, soit une *victime*, soit un *bourreau*, soit *les deux* - quelqu'un qui mérite plus et quelqu'un qui mérite moins. Les arguments relatifs aux externalités, quant à eux, se concentrent sur un *ennemi commun* involontaire. Par essence, les arguments relatifs aux externalités sont fondés sur l'efficacité, ce qui contraste avec les arguments d'équité fondés sur l'équité. En d'autres termes, les arguments fondés sur l'externalité sont largement conséquentialistes, tandis que les arguments fondés sur l'équité sont largement déontologiques. Il s'ensuit que les arguments fondés sur l'externalité pourraient offrir une possibilité de consensus entre des groupes qui sont souvent en désaccord sur les idéaux normatifs.

À notre connaissance, cet article est l'un des premiers à étudier explicitement l'idée de croyances en l'externalité de l'inégalité et le premier à établir un lien empirique direct entre les croyances déclarées en matière d'externalités et les préférences des individus en matière de redistribution. Une littérature abondante a examiné divers autres déterminants des préférences en matière de redistribution, en particulier les idéaux d'équité des individus et leurs préoccupations quant aux coûts d'efficience de la redistribution [e.g. Cappelen et al., 2007, Durante et al., 2014]. Des deux, les idéaux d'équité s'avèrent souvent être la motivation la plus forte [Almås et al., 2020], bien qu'il y ait une certaine variation entre les différents groupes de la population [Fisman et al., 2015]. Des articles ont également étudié le lien entre les préférences en matière de redistribution et les croyances concernant la position relative d'une personne [Cruces et al., 2013, Karadja et al., 2017], les informations sur le niveau d'inégalité et le fonctionnement des systèmes fiscaux [Kuziemko et al., 2015, Stantcheva, 2021], et les croyances en matière de mobilité sociale [Alesina et al., 2018b, Gärtner et al., 2019]. Les préoccupations des citoyens concernant les conséquences de l'inégalité sont rarement abordées dans cette vaste littérature, bien qu'elles aient été proposées comme un motif possible des préférences en matière de redistribution [Alesina and Giuliano, 2011]. Une exception est le travail de Rueda and Stegmueller [2016] qui présente des corrélations entre la peur du crime et les préférences pour la redistribution parmi les individus à haut revenu en Europe occidentale, et explique l'association par un argument théorique basé sur les externalités. En guise de dernier lien avec la littérature sur les préférences en matière de redistribution, nous notons que le consensus que nous trouvons dans les croyances en matière d'externalité de l'inégalité rappelle le consensus entre les partis Norton and Ariely [2011] pour un niveau réduit d'inégalité de la richesse dans le cas de l'absence de friction.

Nous avons également un lien avec une littérature théorique centrée sur l'inégalité en tant qu'externalité. Le premier travail qui considère les effets sociétaux de l'inégalité économique dans un cadre welfariste est Thurow [1971], qui soutient que le premier théorème du bien-être échoue si la distribution des revenus est un bien public pur. Alesina and Giuliano [2011] note que l'inégalité économique peut affecter la consommation individuelle et donc les préférences en matière de redistribution, tandis que Rueda and Stegmueller [2016] considère la criminalité comme une externalité de l'inégalité et montre théoriquement comment elle a un effet sur la redistribution préférée des riches. Une partie de la littérature sur l'aversion pour l'inégalité examine spécifiquement la manière dont ces externalités d'inégalité pourraient influencer les préférences individuelles [Carlsson et al., 2005] et donc l'imposition optimale des revenus dans le modèles discrets [Aronsson and Johansson-Stenman, 2018c, 2020]. [Støstad and Cowell, 2021] formalisent un cadre autour de l'inégalité en tant qu'externalité, établissent des micro-fondations pour diverses externalités d'inégalité, et montrent qu'une externalité d'inégalité peut influencer de manière substantielle des résultats bien connus de la théorie de l'imposition optimale. [32] Cet article, en résumant les croyances largement répandues dans le public concernant ces effets - presque tous les citoyens américains pensent qu'il existe une sorte d'externalité de l'inégalité - montre qu'un planificateur social qui agrège les préférences individuelles pourrait souhaiter inclure ces considérations d'externalité dans son problème d'optimisation. Il serait donc prudent d'examiner plus sérieusement la robustesse - ou la fragilité - des cadres individualistes standard face à ces effets d'externalité.

Il existe également une vaste littérature qui tente d'établir des liens entre l'inégalité économique et divers résultats sociétaux. Un examen complet de cette littérature dépasse la portée du présent document ; en bref, il existe des corrélations indiquant que l'inégalité est une externalité dans diverses dimensions [Wilkinson and Pickett, 2011, Rufrancos et al., 2013, Bergh et al., 2016], mais il est peu probable que des preuves causales à grande échelle soient disponibles en raison de l'absence de variation exogène de l'inégalité économique. [33] Dans des contextes plus restreints, des preuves de causalité peuvent exister ; il a été démontré de manière convaincante que les inégalités économiques affectent le bien-être subjectif [Card et al., 2012] et la productivité [Breza et al., 2018] sur le lieu de travail en raison des préoccupations liées au revenu relatif, et la confiance dans les expériences de laboratoire et d'enquête [Gallego, 2016, Fehr et al., 2020b].

En résumé, cet article examine les croyances sur la façon dont l'inégalité économique change la société et établit un lien de causalité entre ces croyances sur l'externalité de l'inégalité et les préférences en matière de redistribution. À l'aide de deux enquêtes représentatives menées auprès de 6 731 citoyens américains, nous montrons qu'une majorité de personnes interrogées pensent que l'inégalité entraîne des conséquences négatives pour la société, telles que l'augmentation de la criminalité, la détérioration des institutions démocratiques et la diminution de la croissance économique. Nous établissons un lien de causalité entre les croyances des individus en matière d'externalité de l'inégalité et leurs préférences en matière de redistribution en utilisant des traitements d'information vidéo fournis de manière exogène. Grâce à cette méthode et à

[32] Notez que le fait d'appeler ces effets *externalités d'inégalité* est un raccourci ; techniquement, c'est l'inégalité économique *elle-même* qui est une externalité [Støstad and Cowell, 2021].

[33] Ainsi que d'autres préoccupations intrinsèques et des données insuffisantes – voir Støstad [2019] pour une discussion.

d'autres, nous estimons que les croyances en matière d'externalité de l'inégalité ont environ deux tiers d'impact sur les préférences redistributives des individus, au même titre que les opinions générales en matière d'équité économique. Bien que les démocrates soient plus enclins à croire aux conséquences négatives de l'inégalité que les républicains, les convictions sont étonnamment similaires d'un parti politique à l'autre et moins polarisées que les opinions comparables en matière d'équité. Les arguments fondés sur l'externalité de l'inégalité provoquent toutefois moins de colère chez les personnes interrogées que les arguments fondés sur l'équité, ce qui indique des différences structurelles entre les deux types d'arguments.

Chapitre Trois : Un Univers d'Arguments

De nombreux problèmes économiques et sociétaux dépendent de l'évaluation de différents types de déclarations. Un investisseur peut choisir d'investir sur la base de récits spécifiques, par exemple, ou un électeur peut renforcer sa préférence politique à partir d'un discours persuasif. L'efficacité et le contenu émotionnel des différents types de déclarations et d'arguments sont donc essentiels à la compréhension de l'économie. L'évaluation de la nature des différents types d'énoncés est toutefois un défi, car le choix des énoncés à évaluer laisse place à la demande de l'expérimentateur et compromet la généralisation des résultats obtenus.

Dans cet article, nous présentons une nouvelle méthode pour obtenir et évaluer un échantillon relativement impartial de déclarations ou d'arguments. Nous appelons cette méthode "Univers d'arguments". La méthode s'articule autour de trois étapes. Tout d'abord, il s'agit d'obtenir des déclarations à partir des distributions de déclarations qui nous intéressent, au moyen d'une invitation soigneusement formulée qui ne diffère que par la dimension qui nous intéresse. Deuxièmement, l'utilisation d'un ensemble indépendant de répondants à l'enquête pour vérifier la qualité de l'échantillon afin de détecter les déclarations qui ne correspondent pas à l'invitation ou qui ne sont pas souhaitées pour d'autres raisons. Troisièmement, l'utilisation d'un autre groupe indépendant de répondants à l'enquête pour évaluer l'"univers d'arguments" qui en résulte. La méthode réduit les risques de biais du chercheur à des sources bien connues (la formulation des questions et la sélection de l'échantillon).

Nous appliquons cette méthode à la question de la redistribution, en évaluant l'efficacité et le contenu émotionnel de deux types d'arguments redistributifs. Ces deux types d'arguments sont fondés soit sur *les idées d'équité*, soit sur *les conséquences sociétales de l'inégalité*. Nous décrirons les conséquences de l'inégalité comme des "externalités d'inégalité", en suivant Støstad and Cowell [2021] et Lobeck and Støstad [2023].[34]

Les "arguments d'équité" sont basés sur l'équité et se concentrent sur le mérite des individus. Les "arguments d'externalité" sont basés sur l'efficacité et se concentrent sur les conséquences de l'inégalité économique, telles que l'augmentation de la criminalité, la diminution ou l'augmentation de la croissance économique, ou la détérioration de la cohésion sociale.

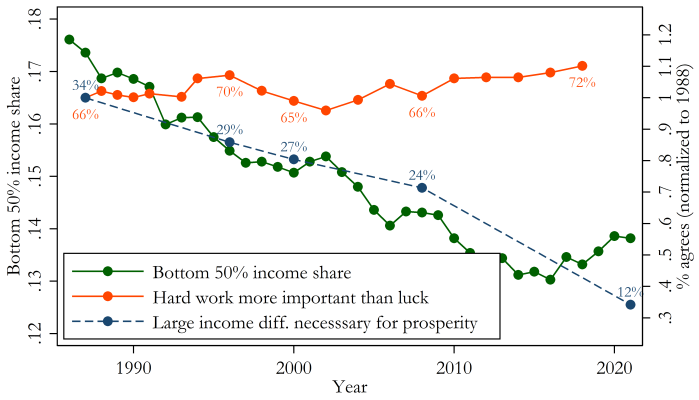
Notre intérêt pour ces arguments est en partie motivé par la Figure 9. Les inégalités ont fortement augmenté aux États-Unis depuis 1987. Les croyances dans les externalités *positives* de l'inégalité ont considérablement diminué au cours de la même période. Les croyances en l'équité du système économique, en revanche, sont restées relativement constantes (voire ont augmenté).

[34] L'inégalité économique est affectée par les décisions de marché des individus. Si l'inégalité économique affecte les résultats sociétaux pertinents, il s'ensuit que l'inégalité économique est une externalité.

En d'autres termes, les convictions relatives aux externalités pourraient être plus malléables que les convictions relatives à l'équité au niveau sociétal. Si ces types d'arguments redistributifs sont fonctionnellement distincts, en conduisant par exemple à des degrés de colère différents, alors cela pourrait expliquer à la fois les différences historiques et internationales dans les débats sur la redistribution.

Notre approche combine trois enquêtes. Dans une première enquête, les répondants ont été invités à rédiger des arguments pour ou contre la redistribution. Chaque personne interrogée recevait deux questionnaires dans un ordre aléatoire, qui étaient identiques à l'exception de la demande que l'argument soit basé soit sur des idées d'équité, soit sur les conséquences sociétales de l'inégalité. Nous avons recueilli un total de 596 arguments auprès de 298 répondants. Nous avons ensuite utilisé une deuxième enquête pour nous assurer que l'échantillon final d'arguments était pertinent et raisonnable. Il restait donc 190 arguments dans notre "univers d'arguments", dont 160 étaient en faveur de la redistribution et 30 contre la redistribution. Nous avons intentionnellement suréchantillonné les démocrates et les indépendants afin de cibler les arguments pro-redistribution, qui constituaient notre principal objectif. Nous avons ensuite montré cet "univers d'arguments" à un échantillon distinct de 4010 répondants afin d'évaluer les arguments sur le plan de la conviction et du contenu émotionnel, en particulier la colère.

Figure 9: Les croyances en matière d'équité et d'externalité au fil du temps



Note: Données de wid.world et du General Social Survey.

Nous avons fait trois constatations principales. Premièrement, les arguments d'équité amènent les personnes interrogées à se déclarer beaucoup plus en colère. Cela est dû aux répondants qui sont d'accord avec les arguments, et en partie au fait que les arguments d'équité font plus régulièrement appel aux émotions que les arguments d'externalité. Deuxièmement, nous constatons que les deux types d'arguments sont généralement convaincants - ce qui renforce la conclusion de Lobeck and Støstad [2023] selon laquelle les croyances en matière d'externalité sont comparables aux opinions en matière d'équité en tant que déterminants des préférences en matière de redistribution. Troisièmement, nous trouvons des indications selon lesquelles les répondants à revenu élevé ($>$ \$75000) trouvent de manière disproportionnée les arguments relatifs aux externalités convaincants - bien que cela ne soit pas statistiquement significatif pour la tranche de revenu pré-spécifiée de \$50000, ou pour la tranche supérieure de \$100000 (où la taille de l'échantillon devient faible).

Enfin, nous ne trouvons aucune preuve que les républicains trouvent l'un ou l'autre type d'argument redistributif (équité ou externalité) disproportionnellement convaincant par rapport aux démocrates. Cela diffère des convictions descriptives de Lobeck and Støstad [2023], où les

opinions en matière d'équité sont plus polarisées que les convictions en matière d'externalité au sein des groupes politiques (un résultat que nous reproduisons).[35]. Nous émettons l'hypothèse que cela est dû au fait que les personnes qui sont opposées à la redistribution par principe ne sont influencées par aucun des arguments utilisés pour motiver cette redistribution. En d'autres termes, bien que la plupart des gens souhaitent une distribution économique plus équitable, la méthode utilisée pour y parvenir est souvent cruciale. En d'autres termes, il se pourrait que de nombreuses personnes soient généralement opposées à *n'importe quel* argument en faveur d'une redistribution pilotée par l'État, même si elles partagent l'objectif général de réduction des inégalités économiques.

En résumé, ces résultats indiquent que les différences historiques et entre pays dans les débats sur la redistribution pourraient s'expliquer, du moins en partie, par le *type* d'arguments redistributifs qui sont courants dans un pays ou à une époque donnés. La polarisation affective autour de la redistribution aux États-Unis pourrait être due au fait que l'accent est mis sur l'équité plutôt que sur les motivations liées aux externalités, par exemple. De même, le soutien aux politiques de redistribution dans la classe supérieure pourrait s'expliquer par le fait que les arguments les plus courants en faveur de la redistribution sont fondés sur des idées d'équité ou d'externalité.

Le présent document s'inscrit principalement dans la vaste littérature consacrée aux préférences en matière de redistribution et aux méthodes d'enquête. La littérature sur les opinions en matière d'équité est abondante ; Cappelen et al. [2007], Almås et al. [2020] et Stantcheva [2021] figurent parmi les nombreux articles établissant un lien entre les opinions des individus en matière d'équité et leurs préférences en matière de redistribution. L'angle de l'externalité est examiné théoriquement par Alesina and Giuliano [2011] et Støstad and Cowell [2021], et empiriquement par Lobeck and Støstad [2023].

L'article le plus proche est Lobeck and Støstad [2023], qui montre que les arguments fondés sur l'équité et l'externalité modifient de manière causale les préférences en matière de redistribution. Cet article montre également qu'une vidéo axée sur l'équité amène les répondants à déclarer plus de colère que trois vidéos axées sur l'externalité. Toutefois, ces résultats ne sont évalués que pour ces traitements vidéo spécifiques ; une multitude de choix de conception spécifiques auraient pu avoir un impact sur les résultats. En revanche, le présent document évalue un ensemble plus large d'arguments échantillonnés de manière non biaisée à partir de la distribution de ces arguments dans la population.

Plusieurs autres articles présentent des modèles de recherche similaires à Lobeck and Støstad [2023] [e.g. Kuziemko et al., 2015, Stantcheva, 2021], où le principal objectif de la recherche a été d'établir l'existence d'un lien de causalité entre certaines croyances (par exemple, les croyances en matière d'équité ou la connaissance des politiques) et les préférences en matière de redistribution. Cet article tente plutôt d'élucider d'autres caractéristiques distinctives de ces liens - telles que leur force et leur contenu émotionnel - en utilisant un échantillon plus large d'arguments qui ne sont pas affectés par les choix de conception de la recherche. En effet, nous évaluons 160

[35] C'est-à-dire que la différence entre les partis à travers diverses questions est systématiquement plus élevée pour les questions d'équité que pour les questions d'externalité. Par exemple, les démocrates et les républicains sont tous deux susceptibles de penser que les inégalités augmentent la criminalité. Cependant, seuls les républicains pensent que le travail acharné mène à la réussite ou que la répartition des revenus est globalement équitable

arguments impartiaux au lieu d'une poignée de traitements vidéo spécifiquement conçus. Cela nous permet de donner à nos résultats une portée plus générale.

Nous nous référons également à la littérature sur les récits et leur relation avec les résultats économiques, notamment Alesina et al. [2018a], Roth et al. [2020] et Andre et al. [2022].

En résumé, nous présentons une nouvelle méthodologie basée sur des enquêtes pour évaluer l'efficacité des catégories d'énoncés qui atténue l'influence du biais du chercheur. Nous appliquons cette méthodologie aux arguments redistributifs, en obtenant un échantillon impartial d'arguments fondés sur *les idées d'équité* ou *les conséquences sociétales de l'inégalité* et en évaluant leur efficacité et leur contenu émotionnel dans le cadre de trois enquêtes menées aux États-Unis ($N_1 = 298$, $N_2 = 215$, $N_3 = 4010$). Notre "univers d'arguments" final compte 160 arguments redistributifs et un total de 32 300 évaluations d'arguments. Les personnes interrogées se déclarent beaucoup plus en colère ($p < 0,002$) en réaction aux arguments d'équité par rapport aux arguments fondés sur les conséquences de l'inégalité. Cela s'explique en partie par le fait que l'argument moyen sur l'équité a un contenu plus émotionnel que l'argument moyen sur les conséquences de l'inégalité. Bien que les deux types d'arguments soient globalement convaincants, nous constatons que les individus situés au sommet de la distribution des revenus sont relativement plus influencés par les arguments relatifs aux conséquences de l'inégalité.

Appendix I.

Appendix to Chapter One

I.A. Discussion on Equations 1.1 and 1.2

Equations (1.1) and (1.2) show the following simplification:

$$U_i(x_i(\bar{\theta}), \bar{\theta}, \overrightarrow{\Psi}(\bar{\theta}), ...) \rightarrow \tilde{U}_i(\tilde{x_i}, \bar{\theta}, ...). \tag{A.1}$$

A skeptical reader may argue that we should rather explore each externality channel individually. For example, if we assume that income inequality increases the amount of crime, one might say that one should strengthen the prevention of crime rather than reduce income inequality, or explore the crime-channel more in depth instead of focusing on inequality itself as an externality.

This does not change the intuition of the problem, however. Channel-specific policy solutions would also carry an associated cost which should be modeled in the general framework. The main argument in this work is that different levels of (in)equality carry a shadow price that need to be taken into account when choosing between tax schedules – and not that the income tax is necessarily the only solution to every inequality externality channel.

What does this mean in a practical example? We return to the example of inequality and crime. Suppose the social planner wishes to deal with inequality externality problems indirectly (such as through crime prevention). The required revenue $R = R_0 + p(c(\bar{\theta}))$ is a function of original revenue requirements R_0 and crime prevention p, which is a function of crime c, which is a function of inequality $\bar{\theta}$. In other words, the required revenue R is a function of inequality θ.

The social planner is then faced with a problem of maximizing social welfare $W = \int_i g_i U(x_i, l_i) di$ under the constraint $R_0 + p(c(\bar{\theta})) \leq \int_{\underline{n}}^{\overline{n}} T(nl) f(n) dn$. Crucially, minimizing inequality also increases "effective" revenue R_0 if total revenue collected R is kept constant. In other words, the social planner's goal once again becomes to maximize revenue $\int_{\underline{n}}^{\overline{n}} T(nl) f(n) dn$ and minimize inequality $\bar{\theta}$. Although the magnitude of the revenue-inequality trade-off has changed – in accordance with how expensive crime prevention programs are compared to reducing inequality itself – the core intuition of the problem has not changed. A naive social planner would choose the suboptimal tax schedule, not taking into account that its choice of tax schedule changes

inequality which affects the effective revenue collected.[1]

The main takeaway is that inequality's societal consequences come with a cost. Whether the social planner directly or indirectly deals with this cost is relatively unimportant to the main conclusions sketched in the remainder of the article. Again, the choice simply changes the *magnitude* of the inequality externality.

As an aside, we also note that a similar simplification as (A.1) is usually made implicitly when including consumption x_i in the utility function. The benefit of consumption to individuals is often not just consumption *per se*, but also what consumption brings them – such as improved health, social status, and so on. As in our case, there are many such potential channels that are usually not explicitly modeled, but can be captured in a vector $\overrightarrow{\Psi}'$. In effect, the following simplification is implicitly made,

$$U_i(x_i, \overrightarrow{\Psi}'(x_i), ...) \rightarrow \hat{U}_i(x_i, ...), \tag{A.2}$$

where $\hat{U}_i$ is the modified utility function – the utility function that is largely used in practice. In effect, a consumption-dependent utility function is a useful shorthand for what is in reality a much more complicated concept. The concept we introduce in this paper simply employs the same method with the *distribution* of individual income.

I.B. Other potential mathematical formulations

It is a natural question to ask whether another type of mathematical structure can keep individualist utility functions while modeling resource inequality's societal effects. Here we consider several other ideas and detail where they succeed or fail to capture the complexity of a resource inequality externality.

Social welfare weights In general, utility-based SWWs cannot approximate the effects of an economic inequality externality. The optimal marginal tax rates in the mechanism design case, shown in (A.28), illustrates one case where even best-designed SWWs would fail to approximate the inequality externality.

There are three main reasons for why SWWs poorly approximate a resource inequality externality. The first of these points holds only when discussing a *resource* inequality externality, as we do in most of this paper. The second and third hold under a utility inequality externality as well.

First, such weights discount *utility*, not resources, which implies that the individual's private labor decision is socially optimal. This is not true under an inequality externality. Second, unlike an inequality externality, SWWs cannot change individual behavior.[2] Third, approximating inequality's societal effects – real-world phenomena – through SWWs would imply a break with welfarist traditions in that the social weights would no longer be a purely philosophical concept.

[1] Such a social planner can still have optimal tax rates above the revenue-maximizing rate. The foregone revenue comes with the benefit of lowered inequality, leaving more revenue for the "standard" revenue requirements R_0.

[2] A natural example would be an agent who increases their work effort to avoid a high inequality externality imposed on low-income agents in a heterogeneous inequality externality framework. This changes the implications of the exercise from a pure self-selection problem [Stiglitz, 1982] to an externality and self-selection problem (our problem).

These points emphasize our larger argument, which is that there are three distinct ways to model the consequences of inequality in a welfarist framework; the cumulative effect of diminishing marginal utility, SWWs, and an inequality externality. The former two are distinct from the inequality externality, occur through different mechanisms, and have different policy implications.

We now present a simple example to illustrate how a resource inequality externality can add nuance that cannot be found when only using social weights and the diminishing marginal utility of income. Imagine a world where one agent has seized the vast majority of income and uses this inequality of income to enjoy disproportionate (and socially damaging) political power. All other agents are equally poor. Suppose we reduce the income of the oppressive ruler slightly, all else equal. We evaluate this change in the presence of (i) *only* risk aversion (diminishing marginal utility), (ii) risk aversion and a weighted social welfare function with non-negative weights, and (iii) risk aversion, a weighted social welfare function with non-negative weights, and an inequality externality.[3]

(i) Social welfare is unambiguously reduced, as the top individual's income decreases.

(ii) Social welfare is either reduced or kept constant – the top individual's income decreases, but they might have zero social weight.

(iii) The effect on social welfare is ambiguous. On one hand, the income of the top individual is reduced, reducing their utility and thus social welfare (if their weight is non-zero). On the other, income inequality is reduced, increasing every other agent's utility. The total effect on social welfare depends on the size of the inequality externality. In extreme cases, such as in this example, overall social welfare could increase.

More generally, concentrated income gains lead to unambiguously non-negative welfare changes in standard models. Considering the current academic and social focus on inequality, this could be a troubling feature.

In Appendix I.B.1 we present, for completeness, a simple proof to show that appropriate utility-based SWWs cannot supplant an inequality externality.

Generalized social welfare weights [Saez and Stantcheva, 2016] The generalized social welfare weights method – or income-based SWWs more broadly – make few predictions for individual behavior. As such, appropriately chosen "modified welfare weights", adjusted to include inequality externality concerns, can approximate the mathematical solutions from a resource inequality externality. However, there are two problems with this approach.

First, the weights become dependent on empirically estimated values such as individual labor elasticities or the local Pareto parameter. The intuition behind the elasticity case is simplest to explain. As the individual contribution to the inequality externality depends on the individual's income, the modified weight – which now takes into account the societal effects of income inequality – needs to account for any changes in the individual's labor decision. This is

[3]The 'standard' case here is no risk aversion, a utilitarian welfare function, and no externality. For example, the first case will consider reducing the income of the top earner in a model with risk aversion, a utilitarian social welfare function and no externality.

mathematically done through introducing the labor elasticity into the modified weight, which implies an unintuitive addition of empirical parameters into an otherwise philosophical concept.

Second, the modified weights can turn negative and thus implicitly break the Pareto principle. This happens when the marginal social welfare of income is negative. This explicitly breaks with the assumptions made in Saez and Stantcheva [2016].

Still, this approach might be useful in some cases, including for the cost-benefit analyses mentioned previously. If modified in such a way, the modeler should be aware that the resulting welfare weights would be different in interpretation from the standard approach, as the modified weights would measure both philosophical issues and externality concerns put together. The weights could also have negative values without breaking the Pareto principle.[4]

An additive resource inequality in the social welfare function If we move away from strict SWWs one could imagine a SWF of the form $\int_i g_i U_i(x_i, l_i, ...)di - \Gamma(\bar{\theta})$, as in Sen [1976], among others. Here U_i is a standard individualist utility function and $\Gamma(\bar{\theta})$ is some function of resource inequality. This can mathematically approximate the solutions from a resource inequality externality if and only if the externality does not affect individual behavior. In other words, this is an accurate mathematical representation of the problem if the externality is fully separable and the number of agents is large. We mention this specifically as the mathematical set-up we use in Section 3 makes these assumptions for simplicity. In general, however, any inequality externality that affects individual behavior cannot be captured through such a modified social welfare function. Such a formulation assumes away most consumption-based inequality externality effects, for example. Intuitively it is also less clear to us whether the social planner should care about inequality effects if these effects do not affect individuals themselves.

I.B.1. Proof: The inequality externality cannot be approximated by utility-based social welfare weights

The social planner aims to maximize:

$$W = \int_i g_i U(x_i, l, \theta(\boldsymbol{x}))di$$

Assume that g_i can have variation (social weights), and that $\frac{\partial U}{\partial \theta} \neq 0$ and $\frac{\partial \theta(\boldsymbol{x})}{\partial x_i} \neq 0$ (an inequality externality exists). x_i is income, l_i is work effort, and $\theta(\boldsymbol{x})$ is inequality as a function of all incomes $\boldsymbol{x}$.

It follows from the social planner's first-order conditions for x_i and l_i that for all $g_i \neq 0$:

$$\frac{\partial U(x_i, l_i, \theta(\boldsymbol{x}))}{\partial l_i} = \frac{\partial U(x_i, l_i, \theta(\boldsymbol{x}))}{\partial x_i} + \frac{1}{g_i} \int_j g_j \frac{\partial U(x_j, l_j, \theta(\boldsymbol{x}))}{\partial \theta(\boldsymbol{x})} \frac{\partial \theta(\boldsymbol{x})}{\partial x_i} dj \tag{A.3}$$

We proceed with a proof by contradiction. Say we want to approximate the effect of the inequality externality with new social weights $\hat{g}_i$ without explicitly including θ in the utility function, otherwise keeping the utility function the same. Denote this new utility function $\hat{U}$. If so, $\frac{\partial \hat{U}(x_j, l_j)}{\partial \theta(\boldsymbol{x})} = 0$ and the second term on the right-hand side of (A.3) is zero. The solution to the social planner's problem would thus involve $\frac{\partial \hat{U}(x_i, l_i)}{\partial x_i} = \frac{\partial \hat{U}(x_i, l_i)}{\partial l_i} \forall \hat{g}_i \neq 0$, which is

[4] The negative weights would imply breaking the Pareto principle in *income*, but not *utility*.

equivalent to $\frac{\partial U(x_i,l_i,\theta(\boldsymbol{x}))}{\partial x_i} = \frac{\partial U(x_i,l_i,\theta(\boldsymbol{x}))}{\partial l_i}\forall \hat{g}_i \neq 0$. However, in the correct solution we are trying to approximate, $\frac{\partial U(x_i,l_i,\theta(\boldsymbol{x}))}{\partial x_i} \neq \frac{\partial U(x_i,l_i,\theta(\boldsymbol{x}))}{\partial h_i}\forall g_i \neq 0$. This implies that $g_i \neq 0 \rightarrow \hat{g}_i = 0$, which cannot be the case. Thus there is a contradiction. This follows from the externality creating a difference between the optimal individual and social work decisions, which cannot be introduced through discounting utility with social weights.

An extension shows that the externality cannot be approximated by the individual variables in the utility function. If x_j is changed, (A.3) implies that it will affect the FOC for i. In the modified solution with $\hat{U}$, it has no effect. To correctly specify $\hat{U}(x_i, l_i)$, one would need x_j or l_j. This would amount to including a distributional parameter $\theta(\boldsymbol{x})$ in the individual utility function, again a contradiction.

I.C. Analytical Solution of the OIT Problem

We first solve the problem in a mechanism design framework, where we fully specify the utility function as,

$$U(x, l, n, \bar{\theta}) = \tilde{\mathcal{U}}\left(u(x) - V(l) - \Gamma(n, \bar{\theta})\right), \tag{A.4}$$

where u is utility from consumption (after-tax income) $x \geq 0$, $V(l)$ is the disutility of work $l \geq 0$, and Γ is disutility from post-tax income inequality $\bar{\theta} \geq 0$ (a society-wide parameter, indicated by the overbar) which is potentially dependent on wage-earning ability $n \geq 0$.[5] The functions $\tilde{\mathcal{U}}$, $u(x)$, $V(l)$ and $\Gamma(n, \bar{\theta})$ are continuous and second-order differentiable in their arguments. For the remainder of the proof we will assume that $\tilde{\mathcal{U}}$ is taken into account by the social planner's social welfare function. The function $u(x)$ is concave in x, $V(l)$ is strictly convex in l, and $\Gamma(n, \bar{\theta})$ has no restriction. We also have that $u_x > 0$ and $V_l > 0$ where subscripts indicate partial derivatives. There are a continuum of agents along the wage-earning ability n, with strictly positive density $f(n)$ and a cumulative distribution function $F(n)$. This functional form allows for potential income effects.

At the heart of the model is n, the exogenous wage-earning ability, unobservable to the social planner. There is a continuum of individuals with n varying according to a an exogenous density function $f(n)$, with a cumulative distribution function $F(n)$. Pre-tax earnings are defined as nl, and total consumption is $x = nl - T(nl)$, where $T(\cdot)$ is the tax schedule. The individual maximizes utility by choosing work effort l given n and $T(\cdot)$. The utility-maximising values of consumption and hours worked are written as

$$x(n), l(n). \tag{A.5}$$

Given the individual's choice, the social planner chooses the tax schedule to maximize the social welfare function. We assume this to be an additively separable function of individual utility. Accordingly the problem is,

$$\max_{T(\cdot)} \int_{\underline{n}}^{\overline{n}} W(U(x(n), l(n), \bar{\theta}))dF(n). \tag{A.6}$$

[5] Allowing Γ to depend on n is our way of introducing potentially heterogeneous inequality externalities without this impacting the individual work decision.

The problem (A.6) is subject to three conditions, the first two of which are standard constraints. First, there is the *revenue constraint* for any required amount R of non-redistributive public goods:

$$R \leq \int_{\underline{n}}^{\overline{n}} T(nl(n))f(n)dn. \tag{A.7}$$

For simplicity we assume that $R = 0$.

Second, we have the *incentive-compatibility constraint* from the possibility that an agent with (unobservable) wage-earning ability n could masquerade as an agent with $\hat{n}$. For any person with wage-earning ability n it must be true that:

$$u(x(n)) - V(l(n)) \geq u(x(\hat{n})) - V(l(\hat{n})) \tag{A.8}$$

where $x(\hat{n})$ and $l(\hat{n})$ are, respectively, the consumption and hours worked if the agent masquerades as someone with ability $\hat{n}$, possibly different from n. The IC constraint (A.8) ensures that the agent self-selects into the appropriate tax bracket.

Third, we need to introduce the role of inequality into the model. Individuals experience an amount $\bar{\theta}$ of after-tax inequality. This inequality is partly determined by F, the distribution of innate wage-earning ability, and partly by the choices made by individuals, captured in (A.5). But it is also partly the result of decisions by the social planner, captured in the tax function T and therefore embedded in (A.5). We can represent this relationship as the following *inequality condition:*

$$\bar{\theta} = I\left(\boldsymbol{x}, F\right) \tag{A.9}$$

where $I\left(\cdot,\cdot\right)$ is an inequality measure, $\boldsymbol{x}\left(\cdot\right)$ is the full set of consumption choices from (A.5) and $F\left(\cdot\right)$ is the distribution function for n.

To complete the model we need an inequality metric $I\left(\cdot,\cdot\right)$. We begin with a specific form of the (absolute) Gini coefficient in after-tax income taken from Cowell [2000]:

$$I_{\text{Gini}}\left(\boldsymbol{x}, F^x\right) = \int_{\underline{n}}^{\overline{n}} \kappa^x(x(n))x(n)dF^x(x(n)), \tag{A.10}$$

where x is after-tax income (consumption) with distribution F^x and

$$\kappa^x(x) = 2F^x(x) - 1 \tag{A.11}$$

is an expression for the weight of the agent in the Gini dependent on the cumulative density of post-tax income.

This form of the inequality metric presents a difficult endogeneity problem when taking derivatives for x, namely that the weight itself depends on the distribution of post-tax income. In short, the inequality weight $\kappa^x(x_i)$ depends on x_i and thus $T(z_i)$, which is problematic since all other $\kappa^x(x_j)$ also depend on $\kappa(x_i)$. It can, however, be modified to a simpler form. If there is rank-equivalency between income and ability, we can use the fact that $F^x(x) = F^n(n)$ to see that $\kappa^x(x) = \kappa^n(n)$. Thus we can re-write the inequality metric as,

$$I_{\text{Gini}}(\boldsymbol{x}, F) = \int_{\underline{n}}^{\overline{n}} \kappa(n) x(n) dF(n), \tag{A.12}$$

where

$$\kappa(n) = 2F(n) - 1. \tag{A.13}$$

Here we have removed the superscripts for notational simplicity. This shows that the absolute Gini in post-tax income can be calculated as a sum of weighted post-tax incomes in the population, where the weight $\kappa(n)$ depends only on the *rank* of the agent in the wage-earning ability distribution $F(n)$, which is constant and exogenous by assumption. Simula and Trannoy [2022], developed simultaneously with this paper, also exploits this rank-invariancy in ability and income to establish novel social welfare weights; here we employ the same trick to allow a post-tax income inequality metric in the utility function in the continuous Mirrlees problem. This vastly simplifies the analytical problem. This assumption is equivalent to assuming that the individuals' second-order conditions hold. The idea is that if the second-order conditions hold, we have that $z'(n) > 0$ [Lollivier and Rochet, 1983]. As shown in Salanie [2011] p. 89, this implies $x'(n) > 0$ which implies rank equivalence. For all the numerical simulations we confirm that this rank equivalence holds.[6]

Using (A.10), condition (A.9) becomes

$$I_{Gini} = \int_{\underline{n}}^{\overline{n}} \left[2F(n) - 1\right] x(n) dF(n). \tag{A.14}$$

One can also use other inequality metrics based on rank-specific weights, such as those in the Lorenz [Aaberge, 2000] or S-Gini families [Donaldson and Weymark, 1980], which simply changes $\kappa(n)$.

As an aside, partly to motivate our specification, we note that if the inequality externality $\Gamma(n, \bar{\theta})$ is linear and we are in a Utilitarian framework, the objective function amounts to the SWF derived in Sen [1976] with an additional labor disutility term. This Sen [1976] SWF is also a cumulation of Fehr-Schmidt preferences over the population [Schmidt and Wichardt, 2019].

To solve the analytical problem we first re-write the incentive compatibility constraint. We note that consumption x, i.e. after-tax income, is a function of wage times hours worked: $x = c(nl(n))$. We also define $\tilde{U}(n)$ as the non-externality part of utility, such that $U(n) = \tilde{U}(n) - \Gamma(n, \bar{\theta})$. The individual maximization implies,

$$\frac{d\tilde{U}(n)}{dl(n)} = 0 = u_x(x(n)) c_{nl}(nl(n)) n - V_l(l(n)), \tag{A.15}$$

and from the IC constraint we have (using either the Mirrlees [1971] trick or the envelope condition):

$$\frac{d\tilde{U}(n)}{dn} = u_x(x(n)) c_{nl}(nl(n)) l \tag{A.16}$$

Taken together these two imply :

[6] In the small perturbation approach we use that inequality weights are unchanged with small perturbations around the optimum given that the individual's second-order conditions hold.

$$\frac{d\tilde{U}(n)}{dn} = \frac{V_l(l(n))l}{n} =: \xi\,(n) \tag{A.17}$$

We can write $T = nl(n) - x(n)$, where $x(n)$ is after-tax consumption. From this and the IC constraint, we observe that the tax schedule implicitly defines both work hours and total individual utility. We rewrite $x(n) = y(l(n), \tilde{U}(n)) = u^{-1}(\tilde{U(n)} + V(l(n)))$. Instead of setting the tax schedule T, then, we can say that the social planner chooses work hour schedules $l(n)$, utility schedules $\tilde{U(n)}$, and the inequality level $\bar{\theta}$.

The Lagrangian of the full problem classified in (A.6)–(A.9) is,

$$\begin{aligned} L = \int_{\underline{n}}^{\overline{n}} W\left(\tilde{U}(n) - \Gamma(n,\bar{\theta})\right) f(n)dn + \lambda\Big(\int_{\underline{n}}^{\overline{n}} \left[nl(n) - y(l(n),\tilde{U}(n))\right] f(n)dn\Big) \\ + \int_{\underline{n}}^{\overline{n}} \alpha(n)\left[\frac{d\tilde{U}(n)}{dn} - \xi(n)\right] dn + \gamma\left[\bar{\theta} - I_{Gini}\right] \end{aligned} \tag{A.18}$$

We note that the incentive compatibility constraint can be simplified using integration by parts. After taking these factors into account, combining the rest of the integrals, and substituting in for I_{Gini}, we have:

$$\begin{aligned} L = \int_{\underline{n}}^{\overline{n}} \Big[\left(W\left(\tilde{U}(n) - \Gamma(n,\bar{\theta})\right) + \lambda\left[nl(n) - y(l(n),\tilde{U}(n))\right] - \gamma\kappa(n)y(l(n),\tilde{U}(n))\right) f(n) - \alpha(n)\xi(n) \\ -\alpha'(n)\tilde{U}(n)\Big] dn + \alpha(\bar{n})\tilde{U}(\bar{n}) - \alpha(\underline{n})\tilde{U}(\underline{n}) + \gamma\bar{\theta} \end{aligned} \tag{A.19}$$

From this we can find the first-order conditions with respect to $l(n)$, $\tilde{U}(n)$, and $\bar{\theta}$, as these variables together will implicitly set the tax schedule.[7] Using the rules for derivatives of inverse functions, we have that $y_l = \frac{V_l}{u_x}$ and $y_{\tilde{U}} = \frac{1}{u_x}$. The first order conditions are the following:

$$\tilde{U}: \quad 0 = \left[W_{U(n)}(U(n)) - \frac{\lambda}{u_{x(n)}}\right] f(n) - \alpha'(n) - \gamma\kappa(n)f(n)\frac{1}{u_{x(n)}} \tag{A.20}$$

$$l: \quad 0 = \lambda(n - \frac{V_l}{u_{x(n)}})f(n) - \alpha(n)\frac{V_{ll}l(n) + V_l}{n} - \gamma\kappa(n)f(n)\frac{V_l}{u_{x(n)}} \tag{A.21}$$

$$\bar{\theta}: \quad 0 = \gamma - \int_{\underline{n}}^{\bar{n}} W_{U(p)}(U(p))\Gamma_{\bar{\theta}}(\bar{\theta}, n)f(p)dp \tag{A.22}$$

$$\alpha(\bar{n}) = \alpha(\underline{n}) = 0 \tag{A.23}$$

Where (A.23) are the transversality conditions. In the FOC for l we have used that $g = \frac{V_l l}{n}$ from ((A.17)), and that $\frac{d\xi}{dl} = \frac{V_{ll}l + V_l}{n}$. We use the new symbol p to denote the productivity n inside the integral in (A.22).

[7] We could use the derivative of $x(n)$ instead, but the methods are mathematically equivalent and this procedure is somewhat more straightforward.

(A.22) implies,

$$\gamma = \int_{\underline{n}}^{\overline{n}} W_{U(p)}(U(p))\Gamma_{\bar{\theta}}(\bar{\theta}, p) f(p) dp \tag{A.24}$$

Here γ is the shadow price of the inequality constraint, which is expressed as the welfare-weighted sum of every individual's marginal disutility of the externality $\Gamma_{\bar{\theta}}$. Under identical $\Gamma(n, \bar{\theta})$ across individuals, this implies that $\gamma = \Gamma_{\bar{\theta}}(\bar{\theta})$. Under a linear homogeneous inequality externality such that $\Gamma = \eta\bar{\theta}$, we have $\gamma = \eta$. Notice that potentially heterogeneous effects of inequality in n do not present strong complications. The resulting γ is simply a sum of the heterogeneous externalities weighted by individual's welfare weights.

Now we move to finding an expression for $\alpha(n)$, the shadow price of the incentive compatibility constraint. We integrate the first order condition for $\tilde{U}$, (A.20):

$$\alpha(n) = \int_{n}^{\bar{n}} \left[\frac{\lambda + \gamma\kappa(p)}{u_{x(p)}} - W_{U(p)} \right] f(p) dp \tag{A.25}$$

And substitute this into ((A.21)):

$$0 = \lambda(n - y_l) f(n) - \gamma\kappa(n) f(n) y_l - \frac{V_{ll} l + V_l}{n} \int_{n}^{\bar{n}} \left[\frac{\lambda + \gamma\kappa(p)}{u_{x(p)}} - W_{U(p)} \right] f(p) dp \tag{A.26}$$

$$\frac{(n - y_l)}{y_l} = \frac{\gamma}{\lambda}\kappa(n) + \frac{u_{x(n)}(V_{ll} l + V_l)}{\lambda f(n) n V_l} \int_{n}^{\bar{n}} \left[\frac{\lambda + \gamma\kappa(p)}{u_{x(p)}} - W_{U(p)} \right] f(p) dp \tag{A.27}$$

We have that $\frac{n - y_l}{y_l} = \frac{n u_{x(n)}}{V_l} - 1 = \frac{t(n)}{1 - t(n)} - 1 = \frac{t(n)}{1 - t(n)}$, so we quickly have the expression for optimal marginal tax rates:

$$\frac{t(n)}{1 - t(n)} = \frac{\zeta_n u_{x(n)}}{f(n) n} \int_{n}^{\bar{n}} \left[\frac{1}{u_{x(p)}} - \frac{W_{U(p)}}{\lambda} \right] dF(p) + \gamma \left[\kappa(n) + \frac{\zeta_n u_{x(n)}}{f(n) n} \int_{n}^{\bar{n}} \frac{\kappa(p)}{u_{x(p)}} dF(p) \right], \tag{A.28}$$

$\zeta_n = \frac{V_{ll} l}{V_l} + 1$ is a term closely related to the inverse compensated elasticity of labor.[8] The first two terms are functionally equivalent to the traditional OIT terms.[9]

By denoting the standard part of the optimal tax function as $\frac{t(n)_{orig}}{1 - t(n)_{orig}}$, we can isolate and evaluate the effect of the inequality externality.

$$\frac{t(n)}{1 - t(n)} = \frac{\gamma}{\lambda} \left[\kappa(n) + \frac{\zeta_n}{f(n) n} \int_{n}^{\bar{n}} \frac{u_{x(n)}}{u_{x(p)}} \kappa(p) dF(p) \right] + \frac{t(n)_{orig}}{1 - t(n)_{orig}} \tag{A.29}$$

The externality introduces two new terms;

[8] With quasi-linear preferences, $\zeta = \frac{1}{E_L} + 1$.

[9] There is a potentially subtle difference in the term containing $W_{U(p)}$. If the SWF is concave, the weights implied by this term are dependent on the inequality externality term itself. Take, for example, $W = \int_i log(U_i) di$. Here the introduction of an inequality externality would change $W_{U(p)}$ and thus the optimal tax rates. The numerical simulations we consider in Section 4 take this into account, but in practice the changes due to this factor are minimal.

- (i) a Pigouvian term, $\frac{\gamma}{\lambda}\kappa(n)$, measuring both the size of the externality itself in terms of public funds ($\frac{\gamma}{\lambda}$) and the contribution of the individuals at the given tax bracket to the externality ($\kappa(n)$, which changes sign across the distribution), and
- (ii) a change to the redistributive benefit of the tax, $\frac{\gamma}{\lambda}\frac{\zeta_n u_{x(n)}}{f(n)n}\int_n^{\bar{n}}\frac{\kappa(p)}{u_{x(p)}}dF(p)$, in effect modifying the SWWs. Beyond standard Mirrleesian parameters, this latter term depends on both the size of the externality in terms of public funds $\frac{\gamma}{\lambda}$ and a measure similar to the total externality weight above the tax bracket, $\int_n^{\bar{n}}\frac{\kappa(p)}{u_{x(p)}}dF(p)$.

Here $\frac{\gamma}{\lambda}$ is the shadow price of inequality in terms of public funds.[10] If inequality is a negative externality (a public bad), $\frac{\gamma}{\lambda}$ will be positive. To rephrase, this is the unsurprising result that equality itself has a cost in a world with a negative inequality externality. Similarly, a Rawlsian social planner will only take into account the inequality externality on the lowest-utility agent. If we assume a linear inequality externality of the form $\Gamma(\theta) = \eta\theta$ then $\gamma = \eta$ – see (A.24). With a squared inequality externality, which we discuss specifically in Appendix I.C.1, the term in the utility function is $\eta(\bar{\theta} - \theta_{opt})^2$ and $\gamma = 2\eta(\bar{\theta} - \theta_{opt})$, which implies that the effect of the externality on the optimal tax schedule is dependent on the distance from the optimal inequality level θ_{opt} .

This solution illustrates both similarities and differences between our approach and the standard Mirrlees externality literature. In Kanbur and Tuomala [2013], for example, where the externality is a flat negative consumption externality, there are also two new terms to the Mirrlees [1971] formula; a Pigouvian term and a SWW modification. However, as the marginal externality effect in Kanbur and Tuomala [2013] is constant across the distribution, the analytical modification to the tax schedule is relatively independent of the location of the tax bracket.[11] This is not true in our specification. The modification to optimal marginal tax rates is now strongly dependent on the location of the tax bracket in the distribution. This location-dependence can be seen in both the marginal externality effect of the agent *in* the tax bracket (κ in the first term), and in the average marginal externality of all agents *above* the tax bracket ($\bar{\kappa}$ in the second term).

We now briefly discuss how income effects affects the solution. Individuals above the tax bracket now also react to tax increases by increasing their labor supply. Mathematically this enters (A.29) through $u_{x(n)}$ decreasing with n and ζ_n changing across the distribution. This is too complicated to easily assess further; we leave it for later works.

To return to the assumptions in the main body, let us assume a linear homogeneous inequality externality ($\Gamma(\bar{\theta}) = \eta\bar{\theta}$) and quasi-linearity in consumption. The resulting utility function is

$$U(x, l, \bar{\theta}) = x - \frac{l^{\left(1+\frac{1}{E_L}\right)}}{\left(1+\frac{1}{E_L}\right)} - \eta\bar{\theta}.$$

The optimal tax rate condition simplifies to:[12]

[10] As noted in Jacobs [2018], the marginal cost of public funds λ is one at the optimum.

[11] Note as well that there is a mistake in the derivation of $\frac{\gamma}{\lambda}$ in Kanbur and Tuomala [2013], which we have corrected here.

[12] Note that with quasi-linearity, $\int_n^{\bar{n}}\kappa(p)dF(p)$ in (A.29) simplifies as $\int_n^{\bar{n}}(2F(n)-1)\,dF(n) = F(n) - F(n)^2$.

$$\frac{t(n)}{1-t(n)} = \eta\kappa(n) + \eta\left(\frac{1}{E_L}+1\right)\frac{\bar{\kappa}(n)}{\alpha(n)} + \frac{t(n)_{orig}}{1-t(n)_{orig}}, \tag{A.30}$$

where we denote the local Pareto parameter $\frac{f(n)n}{1-F(n)}$ as $\alpha(n)$. This is equivalent to (1.5) and is what we employ in the main numerical simulations.

I.C.1. A squared inequality externality function

Our framework is sufficiently general for other functional forms of the MRS, or equivalently $\Gamma(n,\bar{\theta})$, the inequality function from the utility function (see Appendix I.C). Let us use $\Gamma(\bar{\theta}) = \eta(\bar{\theta}-\bar{\theta}_{opt})^2$, such that:

$$U(x,l,\bar{\theta}) = x - \frac{l^{\left(1+\frac{1}{E_L}\right)}}{\left(1+\frac{1}{E_L}\right)} - \eta(\bar{\theta}-\bar{\theta}_{opt})^2 \tag{A.31}$$

The resulting analytical optimal tax rates are:

$$\frac{t(n)}{1-t(n)} = 2\eta(\bar{\theta}-\bar{\theta}_{opt})\left[\kappa(n) + \frac{\zeta}{f(n)n}\int_n^\infty \kappa(p)f(p)dp\right] + \frac{t_{orig}}{1-t_{orig}} \tag{A.32}$$

Comparing these tax rates to (A.29), we see that the effect of the inequality externality is attenuated by a factor of $2(\bar{\theta}-\bar{\theta}_{opt})$. The policy effect of the inequality externality will be larger in societies with high after-tax inequality. We find this intuitive; tax systems responding to inequality will respond more when initial inequality is high. The result is the same when using the small perturbations method.

Also note that this solution is endogenous, as $\bar{\theta}$ depends on the tax schedule. We thus need numerical methods to solve for the optimal tax schedule. This is not a unique feature of this formulation, and also occurs when the social weights are endogenous as in the non-Rawlsian solutions.

We do not perform numerical simulations in this case, primarily because of the complicated nature of estimating a suitable η when we have another unknown variable in $\bar{\theta}_{opt}$.

I.D. Small Perturbation Solution to the OIT Problem

The core part of this approach follows Saez [2001] and Saez and Stantcheva [2016].

We introduce a small tax reform $dT(z)$ where the marginal income tax is increased by $\partial\tau$ in a small band from z to $z+dz$. The reform mechanically increases average tax rates on everyone above this band. This is the mechanical effect of taxation, and collects $dz\partial\tau$ from $1-H(z)$ agents above z under the assumption of no income effects. Thus it collects $[1-H(z)]\,dz\partial\tau$ revenue. For each $dz\partial\tau$ collected, however, inequality also changes. The magnitude of this change per agent above differs based on which agent is considered. Noting that income rank $\kappa(z)$ does not change given that second-order conditions hold, each decrease in one unit of post-tax income at z changes absolute post-tax income inequality by $\kappa(z)h(z)$ (from (1.3)).[13] The mechanical effect thus has a differing equality effect of $-\kappa(z_j)h(z_j)dz\partial\tau$ at each point j above z, where z_j

[13] Note that $\kappa(z)$ is negative at the bottom of the distribution.

is the income of the agent and $h(z_j)$ is the number of agents at this point, and $\kappa(z_j)$ is that agent's weight in the inequality metric. As the income change of each agent above z is equal, we can define the average inequality weight above as $\overline{\kappa}(z)\left[1-H(z)\right]=\int_{\{j:z_j>z\}}\kappa(z)h(z)dj$ and write that the mechanical effect changes income inequality by $d\bar{\theta}_M=-\overline{\kappa}(z)\left[1-H(z)\right]dz\partial\tau$.[14]

Those who are located in the small band between z to $z+dz$ have a behavioral response to the tax change. They work less, and reduce their pre-tax earnings by an amount $\partial z=-\epsilon(z)z\partial\tau/\left(1-\tau(z)\right)$. $\epsilon(z)$ is the elasticity of earnings z with respect to $1-\tau(z)$. There are $h(z)dz$ individuals in the tax bracket who were taxed at $\tau(z)$ before the perturbation, so total revenue decreases by $-dz\partial\tau\cdot\epsilon(z)zh(z)\tau(z)/\left(1-\tau(z)\right)$. This change in total earnings is moderated by an effect $(1-\tau)/\tau$ for the inequality effect, as we are interested in the post-tax income decrease and not the tax revenue decrease.[15] Additionally we must multiply by the agents' weight in the inequality metric $\kappa(z)$. The behavioral response thus has an effect on the post-tax income inequality metric as $d\bar{\theta}_B=-\kappa(z)\cdot dz\partial\tau\cdot\epsilon(z)zh(z)$.

The total revenue effects are:

$$dR=dz\partial\tau\left(1-H(z)-\epsilon(z)zh(z)\tau(z)/\left(1-\tau(z)\right)\right)$$

The direct welfare effect through the individual income channels is $\int_j g_j dRdj$ for $z_j\leq z$ and $-\int_j g_j(\partial\tau dz-dR)dj$ for $z_j>z$. Thus the net individual income-based welfare effect is $dM+dB+dW=dR\cdot\int_j g_j dj-dz\partial\tau\int_{\{j:z_j\geq z\}}g_j dj$.

The total equality effect is $\partial\bar{\theta}=d\bar{\theta}_M+d\bar{\theta}_B$:

$$\partial\bar{\theta}=dz\partial\tau\left(-\overline{\kappa}(z)\left[1-H(z)\right]-\kappa(z)\epsilon(z)zh(z)\right) \tag{A.33}$$

In terms of welfare, the effect is $\int_j g_j\frac{\partial U_j}{\partial\bar{\theta}}\partial\bar{\theta}dj$. We have that $\eta_i=MRS_{x_i\bar{\theta}}=-\frac{\partial U_i/\partial\bar{\theta}}{\partial U_i/\partial x_i}$, and thus the total welfare effect of the inequality change is $dI=\int_j(-g_j\eta_j)\partial\bar{\theta}\cdot dj=-\partial\bar{\theta}\int_j\eta_j g_j dj$.

The total welfare change, including all channels, is equal to zero at the optimum:

$$dM+dB+dW+dI=0.$$

Note that in the main text we denote $dI=dI_B+dI_M$ where dI_B and dI_M correspond to the welfare-weighted versions of $d\bar{\theta}_B$ and $d\bar{\theta}_M$ respectively. Thus, using the expressions for dR and dI, and the expression $\bar{G}(z)\left(1-H(z)\right)=\int_{\{j:z_j\geq z\}}g_j dj/\int_j g_j dj$, we have:

$$\begin{aligned}dz\partial\tau\int_j g_j dj\left[1-H(z)-h(z)\epsilon(z)z\frac{\tau(z)}{1-\tau(z)}\right]-dz\partial\tau\bar{G}(z)\left(1-H(z)\right)\int_j g_j dj\\+\int_j\eta_j g_j dj\cdot\left[dz\partial\tau\left(\overline{\kappa}(z)\left[1-H(z)\right]+\kappa(z)\epsilon(z)zh(z)\right)\right]=0\end{aligned}$$

Dividing by $zh(z)\epsilon(z)\int_j g_j dj\cdot dz\partial\tau$ and re-arranging, we find:

$$\frac{\tau(z)}{1-\tau(z)}=\eta\cdot\kappa(z)+\frac{1-H(z)}{z\cdot h(z)}\frac{\left(1-\bar{G}(z)+\eta\overline{\kappa}(z)\right)}{\epsilon(z)}, \tag{A.34}$$

[14] In the absolute Gini, $\overline{\kappa}(z)=H(z)$.

[15] For the mechanical effect, the tax revenue increase and the individual post-tax income decreases are identical.

where we have used the weighted average of the externality $\eta = \int_i g_i \eta_i di / \int_i g_i di$. By using the local Pareto parameter $\alpha(z) = \frac{z \cdot h(z)}{1-H(z)}$ we find the optimal marginal income tax rates as specified in (1.5).

I.E. Additional notes for Section 3

I.E.1. Numerical simulation specifications

Calibrating the model In the traditional optimal tax literature, tax rates are largely determined by three factors [Mankiw et al., 2009]; (i) the shape of the wage-earning ability distribution, (ii) the social welfare function, and (iii) labor or earnings elasticities.

The first factor is the shape of the wage-earning ability distribution $f(n)$,which is well-known to be important in such simulations [see e.g. Tuomala, 2016]. Our main specification backs out the wage-earning ability distribution from the observed pre-tax labor income distribution. We use the DINA microfiles detailed in Piketty et al. [2018] to measure the U.S. pre-tax labor income distribution in 2019.[16]

We then apply the NBER TAXSIM model to find marginal tax rates for any given tax unit in the DINA files.[17] In applying the TAXSIM model we add the number of dependents, the age of the tax-payer, and marital status for each representative tax unit in the DINA files to calculate corresponding real-world marginal tax rates. We then add an assumed 5% state tax, a 2.9% tax rate for Medicare, and a 2.3% sales tax rate, following Saez et al. [2012] and Hendren [2020].[18] We show a Kernel-smoothed version of the TAXSIM marginal tax rates in Figure E1a. Given these marginal tax rates and the empirical pre-tax income data, we assume individuals have correctly optimized according to the utility function in (1.6) and back out the resulting wage-earning ability of each observation.[19] We then estimate the full post-tax wage-earning ability distribution through a Kernel density estimator with a wage-earning ability bandwidth of \$5,000.[20] We assume a constant Pareto distribution for the last 0.5% of the distribution (above ∼\$600,000 in income), where data is sparse. The local Pareto parameter for this top region is set equal to the value immediately before the cut-off. This yields the final wage-earning ability distributions $f(n)$. In addition to this empirical wage-earning ability distribution, we also present the optimal tax rates for two standard theoretical distributions in Appendix I.E.2.

For illustrative purposes and to accompany the inverse optimum exercise in Section 4.5, we show the local Pareto parameter for pre-tax income in the DINA files, $\frac{z \cdot h(z)}{1-H(z)}$, in Figure E1b. This is calculated by using a Kernel density estimator directly on the DINA files, and is thus not used in the calculation of n.

The second factor shaping the optimal tax rate is the social welfare function. To span the range of non-increasing social welfare functions we use two extremes; (i) a fully Utilitarian SWF,

[16] As the Mirrlees model focuses on labor effort, we focus our analysis on labor income.

[17] Described in Feenberg and Coutts [1993], accessed at https://taxsim.nber.org/ on April 20th 2023.

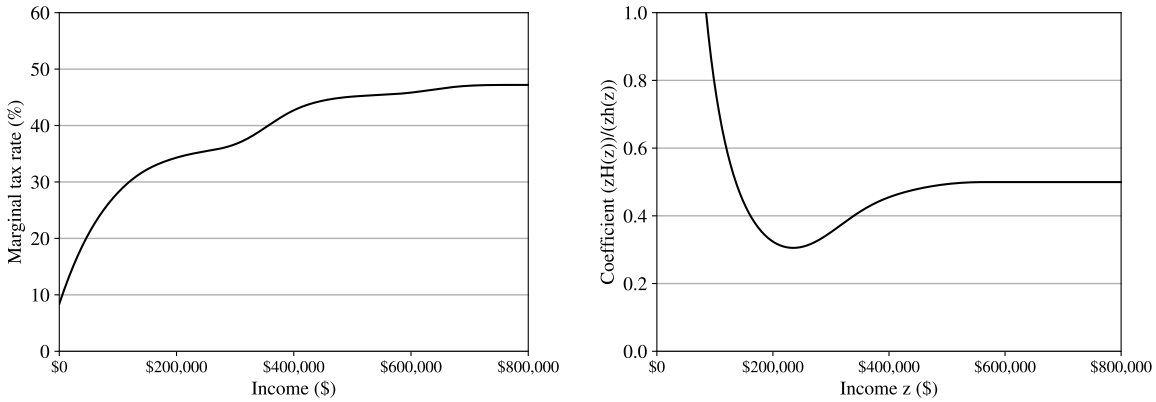
[18] We do not take into account state-based EITC benefits and other deductions; as discussed in Hendren [2020] this is unlikely to significantly affect results.

[19] We use that $(1-\tau(z)) = -U_z/U_x$ from the individual's first-order condition. This together with (1.6) indicates that;

$$n = \frac{x(z)^{1/(1+E_L)}}{(1-\tau(z))^{E_L/(1+E_L)}}$$

[20] This bandwidth corresponds to roughly \$80,000 in the income distribution.

Figure E1: (a) Hazard rate from 2019 U.S. income distribution, (b) 2019 marginal tax rates



Note: 2019 U.S. marginal income tax rates for the simulations, taken from the NBER TAXSIM tool. Left: Right: Hazard ratio $(1 - H(z))/(zh(z))$ for the U.S. pre-tax labor income distribution in 2019. The Kernel estimator bandwidth is \$80,000.

and (ii) the Rawlsian min-max, which implies that the objective function of the government is to optimize the welfare of the worst-off member of society. In comparing to this most inequality-averse SWF we illustrate how the individual inequality concerns from an inequality externality are functionally distinct from the social inequality concerns from SWFs.

The third of these factors are the individuals' labor elasticities. We keep these homogeneous for simplicity in our analysis, assuming that the elasticity of labor supply is constant at $E_L = 0.3$ for all income levels, a reasonable mid-range value from empirical estimates. While this choice is naturally crucial for the optimal tax rates themselves, the numerical effects of introducing an inequality externality is relatively similar across reasonable values of E_L (not shown).

The numerical simulations were performed in Python through an iterative process. We assume an initial tax schedule, set agents' labor supply based on this tax schedule, and then calculate the resulting optimal tax rate. We iterate on this process until an optimum is found. The method is further discussed in the Appendix of Mankiw et al. [2009]. Note that the Rawlsian case can be solved analytically and thus do not require an iterative loop. For every result we check that the individual's second-order conditions hold using two different methods; first we ensure that earnings increases over ability [Lollivier and Rochet, 1983], and second we numerically ensure that the incentive compatibility constraint is satisfied for every agent.

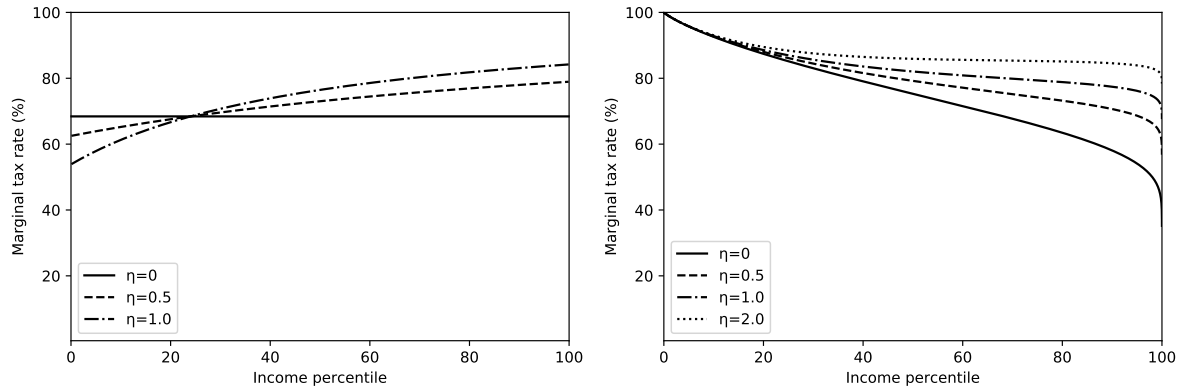
I.E.2. Theoretical ability distributions

We present Rawlsian optimal marginal income tax rates from two theoretical skill distributions in Figure E2, using the Gini as the inequality metric. The first is is a Pareto distribution with $\alpha(n) = 2.0$, which becomes nearly identical to the empirical case at the top of the distribution.[21] The second is a lognormal distribution with $\mu = 2.757$ and $\sigma = 0.5611$, using the values from Mankiw et al. [2009] based on the 2007 U.S. wage distribution.

The Pareto case in Figure E2a illustrates the potentially positive effect of behavioral responses at the bottom. It is socially beneficial for low-income individuals to increase their

[21] Under this Pareto distribution, second-order conditions fail at the bottom for $\eta = 2.0$. This is therefore not plotted.

Figure E2: Optimal Taxation with Inequality Externalities: Theoretical Ability Distributions



Note: Optimal marginal tax rates for various negative Gini-based post-tax income inequality externality magnitudes η_G. The social planner is Rawlsian and the productivity distribution is (a) a Pareto distribution with $\alpha(n) = 2.0$, (b) a lognormal distribution with $\sigma = 0.5611$ and $\mu = 2.757$. Inequality aversion estimates indicate $\eta_G = 1.0$. The solid line, $\eta = 0$, is the standard case of no inequality externality. See Table 2 for further explanation of the inequality externality magnitudes. The $\eta_G = 2.0$ case is excluded from the Pareto simulation because second-order conditions fail at the bottom. The elasticity of labor E_L is 0.3.

incomes – so that inequality is reduced – which leads to a small income subsidy at the bottom as compared to the no-externality case. The goal of this tax subsidy is to make individuals internalize that their increased labor supply leads to positive societal outcomes.

The lognormal case further illustrates the localized effects at the top of the distribution. The standard top marginal tax rate in the lognormal case is 0%. With an inequality externality of $\eta = 2.0$ that increases to 67%. This illustrates the Pigouvian correction at the top, and is salient given the local "zero tax at the top"-result of standard models. This local result is not visible in the graph, but is borne out out in the simulations. At the 99$^{\text{th}}$ percentile the marginal tax rate increases from 39% in the standard case to 79% when $\eta = 2.0$.

I.E.3. Varying inequality metrics

In the main specification we used the absolute Gini coefficient for our measure of inequality. Here we explore two different families of inequality metrics. The first is the top income shares also shown in the main text. The second is the S-Gini, which approximates the Gini with a larger focus on either end of the distribution. The distributional weights implied by both families are plotted in Figure E3.[22]

I.E.3.1. Approximating top income shares

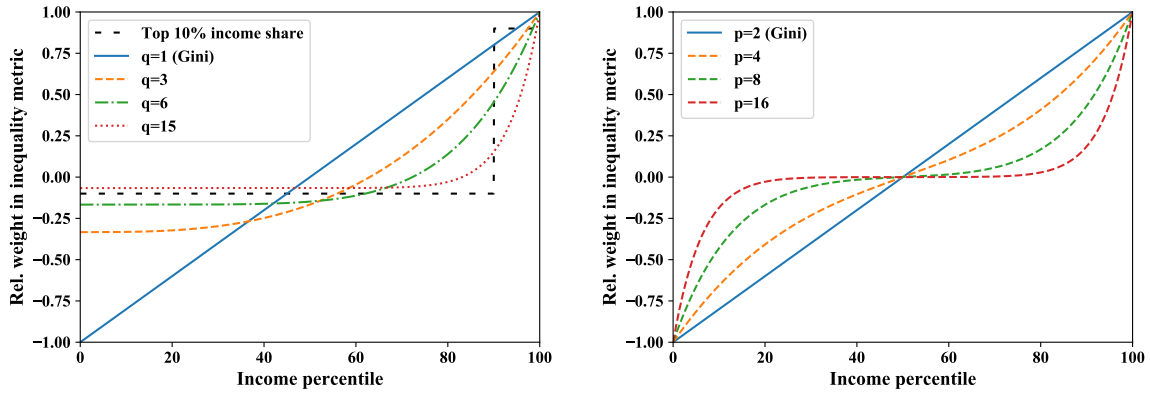
The first family of inequality metrics, also used in the main robustness test, has some of the properties of top income shares. It is,

$$\bar{\theta} = \int_0^{\infty} \left[(q+1)F(n)^q - 1\right] x(n) dF(n), q \in \mathbb{N}. \tag{A.35}$$

When $q = 1$, this becomes the absolute Gini coefficient. In all cases, perfect equality implies

[22]The weights in Figure E3 are normalized such that the top weight is always 1.00. This normalization has no impact on our results due to our re-calculation of η before simulations.

Figure E3: Weights for Families of Inequality Metrics



Note: Consumption weights for inequality metrics used in Appendix I.E.3. For each individual, their impact on the inequality metric is their proportional weight multiplied by their income. In both figures, the Gini is plotted in solid blue. (a) A family of inequality metrics similar to top income shares, as in (A.35). The top 10% income share is plotted in dotted black for reference. (b) The S-Gini family from (A.37).

$\bar{\theta} = 0$ and perfect inequality implies $\bar{\theta} = \mu$ (or $\bar{\theta} = 1$ in the non-absolute family). For increasing q, this indicates an increased focus on the very top of the distribution. The negative externality at the top becomes increasingly concentrated at the very top with increasing q, while the positive externality at the bottom becomes approximately constant for an increasing fraction of the population. In effect, increasing q leads to a metric closer to top income shares, but without the discontinuities that make the analytical problem intractable.

The resulting analytical optimal tax rates with the utility function in 1.6 become,

$$\frac{t(n)}{1-t(n)} = \eta_q \left[((q+1)F(n)^q - 1) + (1 + \frac{1}{E_L}) \frac{1}{f(n)n} \left[1 - F(n)^q\right] F(n) \right] + \frac{t_{orig}}{1 - t_{orig}}. \quad \text{(A.36)}$$

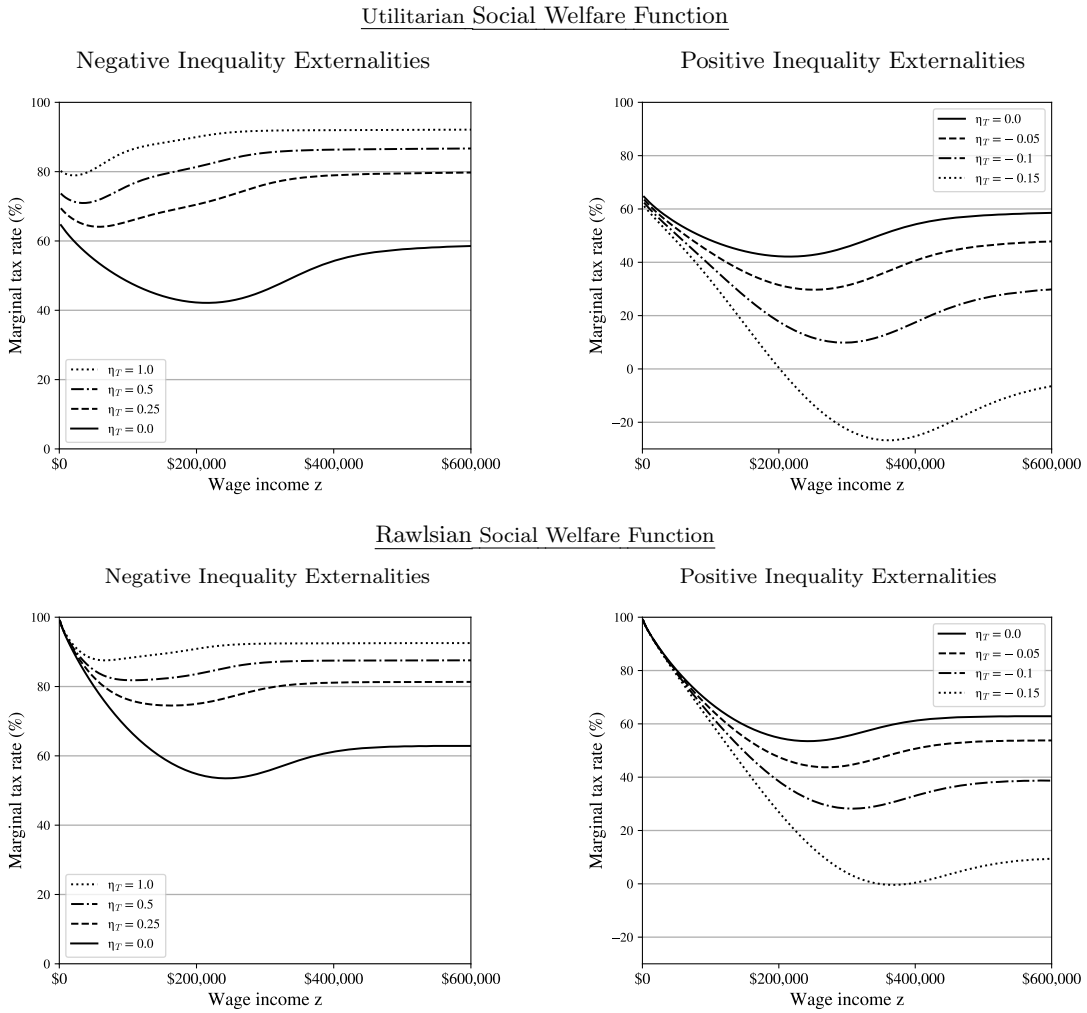
Here η_q is the magnitude of the inequality externality, which is dependent on q when fitting to empirical data. We ensure that values of η_q are comparable over simulations by re-calculating the variable from experimental data for each q.[23]

In Figure E4 we replicate Figure 4 for this inequality metric and $q = 4$. The externality effects are larger at the top and smaller at the bottom when using the top income share metric. With either a Utilitarian or Rawlsian SWF, the optimal top marginal income tax rate goes from 68% in the no-externality case to 90% when $\eta_T = 0.5$ (comparable to $\eta_G = 1.0$ in Figure 4, the value closest to the empirical externality estimate taken from Carlsson et al. [2005]). For the largest negative externality, $\eta_T = 1.00$, the optimal top marginal tax rate is 94%. For the largest positive externality, $\eta_T = -0.15$, the optimal top marginal tax rate is only 26%.

In the Utilitarian case, the effects near the bottom are now relatively small. The negative externalities increase optimal marginal tax rates by around fifteen percentage points at most near

[23] We estimated η with data from Carlsson et al. [2005] in the main text. To remain consistent, we have calculated for each inequality metric q comparable η_q from the experimental values in Carlsson et al. [2005] for all following simulations. This means that, while the value of η_q changes, the underlying estimation comes from the same data. This is true for all metrics.

Figure E4: Optimal Marginal Income Tax Schedules with Top Share Inequality Externalities



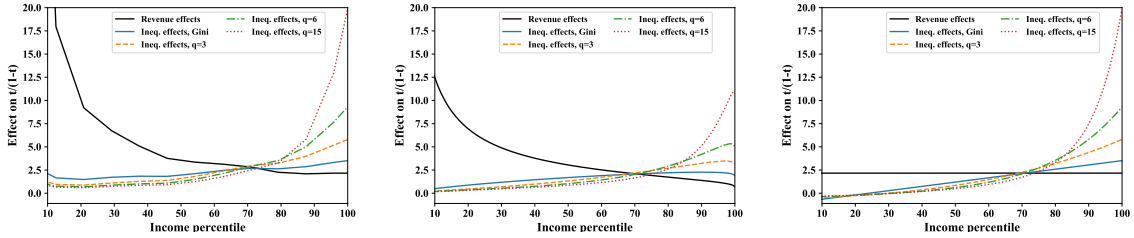
Notes: Optimal marginal tax rates for various top share-based inequality externalities with magnitudes η_T where inequality is either a negative externality (left) or a positive externality (right). The social planner is Utilitarian (above) and Rawlsian (below). The two cases converge when moving towards the top. Empirical estimates indicate $\eta_T \approx 0.5$. The solid line, $\eta = 0$, is the standard no-externality case. Note the different scales of the vertical axes between the negative and positive externalities.

the bottom, whereas the positive externalities have hardly any impact in the region. Around the top, the effects are now larger; the optimal marginal tax rates near the 97$^{\text{th}}$ percentile change from 42% in the no-externality case up to 89% under a negative externality ($\eta_T = 1.00$) and down to -32% under a positive externality ($\eta_T = -0.15$). Negative optimal marginal rates are observed between the 87$^{\text{th}}$ and 99$^{\text{th}}$ percentiles when $\eta_T = -0.15$.

Similarly, the top Rawlsian tax rates can now be negative close to the top. If $\eta_T = -0.15$, optimal marginal tax rates begin at near a hundred percent and go below zero between the 96$^{\text{th}}$ and the 99$^{\text{th}}$ percentiles. Near the bottom, Rawlsian marginal tax rates remain similar to the Gini case.

To further illustrate how increasing q has a large effect on top marginal tax rates, we show the effect of both standard revenue considerations and the new equality considerations on $\frac{t}{1-t}$ with varying inequality metrics in Figure E5. We present this figure for several different underlying ability distributions. The interaction of equality and revenue considerations can make it difficult to interpret values of t, so this graph illustrates the more intuitive impact on $\frac{t}{1-t}$. All social planners are Rawlsian.[24]

Figure E5: Effects on $\frac{t}{1-t}$: Top Income Share Externalities



Note: Effects on $\frac{t}{1-t}$ for various negative inequality metrics $\int_0^\infty [(q+1)F(n)^q - 1]\, x(n) dF(n), q \in \mathbb{N}$. The social planner is Rawlsian. The magnitude of the inequality externality is in each case calculated as the median value from the empirical inequality aversion estimates in Carlsson et al. [2005]. This is done for comparability across inequality metrics. The productivity distribution is (a) the empirical income distribution, (b) a log-normal distribution with $\sigma = 0.39$ and $\mu_{log} = -1$, and (c) a Pareto distribution with $a = 2$. See Figure E3 for an explanation of the inequality metrics. In particular, larger q indicates that top incomes are increasingly weighted. The elasticity of labor E_L is 0.3.

Several points are worth noting. First, as expected, increasing q leads to a more pronounced effect at the top of the distribution in all cases. Second, below the top the effects of changing the metric are small and generally dampen the effect of the externality. Third, equality considerations are relatively constant over different skill distributions; the major factor changing resulting tax rates over skill distributions are revenue considerations. Fourth, equality considerations are proportionally more important than revenue considerations towards the top of the distribution in all three cases. While by nature dependent on the ability distribution and social welfare function, this last point seems likely to hold in many specifications.

I.E.3.2. The S-Gini

The second family of inequality metrics we use is the S-Gini family, which increases the weight of top- and bottom-incomes symmetrically.

[24] Equality considerations would not change with any other SWF due to the homogeneous nature of the externality. Revenue effects would decrease at the bottom and converge to the same at the top.

$$\bar{\theta} = \int_0^{\infty} \left[F(n)^p - (1 - F(n))^p\right] x(n) dF(n), p \geq 2. \tag{A.37}$$

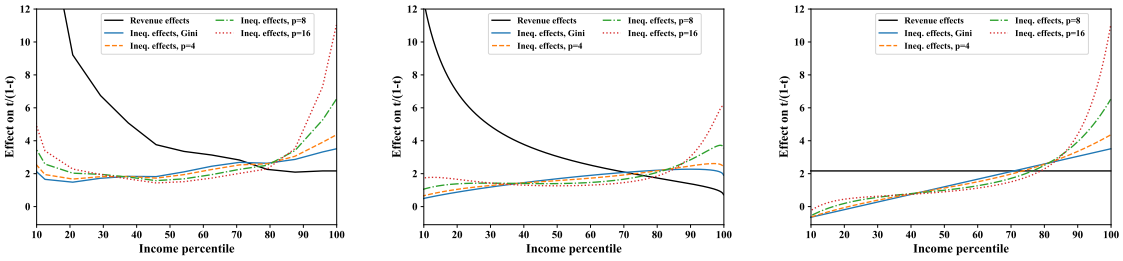
When $p = 2$, this becomes the absolute Gini coefficient. This family also retains the beneficial properties discussed above; perfect equality implies $\bar{\theta} = 0$ and perfect inequality implies $\bar{\theta} = \mu$. For increasing p, the top and bottom is increasingly weighted at the cost of middle incomes. Unlike the previous family, these metrics will always increase if an individual above the median increases their income, as well as decrease if an individual below the median increases their income. The resulting optimal tax rates with the utility function in 1.6 are,

$$\frac{t(n)}{1 - t(n)} = \eta_p \left[(F(n)^p - (1 - F(n))^p) + (1 + \frac{1}{E_L}) \frac{1}{f(n)n} \nu \right] + \frac{t_{orig}}{1 - t_{orig}}, \tag{A.38}$$

where $\nu = \frac{1}{p+1} \left[1 - [F(n)^{p+1} + (1 - F(n))^{p+1}]\right]$.

In Figure E6 we show the effect of changing p on $\frac{t}{1-t}$ with the same methodology as in Figure E5. Increasing p again leads to larger effects towards the top of the distribution and relatively small changes at the bottom. It is notable that the effects at the bottom remain small despite the increased magnitude of the positive externality on these individuals' income. This is driven by the opposition of the mechanical and behavioral channels discussed in the main text. Both equality effects – the internalization of the externality and the increased want for equality – move in the same direction at the top, but work against each other near the bottom.

Figure E6: Effects on $\frac{t}{1-t}$: The S-Gini Family



Note: Effects on $\frac{t}{1-t}$ for various S-Ginis. The social planner is Rawlsian. The magnitude of the inequality externality is held constant for all p at the upper bound of the median value from the empirical inequality aversion estimates in Carlsson et al. [2005]. The productivity distribution is (a) the empirical income distribution, (b) a log-normal distribution with $\sigma = 0.39$ and $\mu_{log} = -1$, and (c) a Pareto distribution with $a = 2$. See Figure E3 for an explanation of the inequality metrics. In particular, larger p indicates that top and bottom income variation is weighted more than middle-income variation. The elasticity of labor E_L is 0.3.

The majority of the new insight noted in the previous subsection also hold for the S-Gini. Unlike in the top income shares, however, the benefits of taxing near the bottom also increase with increasing p. This is a somewhat surprising result. It is due to the mechanical effect being more potent when bottom externalities are very large; in effect, the average inequality metric weight above increases rapidly near the bottom. This leads to the generally large equality benefits from the mechanical effect being even larger than the increased benefits of subsidizing the poor to work more. We caution that this is a particularly model-driven result.

A last caveat; throughout the paper we use a family of *absolute* inequality metrics. This is

done to keep scale independence in the additive utility function while avoiding endogeneity in the optimal tax formula. However, this also implies that the externalities we use induce an incentive for lower average income, which could affect results. The main difference between absolute and non-absolute inequality metrics is that absolute inequality metrics remain unchanged for flat across-the-board income changes, whereas non-absolute inequality metrics remain unchanged for proportional across-the-board income changes. For an exploration of non-absolute inequality metrics, see Aronsson and Johansson-Stenman [2020].

I.F. The Laffer Curve

The central idea of the Laffer curve is simple and true; above a certain tax threshold revenue drops with increased taxation. However, the Laffer curve is often also described as an upper bound on sensible taxation. Laffer [2004] describes this as the "prohibitive range" of taxation, and Manning [2015] argues that "one would not want a rate higher than the Laffer rate".

In the presence of an inequality externality the above statements could be either misleading or false. The externality negligibly changes agent behavior when there is a large number of agents, so the revenue-maximizing rate does not change. However, the welfare-maximizing rate can change, and is in fact often above the Laffer rate given the public benefit of distributional changes.[25] The optimal income tax rate can be higher than the revenue-maximizing rate both at the top (given a negative externality), and at the bottom (given a positive externality). Specifically, the optimal marginal income tax rate is higher than the revenue-maximizing marginal income tax rate if, using the framework in (1.5),[26]

$$\eta\alpha(z)\epsilon(z)\kappa(z) + \eta\bar{\kappa}(z) > \bar{G}(z),$$

that is, if the equality effects of taxation are larger than the welfare effects.[27] $\kappa(z)$ is negative at the bottom and positive at the top of the income distribution, and η changes sign depending on the direction of the externality. Thus the inequality can hold either at the bottom (with a positive externality, $\eta < 0$) or at the top (with a negative externality, $\eta > 0$). [28]

The Mirrlees literature occasionally uses the revenue-maximizing rate as a necessary upper

[25] As an example, consider a society with ten agents, one vastly more wealthy than the other nine. Given the desirability of equality, the welfare-maximizing top marginal rate can be higher than the revenue-maximizing rate, which is zero at the top according to standard results. The Rawlsian simulations in Section 4 provide numerical examples.

[26] In the most general framework, see Appendix I.C, this is equal to,

$$\gamma\left[\kappa(n) + \frac{\zeta u_{x(n)}}{f(n)n}\int_n^\infty \left[\frac{\kappa(p)}{u_{x(p)}}\right] f(p)dp\right] > \frac{\zeta u_{x(n)}}{f(n)n}\int_n^\infty \left[W'(U(p))\right] f(p)dp, \tag{A.39}$$

which represents the same intuition; the equality effects of taxation must be larger than the welfare effects.

[27] This follows from comparing (I.D) to the revenue-maximizing tax rate, which is the same equation when $\bar{G}(z) = 0$ and $\eta = 0$.

[28] In the Rawlsian case, the right-hand side of (A.39) is zero above the very bottom earner. Thus, using the Gini values and a negative externality, the inequality simplifies to

$$\frac{H(z)}{\alpha(z)\epsilon(z)} > 1 - 2H(z), \tag{A.40}$$

which is independent of η and holds for any income above the median. This is intuitive; the Rawlsian rate is the revenue-maximizing rate, and the incentive for equality increases tax rates at least above the median agent. For a positive externality the inequality changes directions.

bound for sensible tax rates. For example, Piketty et al. [2014] states that they "focused on the revenue-maximizing top tax rate, which provides an upper bound on top tax rates". This position would need to be modified in a model with societal effects of inequality.[29]

I.G. Inverse-optimal social welfare weights

Re-arranging (1.5), we can quickly find an expression for $\bar{G}(z)$:

$$\bar{G}(z) = (1 + \Upsilon(z)) - \frac{\tau(z)}{(1 - \tau(z))}\alpha(z)\epsilon(z) \tag{A.41}$$

Using that $\bar{G}(z) = \frac{1}{1-H(z)} \int_z^\infty g(j)dH(j)$, we can multiply by $1 - H(z)$ and take derivatives to find:

$$g(z) = \frac{1}{h(z)} \frac{d}{dz} \left[(1 - H(z)) (1 + \Upsilon(z)) - \frac{\tau(z)}{(1 - \tau(z))} z h(z) \epsilon(z) \right] \tag{A.42}$$

To calculate the inverse-optimal SWWs implied by the U.S. income tax schedule shown in Figure 6 we used (A.42), taking the numerical derivative of the bracketed expression. The pre-tax income distribution and tax specification are detailed in Appendix I.E.1 and shown in Figure E1. We also assumed an elasticity of $\epsilon(z) = 0.3$ and a Gini post-tax income inequality externality. Finally we smoothed the resulting $g(z)$ to 99 quantile bins by taking the weighted mean of data inside each quantile boundary.[30]

In Figure G7a we show the implied SWWs under the same specification for a set of positive post-tax income inequality externalities.

In Figure G7b we show that a positive inequality externality could lead to the implied SWWs from the 2019 U.S. tax system being everywhere decreasing. We use the top-income inequality metric from (1.8) with $q = 24$.[31] The mix of everywhere decreasing SWWs and a belief in top-end inequality as a positive externality could conceivably describe facets of conservative U.S. politics around the latest large-scale tax schedule reform in 1986 (TRA86).

I.H. Different inequality externalities

In this section we calculate the optimal non-linear income tax rates in the presence of other types of inequality externalities, namely (i) pre-tax income inequality externalities, and (ii) utility inequality externalities.

I.H.1. Pre-tax income inequality externality

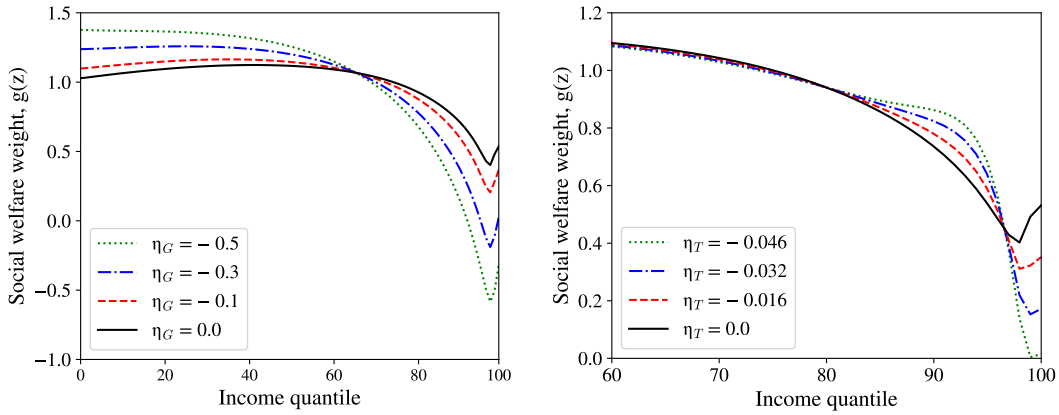
A pre-tax income inequality externality problem is a simpler version of the post-tax income inequality externality problem. We solve it here in the small perturbation framework under no income effects. The majority of the solution is similar. In terms of inequality impacts, the mechanical channel falls away while the behavioral channel becomes stronger. The revenue and

[29] While this is a relatively obvious theoretical finding, we highlight it as the Laffer curve is often discussed in public debates where participants are also likely to believe that inequality is an externality.

[30] We ignored the top 0.5% to avoid our results being affected by the assumption of a constant Pareto distribution at the top as discussed in Appendix I.E.1. This does not significantly affect results.

[31] This indicates a positive inequality externality particularly focused on top incomes. Note that values for η_T are not comparable to values for η_G. Indeed, the magnitudes of η_T have been specifically chosen to illustrate this point.

Figure G7: Implied $g(z)$ from the 2019 U.S. tax system across inequality externalities



Note: Left: Implied social welfare weights $g(z)$ from the 2019 U.S. tax system under various positive inequality externalities η_G (inequality is a social benefit). Right: Implied top social welfare weights $g(z)$ from the 2019 U.S. tax system under different positive top-share inequality externality magnitudes η_T (top inequality is a social benefit). Note that η_G is not comparable to η_T.

direct welfare portions are standard.

We introduce a small tax reform $d\tau_z$ where the marginal income tax is increased by $d\tau$ in a small band from z to $z + dz$. The reform mechanically increases average tax rates on everyone above this band. These agents do not change their work decisions or pre-tax income, so the effect of these individuals on *pre-tax* income inequality does not change.

The behavioral response is driven by agents changing their pre-tax income. The inequality impact of the behavioral responses is thus preserved, and in fact increased. Those who are located in the small band between z to $z + dz$ work less, and reduce their pre-tax income by an amount $\partial z = -\epsilon(z) z \partial\tau / (1 - \tau(z))$.[32] The behavioral response thus has an effect on the post-tax income inequality metric as $d\bar{\theta}_B = -\kappa(z) \cdot dz \partial\tau \cdot \epsilon(z) z h(z) / (1 - \tau(z))$. This differs from the post-tax inequality impact by a factor of $1 / (1 - \tau(z))$.

The total equality effect is only driven by these behavioral responses and thus $d\bar{\theta} = d\bar{\theta}_B$. In terms of utility, this affects every individual as $\int_j g_j \frac{\partial U_j}{\partial \bar{\theta}} \cdot \partial\bar{\theta} \cdot dj$. As we assume a homogeneous inequality externality and quasi-linearity in consumption such that $\eta = MRS_{x\bar{\theta}} = -\frac{\partial U / \partial \bar{\theta}}{\partial U / \partial x} = -\frac{\partial U}{\partial \bar{\theta}}$, the total welfare effect of the inequality change is $dI = \int_j g_j \cdot (-\eta) \cdot \partial\bar{\theta} \cdot dj = -\eta \cdot \partial\bar{\theta} \cdot \int_j g_j dj$.

The total welfare change, including all channels, is equal to zero at the optimum:

$$dM + dB + dW + dI = 0.$$

Thus, using the expressions for dM, dB, dW and other variables from Appendix I.D, we have:

[32] Unlike in the post-tax case, this is already in the relevant metric (pre-tax income) and therefore does not have to be multipled by $1 - \tau(z)$.

$$dz\partial\tau \int_j g_j dj \left[1 - H(z) - h(z)\epsilon(z)z\frac{\tau(z)}{1-\tau(z)}\right] - dz\partial\tau\bar{G}(z)\left(1 - H(z)\right)\int_j g_j dj$$
$$+\eta \cdot \int_j g_j dj \cdot dz\partial\tau \cdot \frac{[\kappa(z)\epsilon(z)zh(z)]}{1-\tau(z)} = 0$$

Dividing by $zh(z)\epsilon(z)\int_j g_j dj \cdot dz\partial\tau$ and re-arranging, we find:

$$\frac{\tau(z) - \eta \cdot \kappa(z)}{1-\tau(z)} = \frac{1-H(z)}{z \cdot h(z)}\frac{\left(1-\bar{G}(z)\right)}{\epsilon(z)}$$

Which implies, after substituting $\alpha(z) = zh(z)/\left(1-H(z)\right)$,

$$\tau(z)\left(1 + \frac{\left(1-\bar{G}(z)\right)}{\alpha(z)\epsilon(z)}\right) = \frac{\left(1-\bar{G}(z)\right)}{\alpha(z)\epsilon(z)} + \eta \cdot \kappa(z)$$

And finally,

$$\tau(z) = \frac{1 + \eta \cdot \kappa(z)\alpha(z)\epsilon(z) - \bar{G}(z)}{1 + \alpha(z)\epsilon(z) - \bar{G}(z)}.$$

The effect of the mechanical inequality channel on the final result has fallen away. The behavioral channel is also stronger, as it is only present in the numerator. The result cannot be approximated by SWWs, whether in utility or income.

Similarly, changing the analytical specification in Section I.C to a pre-tax inequality externality modifies (A.29) to;

$$\frac{t}{1-t} = \frac{\gamma}{\lambda}\left[\frac{\kappa(n)u_x n}{V_l}\right] + \frac{t_i}{1-t_i}. \tag{A.43}$$

I.H.2. Utility inequality externality

We solve the utility inequality externality problem here in the small perturbation framework with an additive utility inequality externality such that the inequality metric is,

$$\bar{\theta}_{\mathrm{U}}\left(\boldsymbol{z}, H\right) = \int_{\underline{U}}^{\bar{U}} \kappa_U(U(z))U(z)dH'(U(z)), \tag{A.44}$$

where $U(z)$ is total individual utility, z is total individual earnings, $H'(U)$ is the density distribution of utility, and $\kappa_U(U(z))$ is some weight in the inequality metric such that $\int_{\underline{U}}^{\bar{U}} \kappa_U(U)dH'(U) = 0$. For simplicity we will refer to a utility function of the form:

$$U(x, l, \bar{\theta}_U) = x - v(l) - \eta_U\bar{\theta}_U. \tag{A.45}$$

The majority of the solution is similar. The revenue and direct welfare effects are standard. We will now focus on the (utility) inequality impacts.

We introduce a small tax reform $d\tau_z$ where the marginal income tax is increased by $d\tau$ in a small band from z' to $z' + dz$. We note that the utility of the agents making behavioral responses

only changes on a second-order basis. We can thus focus on the mechanical effect.

For each $dz\partial\tau$ of revenue collected from those above the bracket, utility inequality changes. To explore the mechanical effect it is useful to first simplify the utility inequality term we need for this specific channel. We can first safely ignore the impact of the mechanical effect on the labor term in the utility function, as the mechanical channel is unrelated to any change in labor choice and the utility function is additive. Further, as $\int_{\underline{U}}^{\bar{U}} \kappa_U(U)dH'(U) = 0$ by assumption and any change in the inequality metric is flatly applied to everyone by the homogeneous externality assumption, we can also ignore the impact the mechanical effect has on utility through the externality term itself. We are using a quasi-linear utility function, and thus the remaining relevant part of utility is simply $x(z)$. Finally, we note that $\kappa_U(U) = \kappa(z)$ as ranks in post-tax income and utility are identical by assumption. We thus use a simplified inequality metric $\bar{\theta}_{U,mech}$ for the mechanical effect calculation,

$$\bar{\theta}_{\mathrm{U},mech}(\boldsymbol{z}, F) = \int_{\underline{z}}^{\bar{z}} \kappa(z)x(z)dH(z), \tag{A.46}$$

which is identical to the post-tax absolute income inequality metrics used in the main text.

With this simplification the derivation of the remainder of the problem becomes nearly identical to that in Appendix I.D. To summarize, the behavioral response channel does not exist in the utility inequality case and the mechanical effect channel simplifies to that of a post-tax income inequality externality. Following the solution in Appendix I.D to its conclusion (excluding the behavioral response channel) we find:

$$\tau(z) = \frac{1 + \eta_U \cdot \bar{\kappa}(z) - \bar{G}(z)}{1 + \alpha(z)\epsilon(z) + \eta_U \cdot \bar{\kappa}(z) - \bar{G}(z)}.$$

Which is identical to the standard case after removing the behavioral response terms. Note that by using the modified SWWs $\bar{G}'(z) = \bar{G}(z) - \eta_U \cdot \bar{\kappa}(z)$ this can be simplified to the no-externality case without the need for $\alpha(z)$ or $\epsilon(z)$ in the modified SWWs.

I.H.2.1. Removing quasi-linearity

Without quasi-linearity in the utility function such that $U(x, l, \bar{\theta}) = u(x) - v(l) - \eta\bar{\theta}$, the relevant inequality metric is:

$$\bar{\theta}'_{\mathrm{U},mech}(\boldsymbol{z}, F) = \int_{\underline{z}}^{\bar{z}} \kappa(z)u(x(z))dH'(U). \tag{A.47}$$

Here there are two significant effects on this absolute inequality metric from the mechanical effect. The first is the reduction of post-tax income (and thus utility) of everyone above the tax bracket. The second is the flat increase in post-tax income from the redistributed revenue.

We begin with the first of these. Each decrease in one unit of post-tax income changes absolute utility inequality by $-\kappa(z)u_x(x(z))h'(z)$ (from (A.46)). The total decrease is thus $\int_{z'}^{\bar{z}} -u_x(x(z))\kappa(z)dz\partial\tau dH(z)$. This is as far as we can go in the general case as the sum of $u_x(x(z_j))\kappa(z_j)$ above z' is not easily simplified.

The flat increase in post-tax income changes utility inequality in a similar fashion, where if total revenue gathered per agent is dR', the total effect becomes $\int_{\underline{z}}^{\bar{z}} u_x(x(z))\kappa(z)dR'dH(z)$. This is again difficult to simplify.

When assuming a quasi-linear utility function the problem simplifies, as the reduction in post-tax income above z' leads to an inequality change of $\int_{z'}^{\bar{z}} -u_x(x(z))\kappa(z)dz\partial\tau dH(z) = -\overline{\kappa}(z)\left[1-h(z)\right]dz\partial\tau$, and the flat increase in income has no effect as $\int_{\underline{z}}^{\bar{z}} u_x(x(z))\kappa(z)dR'dH(z) = dR'\int_{\underline{z}}^{\bar{z}}\kappa(z)dH(z) = 0$. We can thus write that the total utility inequality change from the perturbation is $d\bar{\theta}_U = -\overline{\kappa}(z)\left[1-H(z)\right]dz\partial\tau$, which is equal to the mechanical effect from the standard externality case.

I.I. Further externality micro-foundations

Below we show micro-foundations for three more inequality externality channels; trust, crime, and political capture.

- Trust: Assume that individuals have higher trust $t_{i,j}$ in other individuals who share a set of similar characteristics, where the set of relevant characteristics is denoted as the vector $\overrightarrow{T}$. If income x is part of $\overrightarrow{T}$, or causes changes in individual parameters that are, a change in income inequality $\bar{\theta}$ would decrease individual i's general trust levels $T_i = \sum_i t_{i,j}$. If T_i enters into individual utility $U(x_i, T_i, ...)$, income inequality has an indirect utility effect.

- Crime: Assume that criminal activity gains a fraction α of another agent's income x_j, subtracting a fixed risk cost, where agent j is randomly chosen from some high-income subset. Further assume that the opportunity cost of crime is a wage-paying job with a salary proportional to the agent's income x_i, and that agents will commit crime if it is profitable. We define the Gini coefficient as $\overline{\theta}_G = \sum_i \sum_j (x_i - x_j)$. If $\overline{\theta}_G$ increases, the relative benefit of crime also generally increases, and criminal activity increases with subsequent society-wide utility effects for both victims and perpetrators. As richer individuals are able to spend more income to protect their assets, this effect might be moderated or even overturned.[33]

- Political capture: Assume that the political process is affected by a voting procedure between discrete options $\left\{\overline{V}_1, ..., \overline{V}_m\right\}$ where each agent has a number of votes $v_i(x_i)$ corresponding to an increasing function of their income x_i. Assume further that individual utility $U_i(x_i, \overline{V}_k, ...)$ is dependent on the outcome of this political process, with varying individual preferences. Changing income inequality $\bar{\theta}$ will mechanically change voting outcomes by giving higher-income agents a larger vote share. As the vote outcome affects the individual utility of every agent – positively or negatively – inequality indirectly affects individual utility.

[33] As with all these examples, this is a very simple illustration of a complex topic with several other potential causal strains. See Kelly [2000] for a broader discussion.

I.J. Tables

Table J1: The Effects of a Small Tax Increase on Revenue R and Inequality $\bar{\theta}$

		Bottom incomes	Middle incomes	Top incomes
Behavioral response	Revenue effect	← Decreases R →		
	Inequality impact	Increases $\bar{\theta}$	Small / no change to $\bar{\theta}$	Decreases $\bar{\theta}$
Mechanical effect	Revenue effect	← Increases R →		
	Inequality impact	← Decreases $\bar{\theta}$ →		

Note: The table describes the effect each channel exerts on inequality $\bar{\theta}$ and tax revenue R through a small marginal tax increase in the specified distributional location.

Table J2

Optimal Top Tax Rates, Inequality Externalities and Distribution Parameters

		Inverse top Pareto parameter $1/\alpha$											
		0.25	0.27	0.29	0.31	0.33	0.36	0.40	0.44	0.50	0.57	0.67	0.80
	-0.50	4	7	11	14	18	22	27	32	37	42	49	55
	-0.25	36	38	40	43	45	48	51	54	58	62	66	70
	0.00	**52**	**54**	**55**	**57**	**59**	**61**	**63**	**66**	**68**	**71**	**74**	**78**
	0.25	62	63	64	66	67	69	71	73	75	77	79	82
Sensitivity	0.50	68	69	70	71	73	74	76	77	79	81	83	85
to	0.75	73	73	74	76	77	78	79	80	82	84	85	87
inequality	1.00	76	77	78	79	80	81	82	83	84	86	87	89
η	1.25	79	79	80	81	82	83	84	85	86	87	89	90
	1.50	81	81	82	83	84	84	85	86	87	88	90	91
	1.75	83	83	84	84	85	86	87	88	89	90	91	92
	2.00	84	85	85	86	86	87	88	89	89	90	91	93
	2.25	85	86	86	87	87	88	89	89	90	91	92	93
	2.50	86	87	87	88	88	89	90	90	91	92	93	94
	2.75	87	88	88	89	89	90	90	91	92	92	93	94
	3.00	88	88	89	89	90	90	91	91	92	93	94	94

Note: Top marginal tax rates from (1.9) with varying values of a homogeneous inequality externality and the inverse local Pareto parameter $1/\alpha$ at the top. These values hold for any standard SWF (welfare weights that are non-increasing and non-negative). The elasticity of labor E_L is 0.3. The inverse local Pareto parameter $1/\alpha$ is approximately 0.5 at the top in empirical data (and in the remainder of the paper). The standard no-externality case is in bold.

Table J3

Optimal Top Tax Rates, Inequality Externalities and Labor Elasticities

		Elasticity of labor E_L									
		1.00	0.90	0.80	0.70	0.60	0.50	0.40	0.30	0.20	0.10
	-0.50	0	3	6	10	14	20	27	37	50	69
	-0.25	33	35	37	40	43	47	52	58	67	79
	0.00	**50**	**51**	**53**	**55**	**57**	**60**	**64**	**68**	**75**	**85**
	0.25	60	61	62	64	66	68	71	75	80	88
Sensitivity	0.50	67	68	69	70	71	73	76	79	83	90
to	0.75	71	72	73	74	76	77	79	82	86	91
inequality	1.00	75	76	76	77	79	80	82	84	88	92
η	1.25	78	78	79	80	81	82	84	86	89	93
	1.50	80	81	81	82	83	84	85	87	90	94
	1.75	82	82	83	84	84	85	87	89	91	94
	2.00	83	84	84	85	86	87	88	89	92	95
	2.25	85	85	86	86	87	88	89	90	92	95
	2.50	86	86	87	87	88	89	90	91	93	96
	2.75	87	87	87	88	89	89	90	92	93	96
	3.00	88	88	88	89	89	90	91	92	94	96

Note: Top marginal tax rates from (1.9) with varying values of a homogeneous inequality externality and elasticity of labor E_L. These values hold for any standard SWF (welfare weights that are non-increasing and non-negative). The inverse local Pareto parameter $1/\alpha$ is 0.5 in these calculations. The elasticity of labor E_L is 0.3 in the remainder of the paper. The standard no-externality case is in bold.

Appendix II.

Appendix to Chapter Two

II.A. Prior questions about inequality externality beliefs

As far as we know there are two prior questions in the United States on individuals' beliefs about inequality's externality effects. The first is a question in the General Social Survey asking respondents if they agree with the statement that "*large income differences are necessary for America's prosperity*". We show the trend of this question in Figure 12, overlaid with the bottom 50% income share and a measure of broad economic fairness views in the share of individuals who agree that "*hard work is more important than luck*".

Among larger representative surveys, the International Social Justice Project has asked individuals whether they agree that "*There is an incentive for individual effort only if differences in income are large enough*". This question was also asked in the United States. In Table A1 we show the results from this question in their 1991 and 1996 waves [Wegener et al., 2010] across different countries.

Table A1: Inequality Externality Beliefs from the ISJP (1991-1996)

	Bulgaria	E. Ger	W. Ger	Hungary	Japan	N.lands	Poland	G.B.	U.S.	Russia	Slovenia	Cz. Rep.	Estonia	Slovakia
Does not agree	43%	38%	31%	68%	48%	47%	27%	34%	37%	38%	36%	35%	49%	38%
Agree	57%	62%	69%	32%	52%	53%	73%	66%	63%	62%	64%	65%	51%	62%
Respondents	2,628	2,063	1,787	1,837	708	2,532	1,389	1,246	1,370	2,816	1,294	1,987	1,930	345

Note. Percentage who agree with the statement that "*There is an incentive for individual effort only if differences in income are large enough*". Data from [Wegener et al., 2010]. Data was collected from Bulgaria, E. Germany, W. Germany, Hungary, the Netherlands, Russia, the Czech Republic, and Estonia in both 1991 and 1996. For these countries, data was merged across both years. Options were "Strongly agree", "Somewhat agree", "Neither agree nor disagree", "Somewhat disagree", and "Strongly disagree". The first two are merged into the column "Agree" for this table. Respondents who answered "Don't know" were excluded (on average 6% of respondents).

Outside of Europe and North America there is scarce evidence on the topic. Surveys conducted in China by Whyte [2010] is the sole exception we know of, with three questions on the topic with the following results;

- 51% of respondents agree that "*income gaps threaten stability*",
- 20% of respondents agree that "*income gaps aid national wealth*", and
- 50% of respondents agree that "*income gaps foster hard work*".

II.B. Inequality as an Externality: Theory

II.B.1. The existence of inequality externalities

Section 2 asserts two points. First, that each inequality externality channel can be caused by several different mechanisms. Second, that these channels are potentially simple and can be micro-founded with few assumptions.

To establish the first point we will use the existing literature on economic inequality's impact on crime. Following Kelly [2000], three main theories of how economic inequality increases crime can be sketched:

1. *The economic theory of crime* poses that individuals rationally optimize their resources, allocating time between market labor and criminal activity. Higher economic inequality leads to a higher relative return to crime for the majority of the population. Thus individuals substitute into criminal activity when inequality increases.

2. *The strain theory of crime* poses that individuals who struggle in more unequal economic systems are increasingly frustrated by what they see as their relative (and potentially unjust) failure compared to those around them. This causes stress, alienation, and finally leads at least some individuals to criminal activity.

3. *The social disorganization theory of crime* posits that inequality could decrease family and institutional stability, increase relative poverty, and weaken social networks. If so, there could be both more opportunities for and less risk from criminal activity – thus increasing the amount of crime.

One could also find causal channels for why economic inequality *decreases* crime; suppose that higher economic inequality leads to more segregation or stricter policing, for instance, which leads to less criminal opportunity.

The above is a very brief overview of the nuanced and varied hypotheses that underpin one potential inequality externality, that of crime. We believe this illustrates both the complexity of such theories and why it is infeasible to discuss the mechanisms behind each causal channel in detail.

The second assertion from the main text is that inequality externalities can be relatively simple to micro-found and can be mechanical in nature. Mechanical in this context means that they do not require other-regarding preferences, emotional reactions, or changes in perceived inequality to exist. We illustrate this through the example of political polarization taken from Støstad and Cowell [2021]. Begin with assuming that political opinions O_i are an increasing function of individual income x_i and no other factors (for simplicity). Political polarization is formalized as $\bar{P} = \varphi(\boldsymbol{O})$ and is an increasing function of a distributional metric φ of all opinions in the population $\boldsymbol{O}$. Polarization $\bar{P}$ affects the individual's utility function $U_i(x_i, \bar{P}, ...)$. If the income differences between people increase, the polarization of opinions mechanically increase as well. This generally increases $\bar{P}$ and thus affects $U_i(...)$. It follows that inequality causes more pronounced political polarization and thus also affects individual utility.

For further micro-foundations on inequality externalities we refer to Støstad and Cowell [2021]. Several of the inequality externalities we explore in this work, including crime, trust, political polarization, innovation, and economic growth, are micro-founded there.

II.B.2. Inequality externalities and redistribution in optimal taxation

The following is a mildly modified version of the optimal tax problem found in Støstad and Cowell [2021].

Individuals' utility function U depends on their post-tax income x_i, labor effort h_i through some function $v(.)$, and individuals' experienced state of the world $\Gamma_i(.)$. This $\Gamma_i(.)$ vector includes anything else the individual cares about, for example the level of crime, political polarization, or innovation. These factors may depend on economic inequality θ, which we assume to be a function of all post-tax incomes $\mathbf{x}$.[1]

Similarly to the main text, we assume a utility function such that,

$$U_i = x_i - v(h_i) - \sum_j \gamma_{ij}\alpha_j\theta, \tag{B.1}$$

where α_j illustrates the true dependence of externality channel j (say crime) on inequality, and γ_{ij} illustrates how individual i is affected by this channel. As the social planner is all-knowing, both of these are known. We have also introduced a disutility function v from work effort h_i. We assume no strict other-regarding preferences, which have structurally similar consequences to an inequality externality if these preferences are taken into account by the social planner.

For optimal taxation purposes, the crucial aspect is the total effect of inequality θ on utility through these externality channels. This can be captured in this model by the marginal rate of substitution between income and income inequality $\eta_i = MRS_{x_i,\theta} = -\frac{\partial U_i/\partial\theta}{\partial U_i/\partial x_i} = \sum_j \gamma_{ij}\alpha_j$, which is potentially heterogeneous between individuals.

We suppose that the inequality metric θ can be written as

$$\bar{\theta}\left(\boldsymbol{z}, H\right) = \int_{\underline{x}}^{\bar{x}} \kappa(z)x(z)dH(z), \tag{B.2}$$

where $\kappa(z)$ is the weight of the agent in the inequality metric and the cumulative distribution of all pre-tax incomes z is $H(z)$. The inequality weight $\kappa(z)$ is non-decreasing, continuous, positive near the top of the income distribution and negative near the bottom, and otherwise general. For example, the (absolute) Gini coefficient in post-tax income has a weight $\kappa_G(z) = 2H(z) - 1$.

The social planner sets an income tax $T(z)$ dependent on pre-tax incomes z such that $x_i = z_i - T(z_i)$. This is done through finding the tax schedule $T(z)$ from which no given small perturbation ϵ which changes the tax schedule as $T(z) + \epsilon\Delta T(z)$ leads to welfare improvements. We denote the resulting change in the inequality metric from the small tax increase by $\Delta\bar{\theta}$. The local optimal tax criterion is thus defined as the tax schedule $T(z)$ for which any small budget neutral tax reform in direction $\Delta T(z)$ has $\int_i g_i \left[\Delta T(z_i) + \eta_i\Delta\bar{\theta}\right] di = 0$, where g_i is the SWW of individual i.

[1] Γ could also depend on different inequality metrics θ_i, perceived inequality $\tilde{\theta}(\mathbf{x})$, the individuals' income x_i, and more – we will abstract from these factors for simplicity.

We thus introduce a small tax reform $dT(z)$ where the marginal income tax is increased by $\partial\tau$ in a small band from z to $z + dz$. The reform mechanically increases average tax rates on everyone above this band. This is the mechanical effect of taxation, and collects $dz\partial\tau$ from $1 - H(z)$ agents above z under our utility function (which has no income effects). Thus it collects $[1 - H(z)]\, dz\partial\tau$ revenue. For each $dz\partial\tau$ collected, however, inequality also changes. The magnitude of this change per agent above differs based on which agent is considered. Noting that income rank $\kappa(z)$ does not change given that second-order conditions hold, each decrease in one unit of post-tax income at z changes absolute post-tax income inequality by $\kappa(z)h(z)$. The mechanical effect thus has a differing equality effect of $-\kappa(z_j)h(z_j)dz\partial\tau$ at each point j above z, where z_j is the income of the agent and $h(z_j)$ is the number of agents at this point, and $\kappa(z_j)$ is that agent's weight in the inequality metric. As the income change of each agent above z is equal, we can define the average inequality weight above as $\overline{\kappa}(z)\,[1 - H(z)] = \int_{\{j:z_j>z\}} \kappa(z)h(z)dj$ and write that the mechanical effect changes income inequality by $d\bar{\theta}_M = -\overline{\kappa}(z)\,[1 - H(z)]\, dz\partial\tau$.[2]

Those who are located in the small band between z to $z + dz$ have a behavioral response to the tax change. They work less, and reduce their pre-tax earnings by an amount $\partial z = -\epsilon(z)z\partial\tau/\left(1 - \tau(z)\right)$. $\epsilon(z)$ is the elasticity of earnings z with respect to $1 - \tau(z)$. There are $h(z)dz$ individuals in the tax bracket who were taxed at $\tau(z)$ before the perturbation, so total revenue decreases by $-dz\partial\tau \cdot \epsilon(z)zh(z)\tau(z)/\left(1 - \tau(z)\right)$. This change in total earnings is moderated by an effect $(1 - \tau)/\tau$ for the inequality effect, as we are interested in the post-tax income decrease and not the tax revenue decrease.[3] Additionally we must multiply by the agents' weight in the inequality metric $\kappa(z)$. The behavioral response thus has an effect on the post-tax income inequality metric as $d\bar{\theta}_B = -\kappa(z) \cdot dz\partial\tau \cdot \epsilon(z)zh(z)$.

The total revenue effects are:

$$dR = dz\partial\tau \left(1 - H(z) - \epsilon(z)zh(z)\tau(z)/\left(1 - \tau(z)\right)\right)$$

The direct welfare effect through the individual income channels is $\int_j g_j dR dj$ for $z_j \leq z$ and $-\int_j g_j(\partial\tau dz - dR)dj$ for $z_j > z$. Thus the net individual income-based welfare effect is $dM + dB + dW = dR \cdot \int_j g_j dj - dz\partial\tau \int_{\{j:z_j\geq z\}} g_j dj$.

The total equality effect is $\partial\bar{\theta} = d\bar{\theta}_M + d\bar{\theta}_B$:

$$\partial\bar{\theta} = dz\partial\tau \left(-\overline{\kappa}(z)\,[1 - H(z)] - \kappa(z)\epsilon(z)zh(z)\right) \tag{B.3}$$

In terms of welfare, the effect is $\int_j g_j \frac{\partial U}{\theta}\bar{\theta}dj$. We have that $\eta_i = MRS_{x_i\bar{\theta}} = -\frac{\partial U/\partial\bar{\theta}}{\partial U/\partial x_i}$, and thus the total welfare effect of the inequality change is $dI = \int_j(-g_j\eta_j)\partial\bar{\theta} \cdot dj = -\partial\bar{\theta}\int_j \eta_j g_j dj$.

The total welfare change, including all channels, is equal to zero at the optimum:

$$dM + dB + dW + dI = 0 \tag{B.4}$$

Note that in the main text we denote $dI = dI_B + dI_M$ where dI_B and dI_M correspond to the welfare-weighted versions of $d\bar{\theta}_B$ and $d\bar{\theta}_M$ respectively. Thus, using the expressions for dR and dI, and the expression $\bar{G}(z)\,(1 - H(z)) = \int_{\{j:z_j\geq z\}} g_j dj / \int_j g_j dj$, we have:

[2] In the absolute Gini, $\overline{\kappa}(z) = H(z)$.

[3] For the mechanical effect, the tax revenue increase and the individual post-tax income decreases are identical.

$$dz\partial\tau \int_j g_j dj \left[1 - H(z) - h(z)\epsilon(z)z\frac{\tau(z)}{1-\tau(z)}\right] - dz\partial\tau\bar{G}(z)\left(1 - H(z)\right)\int_j g_j dj$$
$$+ \int_j \eta_j g_j dj \cdot \left[dz\partial\tau\left(\overline{\kappa}(z)\left[1 - H(z)\right] + \kappa(z)\epsilon(z)zh(z)\right)\right] = 0$$

Dividing by $zh(z)\epsilon(z)\int_j g_j dj \cdot dz\partial\tau$ and re-arranging, we find:

$$\frac{\tau(z)}{1-\tau(z)} = \eta \cdot \kappa(z) + \frac{1 - H(z)}{z \cdot h(z)}\frac{\left(1 - \bar{G}(z) + \eta\overline{\kappa}(z)\right)}{\epsilon(z)}, \tag{B.5}$$

where we have used the weighted average of the externality $\eta = \int_i g_i\eta_i di / \int_i g_i di$. By using the local Pareto parameter $\alpha(z) = \frac{z\cdot h(z)}{1-H(z)}$ and assuming that the social value of one dollar at the top is zero (due to the diminishing marginal utility of income) we find the optimal top marginal income tax rates as,

$$\tau(z) = \frac{1 + \eta\Omega(z)}{1 + \alpha(z)\epsilon(z) + \eta\Omega(z)}, \tag{B.6}$$

where $\epsilon(z)$ is the earnings elasticity with respect to $1-\tau(z)$, the local Pareto parameter is denoted by $\alpha(z)$, and the net inequality externality magnitude is a combination of the individual effects such that $\eta = \int_i \eta_i g_i di$.[4] Finally, $\Omega(z) = \alpha(z)\epsilon(z)\kappa(z) + \bar{\kappa}(z)$ contains the remaining effect of the inequality externality where $\kappa(z)$ denotes the weight of the individual in the inequality metric at z and $\bar{\kappa}(z)$ denotes the average of this weight above z.

The main conclusion from this exercise is simple; the total inequality externality magnitude (η) strongly affects the optimal amount of redistribution (represented here by the optimal top tax rate). The externality magnitude, in turn, depends on how inequality affects various channels (α_j) and how this in turn affects individuals (γ_{ij}).

II.C. Survey Details

II.C.1. Survey 1

The survey flow is shown in Figure 3. The survey was divided into three main parts:

1. A demographic section, asking standard questions on gender, age, party affiliation and so on.

2. A video treatment section. This begins with our pre-treatment fairness and externality questions, then sends the respondents into one of six randomly assigned groups. Four of these are shown 1-2 minute treatment videos on externality concepts (3 videos) or fairness concepts (1 video). One group is shown a video on how academics measure inequality using the top 10% income share and the Gini index (active control video). One group does not receive any stimulus (passive control group). Each video group sees "filler questions" after the video.

[4] For a more thorough explanation of these terms see Appendix II.B.2.

3. A section including additional demographic questions, redistributive question questions (main treatment outcomes), and further descriptive questions (including first-stage outcomes).

The full Survey 1 questionnaire can be found here.

II.C.1.1. Data quality and attrition

11,540 respondents land on our consent page. 10,992 respondents consented to taking the survey, and 8,551 pass the required U.S. citizen screening test. Of these, a total of 5,007 respondents (58.6% of initial U.S. respondents) finished the survey. Simple data quality checks remove 7.64% of these respondents from the sample during the demographic section.[5]

Further, 33.9% of respondents end the survey on their own before completion; we will now briefly discuss these respondents. 7.9% of respondents drop out before the information treatment. 17.9% of respondents drop out during the information treatment.[6] While the passive control group had less attrition at this stage – as there is no video treatment and no required questions for these respondents, see Section 5.1.1 – the differences in attrition between the active control and the remaining treatments are small.[7] The remaining 8.2% who dropped out did so after the treatment.

This leaves the 5,007 respondents who finished the survey. We pre-specified to drop the fastest 5% of the fastest respondents at this stage, as is often done in the literature and by survey companies [see e.g. Bellani et al., 2021].[8] We also exclude subjects that spend less time on the screen with the video treatment than the duration of the video, as well as those who were in treatment groups that nonetheless self-reported that they did not watch a video. We also added two extra data quality checks that were not pre-specified. First, we deleted 237 respondents that dropped out of the survey in the middle and then retook the survey, which we identify due to identical IP-addresses. Second, we drop 109 subjects that were flagged due to providing "nonsense" answers to text-based questions (e.g. spam, vulgar phrases or the same non-topical copy-pasted text to all answers). Our results are not fragile to these two steps, which were taken to improve overall data quality.

[5]Respondents are required to pass two of three simple attention checks to continue past the demographic section. All attention check removals happened before any topic-specific questions. These attention checks are very simple and designed to sieve out individuals that do not read the questions at all. We believe relatively rigorous but simple attention check requirements are necessary to optimize the signal from online panel surveys, which are prone to inattentive and non-human respondents. Individuals who failed one attention check but still finished the survey add weight to this argument. These respondents exhibit generally lower correlations across similar questions (e.g. similar externality questions or similar fairness questions) than respondents who did not fail any attention checks. Still, our overall results are similar whether including or excluding these individuals from the final sample. We discuss this for the treatment effect in II.F.5 and show it for the main descriptive results in Figures H3-H4.

[6]This attrition during the treatment is most likely due to either (i) technical issues from the Youtube video, for example accidentally exiting the survey screen, or (ii) inability or unwillingness to correctly answer the simple factual questions that we require respondents to correctly answer before continuing. We are unable to disentangle these two effects.

[7]Individuals who dropped out during the treatment are not significantly more likely to be Democratically leaning or have prior externality beliefs across video groups, for example. Tables I4-I10 shows that demographic controls and pre-treatment questions are balanced across treatment groups, further indicating that any selective attrition is limited. We also control for pre-treatment beliefs and demographic controls in our main treatment results.

[8]Since different treatment groups watch different videos, we drop the 5% fastest subjects within each treatment group.

Overall, this leaves a final sample of 4,371 respondents.

II.C.2. Survey 2

The full Survey 2 questionnaire can be found here.

We conducted Survey 2 as a secondary "robustness" survey with Dynata to ensure the validity of our original results. The survey was conducted between August 7th and October 8th 2022. The main structure of the Survey 2 was a simple questionnaire, where towards the end of the survey respondents were funnelled into one of eight channels on a specific inequality externality.

The first part of the survey asked similar externality-based questions as to those in Survey 1 changed in various ways to explore the robustness of our initial results. We explain the concept of inequality in-depth to respondents, for example, and substitute any mention of "inequality" for "equality" or "differences in income and wealth" for two-fifths of respondents (one-fifth for each). We also asked respondents a simple question to gauge their understanding of inequality itself and explicitly specify our definition of "more inequality" (i.e. we set a reference level of inequality for the externality questions) for the entire survey.

In the latter part of the survey each individual was funneled into a channel focusing on one specific inequality externality. These externality-specific questions included re-asking a specific externality question to check for consistency, allowing individuals more time to ensure a high-quality answer, asking respondents to explain their answer with an open-ended text question, asking whether top- or bottom-based inequalities matter more for the externality, several questions designed to find out whether the reference level changes the direction of the individuals' externality beliefs, and a question which explores whether average income or income inequality is deemed a larger predictor of the outcome in question. The eight externality channels we elicit for in this study are crime, trust, economic growth, innovation, political polarization, corruption, the quality of democratic institutions, and social unrest.

II.C.2.1. Data quality and attrition

The sampling methodology in Survey 2 is similar to that of Survey 1. We use a similar attention check procedure to ensure high-quality responses. Respondents who fail either the U.S. citizen screening at the beginning of the survey or at least two later attention checks were removed from the survey. As before, these attention checks are very simple and designed to sieve out individuals that do not read the questions at all. Unlike Survey 1, we included two attention checks in the middle of the descriptive data collection to further ensure data quality.[9]

6,980 respondents landed on our survey consent page. A total of 6,471 respondents consented to taking the survey, and 5,474 move past the U.S. citizen screening. A total of 2,479 respondents (45.3% of initial U.S. respondents) finish the survey. The data quality checks remove 30.9% of respondents from the sample. A further 23.9% of respondents end the survey on their own before completion. These numbers are likely inflated due to the repetitive nature of the survey. After

[9]These questions were in the same format as other externality questions and asked respondents to answer a specific choice option. We added these questions to ensure that our descriptive statistics had the least amount of noise possible. The question text of the first question;

This question is about the same increase in inequality. Here we just want you to click the answer option at the top. In other words, how do you think more inequality – could you please click the first answer option? Note: Here we just want you to choose the top option to show that you are reading the questions. Thank you.

dropping the fastest 5% of respondents per survey arm the final sample is 2,360 respondents.

II.C.3. Representativity

Table C1 displays the observable characteristics of our two samples. Both surveys explicitly targeted representative quotas for gender, age, political affiliation, and geographical region.

To elicit political preferences, we used the same question that is used by Gallup to monitor political preferences in America.[10] All three final distributions mirror the November 2021 Gallup poll quite closely (31% Republican, 27% Democrat, 41% Independent), each marginally undersampling Independents.[11] Figures C1-C2 show that political affiliation is representative across the 50 U.S. states, although this was not explicitly targeted. On the other dimensions we targeted, Survey 1 is completely balanced on gender and census region, and matches the age-group distribution of the overall population well. Survey 2 slightly oversamples men, older individuals, and those who live in the Census region *West*.[12]

Though we did not explicitly target these dimensions, we are also interested in having diverse socio-economic representation. We have significant variation in household incomes, and particularly Survey 2 approximates the U.S. income distribution well. Our surveys are less representative on racial dimensions, as they oversample white Americans. Hispanics and Latinos are particularly underrepresented in our study (16.8% in the overall population versus 7.0% in our pooled sample). Similar to other studies using similar access-panels, our samples are more educated than the average American, as roughly half of the respondents have at least a college degree versus 36% in the overall population.

Note that the oversampling of college-educated individuals could affect our results, as such respondents are more likely to hold a negative externality viewpoint than non-college graduates. The net effect is relatively small, however. On average, a college-educated respondent is ~ 5 p.p. more likely to hold negative externality beliefs. Our merged sample has 19 p.p. more college-graduates than a fully representative sample; taken at face value, the net effect of this oversampling is a roughly 1 p.p. increase in the share of negative externality beliefs in our data.

II.C.4. Eliciting externality beliefs

We elicit externality beliefs using various methods. Questions on individuals' *general* and *specific* inequality externality beliefs are asked in both Survey 1 and Survey 2. These beliefs are explored in closed-form multiple choice questions and open-ended text questions. Most questions are asked to all or a majority of respondents within a survey. The main exception is in Survey 2, which also funnels respondents into eight different survey strands. Each survey strand poses detailed questions on one specific externality channel.

[10] "In politics, as of today, do you consider yourself a Republican, a Democrat or an independent?"

[11] This poll was the most recent poll when the first survey was conducted. Note that there is significant fluctuation in this distribution on a month-to-month basis (c.f. https://news.gallup.com/poll/15370/party-affiliation.aspx). The year-long average is 27% Republican, 30% Democrat, and 43% Independent.

[12] These discrepancies as well as the undersampling of non-white respondents in Survey 2 come from a technical quota error on the survey providers' part. In short, the survey provider accidentally increased the sample size from 1,700 to 2,360 but did not keep quotas in mind for these extra respondents. This made the total sample from Survey 2 somewhat less representative, as the additional respondents were not subject to the designed quotas. We decided to keep the larger sample as it is balanced on political affiliation, and the over-sampled observables (male and white respondents, specifically) generally do not have large effects on our outcome variables. Our results are robust to re-weighting for full representativity.

Table C1: Survey demographics compared to the 2021 U.S. adult population

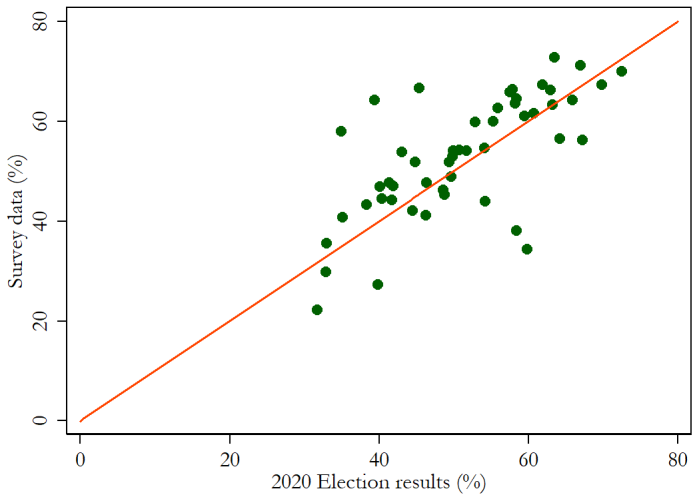
	2021 U.S. share	Survey 1	Survey 2	Merged sample
Republican	31%	32%	33%	32%
Democrat	27%	30%	28%	29%
Independent	41%	39%	39%	39%
Male	49%	50%	54%	53%
Female	51%	50%	45%	46%
White	64%	77%	75%	76%
Black	12%	9%	8%	8%
Neither black nor white	24%	14%	17%	16%
Income: 0-25k	18%	22%	18%	19%
Income: 25-50k	20%	29%	23%	25%
Income: 50-100k	29%	30%	33%	32%
Income: 100k and more	33%	19%	26%	25%
Age 18-29	18%	14%	11%	12%
Age 30-39	17%	17%	16%	17%
Age 40-49	16%	17%	17%	17%
Age 50-59	16%	14%	16%	16%
Age 60-69	17%	17%	22%	21%
Age 70 and above	17%	21%	17%	19%
4-year college degree or more	36%	50%	58%	55%
Employed	59%	47%	51%	50%
Unemployed	4%	9%	7%	8%
Outside the labor force	38%	43%	42%	43%
South	38%	38%	30%	32%
West	24%	24%	32%	30%
North-East	17%	16%	16%	16%
Midwest	21%	21%	22%	22%
Respondents		4371	2360	3922

Note. This table represents respondent demographics of Survey 1 (N= 4,371), Survey 2 (N=2,360), and the merged descriptive sample (the control group of Survey 1 and all of Survey 2) compared to the share among 2021 U.S. adults for the respective characteristic. Data on the U.S. population is from the U.S. Census Bureau, the U.S. Bureau of Labor Statistics, Gallup,

In eliciting externality beliefs we took extensive measures to avoid biases arising from respondents misunderstanding the question, anchoring, or phrasing. In general, questions are always designed to be unbiased and symmetric around a neutral answer option. The order of multiple choice answers was randomly flipped on the question level to avoid anchoring bias whenever possible. We also varied the phrasing respondents face on a question-by-question basis in Survey 1 and throughout the survey in Survey 2. The standard phrasing in both surveys uses the word "inequality". In Survey 1, respondents instead saw the phrasing "differences in income and wealth" in one-third of the specific externality questions. In Survey 2, 20% of subjects received identical questions but with "inequality" changed to "differences in income and wealth." An additional 20% of subjects had "inequality" changed to "equality".[13] Survey 2 also rigorously defines the chosen distributional concept (e.g. inequality) to respondents, including a small quiz

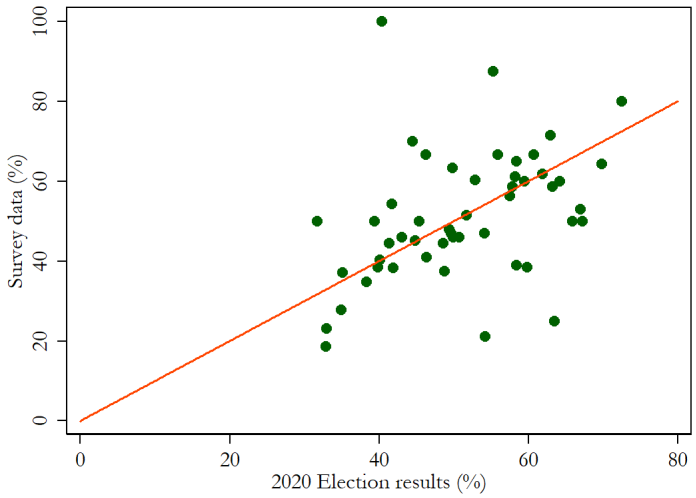
[13]Many of our questions ask how "more inequality" changes society; these respondents are instead asked how "more equality" changes society.

Figure C1: Relationship between state-level political leaning in the survey and 2020 state-level election outcomes in Survey 1



Note. This figure plots state-level shares of respondents stating that they lean towards the Republican party in Survey 1 against the state-level share of votes going to the Republican party. Washington D.C. is included (the left-most data point). The diagonal line characterizes the points where both would coincide. 43 of 50 data points from the survey are within 1.96 standard errors to the 2020 election result. In making this comparison, note that we collect responses two years after the election and do not screen on likely voters.

Figure C2: Relationship between state-level political leaning in the survey and 2020 state-level election outcomes in Survey 2



Note. This figure plots state-level shares of respondents stating that they lean towards the Republican party in Survey 2 against the state-level share of votes going to the Republican party. Washington D.C. is included (the left-most data point). The diagonal line characterizes the points where both would coincide. 47 of 50 data points from the survey are within 1.96 standard errors to the 2020 election result. In making this comparison, note that we collect responses two years after the election and do not screen on likely voters. Note also that the sample size is significantly lower in this survey ($N = 1873$), and that some data points have very few associated respondents. Delaware, the data point at 100% in survey share, has only two respondents.

which allows us to check respondent comprehension of distributional concepts.[14] Survey 2 also ensured that respondents answer accurately by asking them to confirm their previous choice;

[14]The share of individuals who believe in a given negative externality is often somewhat stronger ($5 - 10\%$) among those who answer this quiz correctly.

almost all respondents (> 97%) confirm their choice.[15] We discuss these and other robustness checks in Section 4.3 and Appendix II.D.5.

II.C.5. Designing the general externality question

We ask somewhat different questions to elicit general externality beliefs in each survey.

Due to the complex nature of the views we wish to elicit, the two questions of *does inequality affect society* and *is this effect positive or negative* were asked separately in both surveys. The order between these two questions was swapped between surveys to minimize any potential effect of noise or phrasing. In Survey 1, respondents were first asked to choose between 5 options, ranging from "A lot to the better" over "Neither / no change" to "A lot for the worse." If subjects chose "Neither / no change", they were asked a follow-up question to find out whether they chose this option because they believe that inequality has "no effect on society" or because they think the "good and bad effects cancel each other out".[16] In Survey 2, this order was swapped; respondents were immediately asked a "Yes"/"No" question about whether inequality affected society. If subjects chose "Yes, economic inequality affects society", they were then asked whether the changes would be overall positive or negative.

The way these questions were phrased also differed across surveys. The main Survey 1 question is short, designed to be easily understood, and reads "Generally speaking, do you think **more economic inequality** changes society **for the better** or **for the worse**?". This brevity comes at the cost of imprecision; in Survey 2, the question is much longer and clarifies any unclear points explicitly (see Appendix II.C.5.1 for the question in verbatim).[17]

We show the separate results for each survey in Table C2.

II.C.5.1. Survey 2 General externality belief question (Part 1)

Note: This question comes directly after a question which introduces the distributional concept (inequality, equality, etc.) and shows the income distributions shown below. Randomized phrasing is shown in brackets.

This question is about how **you think economic [inequality/equality/differences] changes society.**

Below we are showing you the same two income distributions as earlier. The correct answer was that society (B) [is more unequal/ is more equal[18]/has more economic differences].

[Insert figure]

Here's some more information: **Society A** has a large middle class and few with relatively small or large incomes. The richest tenth of society earns 5 times as much as the poorest tenth

[15] We ask subjects to confirm both their general and specific externality beliefs. These questions allow respondents to state that they either agree with their choice, that they disagree with it and want to change it, or that they answered randomly.

[16] They were also able to answer "I don't know"; the 12 respondents who answered this were pooled with the "Good

[17] Specifically (i) the initial inequality level, shown through diagrams and words to be roughly at a Scandinavian level (without explicitly naming countries), (ii) the level of the change in inequality, which is an inequality increase to roughly the level of the United States, (iii) the exogenous nature of the inequality shock, (iv) explicitly noting that we are interested in changes in factors of society, using examples such as crime and economic growth, (v) explicitly noting that the question is not about individual income or fairness concerns.

[18] Order of the distributions is switched for the equality-phrasing

Table C2: General Externality Beliefs: How Does More Economic Inequality Change Society?

	Survey 1	Survey 2
A lot for the better	4.1%	3.1%
Somewhat for the better	10.9%	7.6%
Good and bad effects cancel	20.8%	10.3%
Somewhat for the worse	34.8%	36.7%
A lot for the worse	24.8%	26.4%
Inequality does not affect society	4.5%	15.9%
Total respondents	919	2360

Note. Survey 1 question text: *"Generally speaking, do you think **more economic inequality** changes society **for the better** or **for the worse**?" For Survey 1, only data from the control group is shown. For Survey 2, the table contains data from all respondents (i.e. "inequality" phrasing, "differences" phrasing, and "equality" phrasing). See Appendix II.C.5.1 for the full question text.*

of society.

Society B has a small middle class and many with relatively small or large incomes. The richest tenth of society earns 30 times as much as the poorest tenth of society.

There is a low amount of extreme poverty in both countries.

Now imagine that the income distribution in a society moves from (A) to (B). In other words, the society becomes [**more economically unequal / more economically equal / has larger economic differences**]. The change is because of something outside the society, such as technological change in another country.

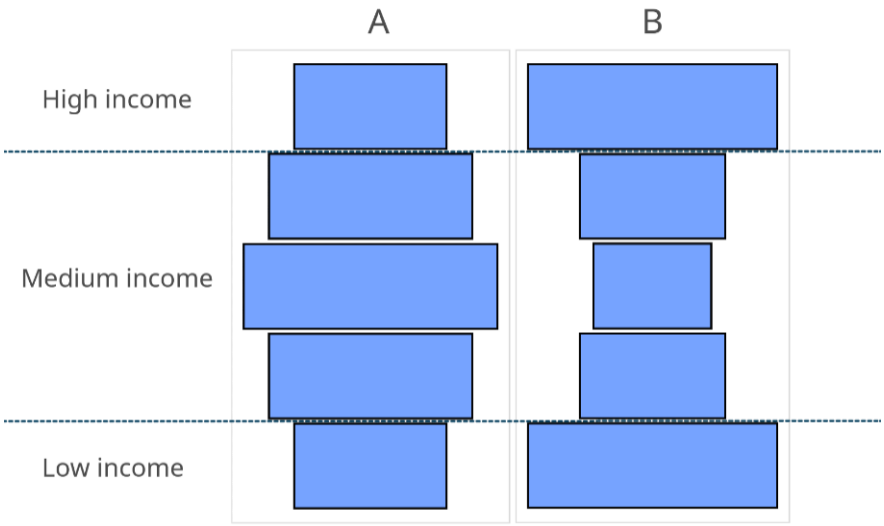
One could imagine that this either **changes** or **does not change factors in society** - such as economic growth, crime, general trust, innovation, the quality of democratic institutions, and so on. Note that this question is not about whether you think the new distribution is more or less unfair, or about the direct changes in individuals' economic situation, but about **potential changes in how the society functions as a result of [increased economic inequality / more economic equality / increased economic differences]** .

We are interested in whether you think **any** such changes occur (whether they are positive or negative).

All in all, do you think society would function differently **at all** after [becoming more economically unequal/becoming more economically equal/such an increase in economic differences within the population]?

- Yes, economic [inequality/equality/differences] affects society. The society would change
- No, economic [inequality/equality/differences] does not affect society. The society would remain the same.

Survey 2 General externality belief question (Part 2) *Note: Randomized phrasing is shown in brackets. This question was only shown to respondents that clicked "Yes" in the pre-*



vious question.

This question is about the same increase in economic [inequality / equality / differences within society] (the transition from society A to B).

All in all, do you think that the changes in society as a result of such an increase in economic [inequality / equality / differences within society] would be positive or negative?

(When thinking about your answer, try to ignore the direct effects on individuals' economic situation and focus on changes to society as a whole. Also note that this question is not about whether new distribution is more or less unfair. If you do not believe economic [inequality / equality / differences within society] affects society, select the answer option in the middle here and in subsequent questions.)

- [More economic inequality / More economic equality / Larger differences in income and wealth] → Society functions **much better**
- [More economic inequality / More economic equality / Larger differences in income and wealth] → Society functions **somewhat better**
- [More economic inequality / More economic equality / Larger differences in income and wealth] → Society functions **as well as before**
- [More economic inequality / More economic equality / Larger differences in income and wealth] → Society functions **somewhat worse**
- [More economic inequality / More economic equality / Larger differences in income and wealth] → Society functions **much worse**

II.D. Further Descriptive Results

II.D.1. Size of specific externality channels

In this subsection we discuss the relative and absolute magnitudes of each externality channel. We note, however, that eliciting the absolute magnitude of each externality channel is chal-

lenging; a precise exploration would require respondents to understand and relate changes in outcomes to changes in inequality metrics. As such, the exploration will remain relatively broad.

Which externality channels matter the most? To examine the impact of each channel we ask respondents to delegate 100 points to the externality channels that "matter the most", separating respondents who believe in an overall positive or negative inequality externality.[19] The average responses to this question are shown in Figure H15.

Respondents indicate that crime and corruption are the most important negative externalities, and that economic growth and innovation are the most important positive externalities. We note that answers are to the negative and positive externality versions are clearly different. Most notably, the economic factors are deemed the least important negative externalities and most important positive externalities.

Are externality channels meaningful? In Survey 2 we also ask respondents whether they deem a specific inequality externality "meaningful".[20] Results here are qualitatively similar. The three most meaningful channels are considered to be the quality of democratic institutions (70%), crime (67%), trust (66%), and corruption (65%).[21] Answers also indicate that respondents believe these issues are important; a majority of respondents believe the given externality is "generally meaningful" (30%) or "very meaningful" (32%).

II.D.2. Top- or bottom-inequalities

Inequality externalities could depend heavily on the *type* of inequality. Specifically, inequalities near the bottom – the amount of relative poverty[22] – and inequalities near the top could affect different channels differently. To explore this we ask respondents in Survey 2 what type of inequality matters more for any given externality.[23]

The answers are shown in Table D1. Generally respondents believe both top- and bottom-inequalities matter. However, bottom-inequalities are generally considered more important than top-inequalities, particularly for trust and crime. The only exception is corruption, where top inequalities are deemed most impactful. Finally, in Appendix II.D.4 we discuss the results from a question allowing individuals to predict the amount of various outcomes (crime, trust, and so on) given the level of average income and economic inequality.

[19] Full question: "*When thinking about how inequality [negatively / positively] affects society, which dimensions do you think matter the most, generally speaking? Please indicate what dimensions you think matter the most by giving scores below that add up to 100*". This question only makes sense if the respondent thinks inequality has at least one negative or positive externality, and we only ask the negative or positive externality version to those who answered that inequality generally affects society negatively or positively, respectively, in the general externality question shown in Table C2. We also allow respondents to self-select out of the question by stating that changed their mind. There is thus a selection effect; each group should not be seen as a representative sample, but instead as the subsection of respondents who believe inequality affects society negatively or positively, respectively.

[20] In the sense that it is *"something politicians and policy-makers should be focused on, or [...] ultimately not very important"*.

[21] Note that the quality of democratic institutions is not included in Figure H15 due to a coding error.

[22] We note that relative poverty, unlike absolute poverty, is an inequality metric.

[23] Example question text: *What do you think matters more for how economic inequality changes the amount of social unrest? · Economic differences near the bottom, meaning how many relatively poor people there are and how little they have, or · Economic differences near the top, meaning how many relatively rich people there are and how much they have.*

Table D1: What Matters: Inequalities Near the Top or Bottom?

	Pol. polar.	Crime	Corr-uption	Inno-vation	Social unrest	Econ. growth	Trust	Dem. inst.
Both	48%	40%	41%	39%	36%	36%	36%	40%
Bottom inequality	33%	49%	22%	38%	51%	38%	43%	36%
Top inequality	11%	7%	29%	15%	9%	14%	9%	18%
Don't know	8%	4%	8%	8%	4%	12%	11%	6%
Sample size	247	251	228	212	249	226	226	214

Note. The share of respondents who think economic differences near the bottom or top matters more for one randomly chosen inequality externality; data is from Survey 2.

II.D.3. Consistent inequality externality beliefs across inequality levels

A key assumption for our main results to be easily interpretable is that the direction of individuals' inequality externality beliefs do not vary across the level of inequality. If they are not, a respondent might for example think that more economic inequality increases the amount of economic growth if inequality is low and decreases the amount of economic growth if inequality is high. To explore this we ask Survey 2 respondents directly whether they think the same relationship holds "*no matter whether the country is initially very equal, very unequal, or anything else*". The large majority (81%) expresses unchanging externality beliefs. The externalities with the largest share of changing beliefs are innovation (25%) and economic growth (24%). The full data is shown in Table I16.[24]

II.D.4. Inequality or average income?

A pertinent question is whether respondents think the average income or level of economic inequality is more impactful in determining the levels of the outcomes we elicit. It is difficult to create an easily understood question on this topic; such a question would also be prone to experimenter demand. We thus examine this topic from an indirect approach. To do so we ask respondents to predict the level of a given outcome in an average country with a [low/high] level of average income and a [low/high] level of economic inequality. We then analyze how the changes in given average income/inequality changes the predicted outcome.

We show the results in Figure H19. Both the average income and level of inequality are generally strong predictors for the outcomes we elicit (crime, trust, and so on). Economic inequality is particularly predictive for the level of political polarization and corruption, whereas the average income is particularly predictive for economic factors. Indeed, high economic inequality is on average *positively* correlated to high economic growth and innovation, in opposition to our prior results. Meanwhile, respondents do not take the level of average income into account when predicting the level of political polarization.

In sum, respondents believe that both the level of average income and the level of economic inequality are strong and distinct predictors for other outcomes. Note, however, that this analysis

[24] Among respondents who have changing beliefs in innovation and growth, follow-up questions indicate that they believe in a positive externality at low inequality levels and a negative externality at high inequality levels.

does not explore respondents' *causal* beliefs and is a purely correlational exercise. For example, respondents believe high-income countries have high economic growth and not necessarily that high incomes *cause* high economic growth. As such we suggest interpreting Figure H19 with caution.

II.D.5. Phrasing results

The robustness check that has the largest effect on the descriptive statistics is to substitute out the phrase "more inequality" with "more equality". Under this phrasing the share of negative specific externality beliefs are on average $\sim 12\%$ lower.[25] This does not seem to be due to the word "inequality" itself, as using "differences in income and wealth" has a much smaller effect, but rather due to the difference in the distributional change (the effect of "more equality" vs. "more inequality").

This section discusses this change. We modify the word "inequality" in the survey for various subsets of respondents to explore whether the word itself (and its potentially loaded nature) affects results. Instead we use either "equality" or "differences in income and wealth" throughout the survey for 20% each of Survey 2 respondents, and "differences in income and wealth" for one-third of questions in Survey 1 on a question-by-question basis.

In Survey 1, one-third of respondents per question saw the phrasing "*How do larger differences in income and wealth within the population...*" instead of "*How does more economic inequality...*". This phrasing was randomly assigned on a question-by-question basis with the goal of exploring whether the phrasing of the question significantly impacted answers.

We further explored this topic in Survey 2. There 20% of respondents were shown an "equality" phrasing and 20% were shown a "differences in income and wealth" phrasing throughout. Respondents in the "differences in income and wealth" phrasing strand, for example, do not see the word "inequality" anywhere in the survey. Respondents were explained each concept using the same diagrams.

Note that respondents who received the "equality" phrasing were asked how "more equality" changes the relevant factors, which changes the *direction* of the question. As an example, a negative externality belief under the "inequality" phrasing would be "*More inequality* $\rightarrow$ *More crime*". The same belief under the "equality" phrasing would be "*More equality* $\rightarrow$ *Less crime*".

General externality beliefs Neither the "differences" nor the "equality" phrasing had a significant effect on general externality beliefs (statistically insignificant $>$ 2 p.p. changes).

Specific externality beliefs Specific externality beliefs are generally not constant across phrasing choices. We show this in Figures H3-H4 and detail the results below.

First, the "differences" phrasing. This phrasing choice has a small but non-negligible effect on results in Survey 1 (where it was used on a question-by-question basis). In most questions it shifts averages by roughly 2-4 percentage points. The largest phrasing effect is for economic growth, where about 8% of individuals shift their response away from inequality decreasing growth under the "larger differences" phrasing (55% to 47%). In Survey 2, where the phrasing change was

[25] The share believing in negative inequality externalities of political polarization and corruption are particularly impacted. These shares decrease from 71% to 44% (political polarization) and 69% to 47% (corruption) respectively.

employed throughout the survey, changes are similar or smaller. No specific externality belief average shifted more than 5% from the baseline under this phrasing in Survey 2.

The "more equality" phrasing has a larger effect. It particularly affects the inequality externality beliefs regarding political polarization and corruption, where the proportion of those believing in the negative externality change from 70% to 44% and 68% to 47% respectively when changing "more inequality" to "more equality" (a decrease of 26% and 21%).[26] Despite this, the negative externality belief is still held by close to a majority in both cases. Other shifts are generally smaller and always below 15%. Respondents are less likely to choose the negative externality option in this setting for six out of eight outcomes (with a small effect in the opposite direction for the two economic outcomes).

Although phrasing choices have a significant effect on our results, the negative externality option is still the most popular for any combination of phrasing choice and outcome. It follows that our main results are robust to these changes. This exercise also implies that our results are not caused by the nature of the word "inequality" itself, as the "differences in income and wealth" phrasing do not change results in a noteworthy way. We hypothesize that the larger effect of the "equality" phrasing could be at least partly due to respondents thinking these problems are persistent and that an increase in economic equality – in other words, a reduction in economic inequality – is unlikely to solve them immediately or at all.

II.D.6. Survey bias

At the end of the survey, respondents were asked whether they considered the survey biased in an either left-wing or right-wing fashion. The large majority of respondents (72.0%) did not think the survey was biased in either direction. More respondents answered that the survey was left-wing biased (21.5%) than right-wing biased (6.5%).

The percentage of respondents who believe the survey was left-wing biased is lower in the control groups (19.1%) than in the treatments (22.3%), but there is no statistically significant difference over treatment groups. All treatment groups are between 21% and 23%. This is shown in Table I52. All main treatment effects are robust to including a dummy for left-wing bias as a control. The corresponding statistics in the Survey 2 is 16.6% left-wing biased, 5.4% right-wing biased, and 78.1% unbiased.

II.E. Further Details: Information treatment

To study the causal effect of inequality externality beliefs on redistributive preferences an information provision experiment was integrated into Survey 1. The survey was divided into three parts; the structure is shown in Figure 3. In Part 1, we elicit sociodemographic information that is needed to check for representativeness.[27,28] Part 2 presents subjects the intended information. Part 3 elicits respondents' preferences for redistribution and inequality externality beliefs, which

[26] Believing in the negative externality in this case implies that respondents answer that more equality leads to less political polarization or less corruption.

[27] This information is also used to check for selective attrition across the treatment groups, which we do not find any significant evidence of and discuss in Section II.C.

[28] We also elicit respondents' trust in the federal government and beliefs about whether people work less when taxed more. These latter attitudes have shown to be important drivers of redistributive policy preferences that are independent of fairness concerns or externality beliefs. For that reason, we elicit these views *before* the information intervention.

constitute our main reduced-form and first-stage outcomes respectively.

II.E.1. Information treatment

The information intervention in Part 2 is our main treatment variation. All subjects are first asked to answer two questions about (i) their general inequality externality beliefs, and (ii) their broad fairness beliefs.[29] These questions have two functions. First, to measure pre-treatment first-stage outcomes (when they should be equal across groups). Second, as a lead-in to the video information treatment.[30]

After these questions the sample is split into four treatment groups and two control groups. Videos are shown to the subjects in all four treatment groups and in the active control group. Each treatment video aims to shift *either* respondents' inequality externality beliefs *or* economic fairness views. We will discuss these in further detail shortly; for now it is sufficient to note that there are three inequality externality treatment groups and one fairness treatment group. The two control groups (one "passive" with no stimuli, one "active" with a control video) are as large as each of the four treatment groups when combined, which they eventually are on pre-specified criteria. We detail this *dual control group approach* further below.

II.E.2. Videos

This section describes the video content in each of the treatment arms and control groups. Note that after watching the video, respondents answer three very simple control questions to ensure that they understood the information provided in the video. We require respondents to answer these questions correctly to proceed with the survey.[31] They are then showed the filler questions.

Treatment group 1: Crime as an inequality externality This treatment group receives information on the relationship between crime and inequality using data from the World Bank and the World Inequality Database. As shown in the screenshot in Figure E1, the video first presents subjects with a scatter plot and a fitted line that characterizes the relationship between inequality and homicides. The next graphic characterizes the magnitude of the correlation. It shows that very equal countries have, on average, between one and two homicides per year per 100,000 people, while very unequal countries have, on average, between ten and twenty homicides per year per 100,000 people. The respondents are then told that researchers still argue about whether this means that inequality *causes* more crime – highlighting that these correlations need not imply causation. The video ends with a statement that most research on this topic has confirmed the correlational relationship and finds that it holds for alternative metrics of crime such as property crime and robberies.

The filler questions ask the respondents about whether they experienced or perceived more crime in places they lived or travelled to with higher levels of inequality. It thus creates a direct link to the video by asking the subjects whether they themselves experienced this relationship.

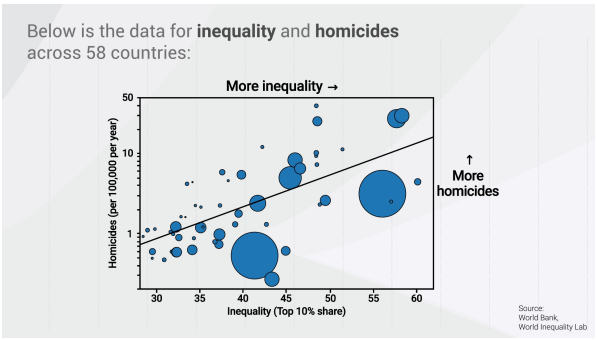
[29] These questions are "*How much do you agree with the following statement? Working-class Americans are generally paid less than their productivity.*" and "*How much do you agree with the following statement? Countries with more economic equality usually function worse.*"

[30] Subjects in treatment groups are introduced to the video with the following prompt: *"We will now show you some information regarding the last question you answered. Please watch the video below."*

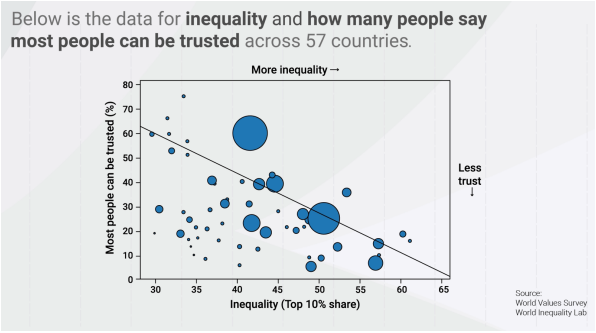
[31] Respondents who answer incorrectly are able to change their answers after being presented the video screen again (with the option to re-watch).

Figure E1: Treatment Video Screenshots

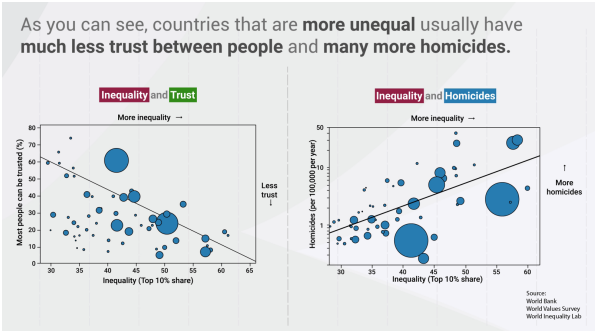
Crime



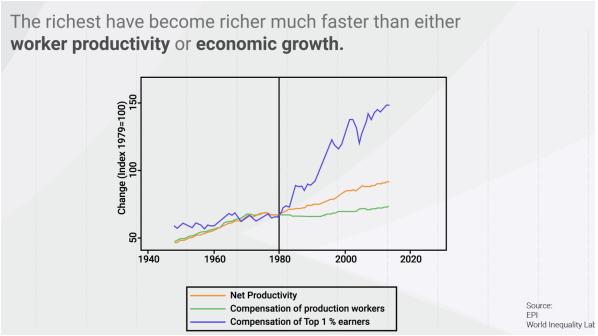
Trust



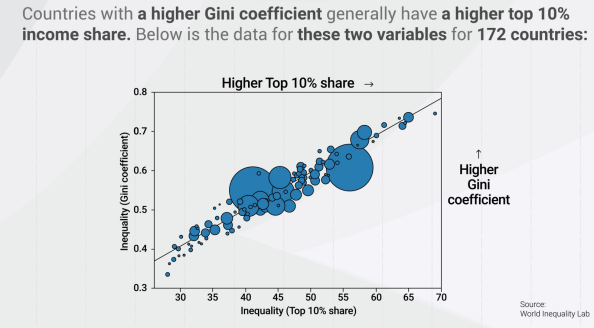
Full Externality



Fairness



Active control



Note. These are screenshots from the five videos used in the survey experiment. One video was shown to each respondent, except for the 10% of respondents in the passive control group. Click the following links for the full videos: Crime – Trust – Full externality – Fairness – Active control

Treatment group 2: Trust as an inequality externality This treatment is structurally similar and uses a correlation between inequality and generalized trust (the number of individuals that say that most people can be trusted in their country) using data from the World Inequality Database and the World Value Survey. The remainder of the video and the filler questions are intentionally similar to the crime video; the style and phrasing remains the same (with modified numbers and alternative metrics) at all times.

Treatment group 3: Full externality treatment While treatment groups 1 and 2 tackle one externality channel each, treatment group 3 is designed as an all-encompassing externality treatment. It thus aims at providing more comprehensive information on whether societies with high economic inequality usually function better or worse. By presenting broad evidence that highlights the *negative* effects of inequality and by showing that the evidence for positive externalities is rather limited, the treatment makes the strongest case for the negative consequences

of inequality between our three videos. As shown in the screenshot in Figure E1, the first part of the video shows the same information that we present in treatments 1 and 2. It then shows that there is no relationship between inequality and economic growth nor between inequality and innovation (measured by the number of patents).[32] Respondents are then told that these correlations need not imply any causal relationship, and that researchers disagree on the topic – some do not believe inequality causes society to function worse, while others believe economic inequality harms society through these and other channels (the video briefly mentions social unrest, corruption, and political polarization). The video ends with a quote from Amartya Sen, quoted as a nameless Nobel-winning economist, that "*virtually all the problems in the world come from inequality of one kind or another.*" Following the video, the filler questions in this treatment ask respondents whether they have generally experienced that more unequal places function better or worse than more equal places.

The full externality treatment is designed as the strongest externality treatment at the cost of precision. Realistically, redistributive preferences are composed of fairness concerns, externality beliefs, and several other factors. Crime or trust are only one part of each of these externality concerns. If our respondents are rational, even a large shift in the belief in a crime externality might be an overall small shift in their redistributive preferences which is not detectable even with a large sample size. The full externality treatment solves this issue by informing subjects about inequality externalities on a broader scale.

Treatment group 4: Fairness treatment The fourth treatment group receives information on how the wage-productivity gap has evolved since 1975, as shown in the screenshot in Figure E1, using data from the Economic Policy Institute. The stimulus includes information that blue-collars' wages stagnated while their productivity increased since the 1980s. The income share of the top 1% earners (from the World Inequality Database), on the other hand, increased sharply, indicating that the economic gains from the increase in productivity went for the most part to the richest Americans. The filler questions ask subjects to recall whether people they know that were employed in 1950 and 1980, and whether they thought these people were paid closer to what they produced than people with similar jobs today.

The treatment intends to give respondents information about the *fairness* of the economy, and functions as a comparable benchmark to the inequality externality treatments.

Control group 1: Active control The active control group receives a video that is structured similarly to those on trust and crime. The general video topic is how economic inequality metrics differ and how this can affect research on economic inequality. The video informs subjects about the difference between the Gini index and the top 10% income share. The video does not contain any information that is strictly speaking relevant for redistributive preferences, but does give individuals stimuli about inequality itself. Thus subjects are primed to think about inequality without revealing any information about inequality externalities or the fairness of the prevailing income distribution. The comparison across this control group and the treatments thus seeks to isolate the role of information. Filler questions to this treatment ask subjects to reflect on whether they (i) have already thought about the measurement of inequality and (ii) whether

[32] This was included due to our original assumption that individuals would believe in a positive association between these variables.

they knew that researchers had different ways of measuring inequality before the survey began.

II.E.3. Dual control groups

There are benefits and drawbacks to both a passive control group, where respondents see nothing, and an active control group, where respondents see information on a similar but unrelated topic. The main drawbacks of a passive control group are two-fold. First, if respondents see no stimuli/video, their overall attention and survey fatigue will likely differ from respondents who saw treatments in the post-treatment part of the survey. This could bias results due to attention issues and cause attrition problems. Second, priming from the information treatments – hearing the word "inequality" in our treatments, for example – could conceivably drive outcome differences. In surveys with only a passive control group these are untestable hypotheses. From these perspectives, an active control group is preferred. However, an active control group, no matter how well-designed, could always unintentionally convey information that affects the outcomes of interest. In experiments with only an active control group the existence of such unintentionally conveyed information is itself an untestable hypothesis. It is clear, then, that either method has significant drawbacks.

To solve these issues we introduce what we call *dual control groups*. This involves splitting the control into two groups; one active control and one passive control. The main outcomes are then compared across these two control groups and checked against pre-specified merging criteria. In the case that there are no significant differences across the control groups, the concerns stated above can safely be ignored and the two control groups can be merged.[33] The researcher can thus test whether attention effects and priming about the relevant concept (inequality, in our case) have important effects, which presents an improvement over the uncertainty involved in single control groups.

The idea of several control groups is not new; multiple control groups in observational studies have been discussed extensively [Rosenbaum, 2002]. Intentionally designed dual control groups in information experiments are less common, however, and have as far as we know not been employed or formalized before. The usage of pre-specified merging criteria in such groups is also as far as we know novel. Below we present a short description of each control group.

Control group 1: Active control The active control group receives a video that is structured similarly to those on trust and crime, informing subjects about the difference between the Gini index and the top 10% income share. The video does not contain any information that is strictly speaking relevant for redistributive preferences, but does give individuals stimuli about inequality itself.

Control group 2: Passive control This group does not receive any stimuli or filler questions.

Outcome differences between the two control groups were small and satisfied the pre-specified merging criteria. The control groups were thus merged and will be discussed as one larger control group for the remainder of the paper. The merging criteria and further discussion can be found in Appendix II.F.1.

[33] The only exception to this is in the very unlikely case that the attention effects from the passive control and unintentional information from the active control group has exactly the same effect.

II.E.4. Secondary survey

The short survey time (∼19 minutes on average) and lack of a follow-up survey presents potential issues with external validity for our survey experiment. The treatment effects we present are within-survey estimates, and we do not know whether these shifts are temporary or permanent in nature. This is an intrinsic limitation of our work which is also shared by other survey projects where obfuscated follow-ups [Haaland and Roth, 2020] are not possible. To remedy this we have taken steps to create an approximation of an obfuscated follow-up survey within the survey itself. We thus introduce what we call the *secondary survey* approach.

We motivate this through the reduction of experimenter demand and priming effects, which are key problems in traditional video information treatments. These issues are often partly circumvented by adding unrelated questions between the information treatment and outcome variables. However, this introduces a different issue; respondents who are presented the information do not understand *why* they were presented the information. This could lead to confusion or suspicion of future questions which could bias outcome data. To avoid this we design what we call a secondary survey, or a logical flow of questions that explains the information treatment while disguising the true purpose of the survey.

In the present experiment, the secondary survey relies on what we call "filler questions". These "filler questions" immediately follow the video and are directly related to the video content. All these questions focus on personal experiences related to the video topic. In the crime treatment, an example of one such question is the following: "*Have you lived in more than one place in your life? If so, think back – do you think the places with more economic inequality had more crime, generally speaking?*" These questions are designed to hide the purpose of the study by being directly related to the videos (and so explaining why the respondents had to watch them) while being unrelated to the true intent of the survey.[34] They thus create the impression that the videos are shown to lead into these filler questions and have no direct link with the rest of the survey.

To emphasize this connection we immediately end Part 2 of the survey after the filler questions, notifying respondents of this. We then start Part 3 with an introduction screen, upon which we continue with several unrelated demographic questions to create the appearance of each survey part being functionally independent. Our true treatment effects are all based on questions in Part 3 (see Figure 3). The respondents have thus seen a self-contained *secondary survey*, which should minimize experimenter demand and priming effects. While many survey experiments employ some structural break between treatment and outcome, we believe the formalization of this broader concept is a beneficial addition to the literature.

We note that we cannot guarantee that the filler questions themselves do not change individuals' beliefs about the video topic. The crime question above could conceivably change individuals' beliefs about how economic inequality affects crime, for example. This is less problematic than it might seem, however, as the origin of respondents' opinion change – whether from the video or the filler questions – is of second-order importance to our research questions.[35]

[34] The specific filler questions are discussed further in Appendix II.E.2.

[35] First-stage responses and mediation analysis in Section 5.2.2 show that the treatment mechanism appears to go through a shift of the intended beliefs.

II.E.5. Theoretical mechanism of the information experiment

What theoretical mechanism drive a potential treatment effect? We use the preference function (2.1):

$$\cup_i = x_i - \sum_j \gamma_{ij} \mathbb{E}_i(\alpha_j)\theta + \Upsilon_i. \tag{B.7}$$

Individual i's stated redistributive preferences depend on their income x_i, which is not changed from a video treatment, the net effect of their inequality externality beliefs $\sum_j \gamma_{ij}\mathbb{E}_i(\alpha_j)\theta$, and other determinants Υ_i. We assume these other determinants can be broken down as $\Upsilon_i = G_i + X_i$, where G_i denotes broad economic fairness views and X_i denotes a set of other characteristics including attention, mood, immutable characteristics, and so on. A video information treatment T_q, where q determines the type of information treatment – T_α denoting the externality treatments – can affect inequality externality beliefs, broad economic fairness views, or other determinants;

$$\frac{d\cup_i}{dT_q} = -\sum_j \gamma_{ij}\theta \frac{\partial \mathbb{E}_i(\alpha_j)}{\partial T_q} + \frac{\partial \cup_i}{\partial G_i}\frac{\partial G_i}{\partial T_q} + \frac{\partial \cup_i}{\partial X_i}\frac{\partial X_i}{\partial T_q} \tag{B.8}$$

These are the three main channels through which any of our treatments can affect redistributive preferences. We are specifically interested in whether $\gamma_{ij} \neq 0$ for at least some j, which would imply that inequality externality beliefs $\mathbb{E}_i(\alpha_j)$ are a causal determinant of redistributive preferences for individual i.

As the active and passive control are similar in outcomes (see Appendix II.F.1), we can be confident that an inequality-related video generally has only a limited (if any) affect on redistributive preferences if the topic of the video is not related to redistribution. This is because the active control video discusses differences between inequality metrics, an inequality-related topic that is (theoretically) orthogonal to preferences for redistribution. This implies that $\frac{\partial RP_i}{\partial X_i}\frac{\partial X_i}{\partial T_q} \approx 0$. In other words, showing respondents a video about inequality-related issues does not significantly change their redistributive preferences due to attention effects, priming, or any other change to the broad set of characteristics defined as X_i.

If the externality treatments T_α have limited spillovers on fairness views, we also have that $\frac{\partial G_i}{\partial T_\alpha} \approx 0$. If finally the externality treatments affect externality beliefs themselves, we have that $\frac{\partial \mathbb{E}_i(\alpha_j)}{\partial T_q} \neq 0$. From Equation B.8 we can then conclude that,

$$\frac{d\cup_i}{dT_\alpha} \neq 0 \rightarrow \gamma_{ij} \neq 0 for some j. \tag{B.9}$$

Thus, a significant treatment effect – or that $\frac{d\cup_i}{dT_\alpha} \neq 0$ – imply that at least some inequality externality beliefs causally affect redistributive preferences.

II.E.6. Treatment outcomes and first-stage beliefs

Treatment outcomes Below is the four redistributive preference outcome questions in full. The questions were presented in this order.

1. **"Wants redistribution":** How much redistribution of income do you prefer across citizens in the U.S.? *No redistribution means that the initial level of inequality is kept. Full redistribution means that all citizens should have the same income.*

- Slider 0-7, 0="No redistribution" to 7="Full redistribution"

2. **"Increase top taxes":** In your view, which average income tax rate should the richest 10% of households in the U.S. pay?

- Seven options, e.g. "25-35%: I want to tax them at roughly what they are taxed now."

3. **"Gov. reduce ineq.":** To what extent do you agree or disagree with the following statement: *The government should take measures to reduce differences in income levels.*

- Totally disagree / Disagree / Neither agree nor disagree / Agree / Totally agree

4. **"Ineq. is serious issue":** How big of an issue do you think income inequality is in America?

- Not an issue at all / A small issue / An issue / A serious issue / A very serious issue

The redistributive preference index ("RP Index") was pre-specified as the sum of dummy versions of these four outcomes. The binary split of each outcome was pre-specified with the goal of keeping even 50-50 splits. The index was then standardized such that the units in Table 1 are in population standard deviations.

First-stage beliefs The first-stage beliefs we discuss in the main text are represented by seven questions, all asked after individuals' redistributive preferences in the following order:

1. **General externality question:** This question is about what economic inequality does to society. Generally speaking, do you think more economic inequality changes society for the better or for the worse?

2. **Open-ended text question:** How do you think economic inequality changes society? For this question we want to hear your ideas and opinions more broadly. Some example answers would be "Society would become more/less ____" or "____ would increase/decrease" (where you write whatever you think instead of ____). But these are just examples; feel free to use your own words! Remember that there are no wrong answers, and that we appreciate it if you put some thought into the response.

3. **Crime externality:**[36] Please pay very close attention to this question. How does more inequality change the amount of crime in a country? Note: When we say the amount of crime we mean the overall crime rate, including homicides, robberies, property crime and more.

4. **Trust externality:** Please pay very close attention to this question. How does more inequality change the overall level of trust in a country? Note: When we say the total level of trust we mean the strength of a country's social fabric. Some examples are whether most people

[36] We note that respondents were asked to be particularly attentive when shown the questions regarding crime and trust. This can be seen in the questionnaire (Appendix II.C.1). We made this design choice to maximize attention and minimize measurement error. We acknowledge that this reminder may induce a demand effect in these responses. This would not affect any questions placed earlier in the questionnaire, notably (i) our main outcome questions (on redistributive preferences), and (ii) the open-ended text question detailed below, which provides further support that the videos targeted the relevant belief.

trust others, whether people cooperate with each other, how many people return lost wallets, and so on.

5. **Fairness:** Do you feel that the distribution of money and wealth in this country today is...

   - fair, because everybody gets what they are entitled to, or
   - unfair, because some get much more than they are entitled to, while others get too little?

6. **Luck vs. Effort:** Which has more to do with why a person is rich?

   - Is someone rich because he or she worked harder than others, or
   - because he or she had more advantages than others?

   (Please pick the one closest to your views, even if it does not match your view perfectly.)

7. **Growth externality:** How does more economic inequality change the rate of economic growth in a country?

II.F. Further Data: Information Experiment

II.F.1. Dual control groups

In this section, we compare the respondents' characteristics and outcomes across the two control groups. We pre-specified to merge these to groups conditional on being sufficiently similar. Specifically, we pre-specified the following decision rule:

"If the active and passive control group are sufficiently similar, we will merge them for the main analysis. This decision will be made upon not reaching all the three following criteria.

- There is no 1% statistical difference in the index outcome variable between the active and passive control.
- There is not a 5% statistical difference in at least three of the four redistribution dummy variables listed above.
- There is not a 5% statistical difference in at least three of the four externality dummy variables listed above.

If one of these criteria are reached, we will present regressions with both control groups as separate categories."

Table F2: Dual control: Balance table for redistributive preferences

Variable	(1) Passive Control	(2) Active Control	(3) Difference
RP Index	-0.111	-0.045	0.067
	(0.965)	(0.984)	(0.065)
Wants redistribution	0.370	0.360	-0.009
	(0.483)	(0.481)	(0.032)
Increase top taxes	0.537	0.622	0.085***
	(0.499)	(0.486)	(0.033)
Gov. reduce ineq.	0.480	0.508	0.028
	(0.500)	(0.501)	(0.033)
Ineq. is serious issue	0.515	0.508	-0.007
	(0.500)	(0.501)	(0.033)
Observations	538	394	932

Note. This table represents mean (standard deviations) for redistributive preference measures of respondents in the active (column 1) and passive (column 2) control groups. Column (3) characterizes the difference across the two. The pre-specified criteria to merge these two control groups for the main analysis is satisfied. *Significance levels*: *10%, **5%, ***1%.

As shown in Table F2 the index is not significantly different across the two groups. From the redistributive preference variables, only the variable on top tax-rates is significantly different across the two groups.[37] The other variables are not significantly different between control groups; the differences are also relatively small and in opposing directions. As pre-specified, we thus merge the two groups for the main analysis.

We also compare first-stage post-treatment outcomes (inequality externality beliefs and fairness views) across the two groups and find no significant difference between the two groups on any of these outcomes (see Table I4). This indicates that the difference for the top tax rate could be spurious, as other strong predictors of redistributive preferences such as fairness views are balanced across the two groups. It is also possible, however, that the quantification of the top 10% income share in the active control video made respondents who saw this video prefer a higher income tax rate for the same top 10%. If so, this would bias our main treatment effects on this variable downwards. We note that such unexpected effects are a good motivation to use dual control groups.

As shown in Table I5, there are no significant differences between the two groups on any pre-treatment dimension. Table I6 shows that they are also comparable on various sociodemographic characteristics.[38]

Overall, the results show that the two control groups are sufficiently similar to be merged

[37] This could be simple statistical noise; it is also possible that mentioning the top 10% income share shifted individuals' top tax rate preferences. We note that unexpected discrepancies like these are a strong motivation for the dual control group method.

[38] We find that the two groups are mostly balanced apart from a few exceptions. Subjects in the active control group are less likely to be neither black nor white, and are somewhat differently allocated into the three income groups. Note that these differences are not large and including them as control variables does not affect the differences in redistributive preferences or first-stage outcomes. Beyond that, passive control group subjects are not more or less likely to pass all three attention checks build into the survey than active control groups. Neither are they more nor less likely to pass an attention check that was administered *after* the treatment.

and can be treated as one control group. While there are few idiosyncratic differences across the two groups, they are non-systematic and likely to be spurious, reflecting the fact that we are testing many hypotheses at once.[39] Following our pre-analysis plan, we thus merge the two groups.

II.F.2. Balance across control and treatment groups

This section checks the pre-treatment balance of control and treatment groups. As shown in Table I7, the crime and control groups are balanced on nearly every dimension. There is one important exception; subjects in the crime treatment group have significantly higher perceptions that unequal countries usually function worse. However, including this perception as a control variable in outcome regressions does not affect the results of the analysis.

Table I8 compares observable characteristics across the trust and control groups. The two groups are completely balanced on observables.

Table I9 compares observable characteristics across the full externality treatment and the control group. The full externality group has somewhat fewer individuals in high income households but more individuals from middle-income households. Respondents in this group are also somewhat more likely to be highly educated and to believe that working-class Americans are paid less than their productivity. The main results do not change when including or excluding these data points as controls. The correlations are also less significant than the first-stage and outcome treatment effects, and respondents in this group do not statistically differ from the control group in other fairness views elicited either pre- or post-treatment.

Table I10 compares observables across Fairness and Control group. The two groups are balanced on all covariates with the exception of gender (slightly more in the Fairness group) and the number of individuals from middle-income households (slightly more in the Fairness group).

II.F.3. First-stage beliefs: Open-ended text question

The open-ended text question (shown in Appendix II.E.6) asks respondents to write about how they think inequality changes society without prompting them specifically in any further direction. The share of answers that include the words "crime" or "trust" strongly increases in the corresponding treatment groups (shown in Table F3). As an example, the word "crime" is used by about 15% of the crime and full externality treatments, and only about 4% of any other treatment or control group. To ensure that this is not driven by respondents simply describing the video, we check the equivalent for the word "video", which is barely mentioned by respondents in any group (0.18% of all respondents). This also holds for other similar words ("Youtube", "infographic", and so on). This highlights that the video as such is barely discussed in the answers; instead respondents discuss the informational content itself. This indicates the success of the *secondary survey* we describe in Section 5.1, softening concerns about experimenter demand.

II.F.4. Mediation analysis

This section describes the mediation analysis results in Table I24. This table includes post-treatment first stage outcomes in the regression of redistributive preferences on treatment variables. Compared to the treatment effect of a regression without post-treatment beliefs, the

[39] The potential exception to this being the top tax result.

Table F3: Share of subjects mentioning "crime", "trust" or "video" in open-ended question

	Mentioned crime (%)	Mentioned trust (%)	Mentioned video (%)
Crime tr.	**17.04**	0.32	0.43
Trust tr.	4.48	**6.30**	0.12
Full ext tr.	**13.23**	**3.71**	0.37
Fairness tr.	4.13	0.23	0.00
Control (passive)	4.46	0.32	0.00
Control (active)	4.57	0.00	0.00

coefficients of the treatment dummies decrease when the post-treatment first stage outcomes are included. For each treatment effect this effect is driven by the post-treatment belief in question (e.g. externality belief for the externality treatment). This strongly indicates that at least part of the treatment effect is driven by changes in first-stage beliefs.

More concretely, the treatment effect of the full externality treatment on our redistributive preference index was 10.7 percent of a standard deviation if we do not control for post-treatment externality beliefs and decreases to 5.5 percent of a standard deviation once we control for post-treatment externality beliefs. This implies a reduction in the magnitude of the treatment effect of nearly 50% (p = 0.002, t-test). The reduction in the magnitude of the fairness-treatment's treatment effect is similarly large. Before controlling for beliefs, the magnitude of the fairness treatment was 20.8 percent of a standard deviation and then decreased to 13.5 percent of a standard deviation (p=0.000, t-test). This provides evidence that our reduced form treatment effect is mediated through a shift in beliefs, as intended by the treatment itself.[40] When also including the opposite post-treatment belief (e.g. externality beliefs for the fairness treatment) treatment effects remain similar (5.5 → 5.8 for the full externality treatment and 13.5 → 12.2 for the fairness treatment), indicating that the treatment effect is not driven by spillovers.

II.F.5. Robustness of treatment effects

Population weights Even though we targeted representativity along several observable dimensions, we slightly over- or under-sample populations with some characteristics as described in Section 3.3. To establish representativity ex-post, we replicate our key analyses by reweighting along gender, race, age-groups, party, holding a college degree, income group, and geographic region. Regressions in Table I21 regress redistributive preferences on our treatments; Regressions in Table I23 regress posterior beliefs on treatments; and Regressions in Table I33 replicate the horse-race regressions using population weights. The results for the latter two regressions are nearly identical. For the former, reweighting has only small effects on the magnitude of the significant treatment effects. As standard errors increase under the reweighting procedure,

[40] A complete disappearance of the treatment effect is unlikely given that beliefs are generally measured with noise and that our first-stage belief measurements are bounded. An example of this would be an individual who already thought inequality increases crime before the survey; after watching the full externality video she becomes increasingly convinced of the importance of this causal channel, which shifts her redistributive preferences. Her response to the first-stage crime question is the same (*"More inequality → A lot more crime"*). However, her beliefs have changed, which then affect her redistributive preferences.

certain clearly significant treatment effects in the original weighting are, however, no longer 5% significant in the reweighted data.

Keeping all respondents As prespecified, we dropped the 5% fastest and slowest respondents, as well as those that spent less time watching the video than the length of the video. Additionally, we dropped respondents with unusual or strange responses to open text questions. We replicate our main regressions keeping these respondents. As shown in Tables I34, I35 and I36, we do not find any meaningful differences compared to the analyses using our main sample.

Failing any attention check We also replicate our main regressions while excluding all respondents that failed *at least* one attention check. While the first-stage effects and the horserace regressions remain very similar to our main specification (Table I38 and I39 respectively), the effect of the full externality treatment on RP-Index becomes marginally significant as shown in Table I37. Given that controlling for passing or failing an attention check does not result in any differences, as shown in Table I40, this is likely due to the lack of power that results from dropping one-third of our sample.

Specifying only one control group As shown in Appendix II.F.1 we merge our two control groups given that they are sufficiently similar on a set of pre-specified criteria. As a robustness check, we first-stage and reduced form treatment effect regressions but drop either the active or the passive control group in Tables I41, I42, I43, and I44. The treatment effects are slightly stronger when only considering the passive control group as the baseline compared to when only specifying the active control group as the baseline, and overall results are robust to either specification.

We briefly discuss the full externality group specifically as this is important for our main hypothesis. Results are qualitatively unchanged and slightly stronger in magnitude when using only the passive control group. When using only the active control group, treatment effects still go in the expected direction. The RP-index treatment effect of the full externality treatment is only marginally significant in this setting, however, due to lower statistical power from the smaller control group. The magnitude of the effect is slightly smaller but comparable to the standard full control group specification.[41]

Not controlling for observable characteristics We replicate our main regressions without controlling for any observable characteristics. As shown in Tables I45 and I46 reduced form and first-stage treatment point-estimates are nearly identical to our main specification in magnitude and significance. This is expected given our randomized treatment design.

Different sets of controls We pre-specified a vector of control variables to evaluate the treatment effects. The results do not change significantly when we change this vector to any other reasonable permutation (as expected from our randomized experiment design). Notably, our results do not changes if we include prior externality beliefs in the set of controls. Due to the large number of such permutations we do not explicitly show these results.

Using non-dichotomized outcome variables In our main specifications, we dichotomize our outcomes and explanatory variables when applicable. In Tables I47 and I48 we replicate our

[41] We note that the high top tax rate result from the active control group leads to a *negatively* significant top tax rate result for the full externality treatment.

main regressions without dichotomizing any outcomes or control variables and, furthermore, we recompute the RP-Index based on non-dichotomized beliefs. As shown in the tables, the results are nearly identical to those presented previously.

Multiple hypothesis testing In the main regression tables (Table 1) we run a total of twenty tests for statistical significance. On this scale, Type I errors can become a serious problem and lead to erroneous inference of statistical significance. To correct for this we use the false discovery rate (FDR) sharpened q-values as described in Anderson [2008]. FDR sharpened q-values are classical p-values that are corrected for the expected number of significant treatment effects that are truly null effects. Where a p-value threshold of 0.05 gives a false positive rate of 5% among all treatment effects that are truly null, a q-value threshold of 0.05 gives a false *discovery* rate of 5% among all *significant* treatment effects. This correction has no significant effect on our conclusions. None of the treatment effects with $p < 0.05$ in our original specifications have q-values above 0.05, indicating that this is a negligible concern. The results of this correction are shown in Table I51.

II.F.6. Fairness video

The fairness video is described in Appendix II.E.2.

As shown in Appendix Table I20, the treatment has strong first-stage effects on broad economic fairness views. These first-stage effects are evenly distributed among party affiliations and incomes (not shown). The treatment also has a significant first-stage effect on general externality beliefs, however, which may indicate that learning about income distribution dynamics also affects individuals' externality beliefs.[42]

As shown in Appendix Table I18, the treatment has highly significant effects on redistributive preferences across all four redistribution outcomes, as well as the main index itself. This is in itself a meaningful result, as changing survey respondents' redistributive preferences is often challenging [Kuziemko et al., 2015]. The treatment effect is sizable at approximately $\frac{1}{4}$ of the difference between Republican-leaning and Democrat-leaning respondents.

We note that one of our outcomes – inequality being a serious issue – was also asked in Stantcheva [2021]. The treatment effects of the videos in that work (2% - 9%) are similar to those we find (2% - 12%). The main redistribution treatment effect in Stantcheva [2021] is 9%, which is similar to the treatment effect in our fairness treatment (12%).

II.G. Further data: Comparing inequality externality beliefs to other determinants

II.G.1. Ranking motives behind preferences for redistribution

This section discusses Figure 6, which provides direct evidence of respondent's redistributive motives under the assumption that they are able to discern and report these motives.

The motive that attains the highest average support is income maximization. This is closely followed by a diminished marginal utility (DMU) argument that a dollar is worth more to the rich than to the poor. Negative externalities ("*Inequality changes society for the worse (more*

[42] It is also possible that some respondents simply interpreted this question as being focused on fairness issues, as it asks whether "more economic inequality changes society for the better or for the worse".

inequality $\longrightarrow$ *a worse society through various ways*)") are the third most important motive, attaining an average of 18 points. A broadly framed fairness motive ("*High inequality is unfair*") actually ranks slightly behind the inequality externality motive (p = 0.001, t-test). A general aversion against taxation, positive externality concerns, and efficiency concerns attain only weak average support from our sample. This last point on efficiency concerns is consistent with the findings in Table 2 and in Stantcheva [2021], among others; efficiency concerns do not seem to be strong determinants of U.S. citizens' redistributive preferences.

What does this tell us about the relative importance of externality concerns and fairness views? First, we note that inequality externality concerns rank as one of the most important motives within our sample. This is remarkable, and cannot be explained by experimenter demand given that this is essentially the first question about inequality externalities that respondents are faced with in the survey.[43] Second, negative externality concerns are similar in magnitude as broad but explicit fairness views. When comparing a combining positive and negative externality grouping with a combined "broad" fairness classification including both DMU and the fairness motive, general externality concerns are about three-quarters (74%) as important as fairness motives as a redistributive determinant - thus echoing the results from the two methods described above.[44]

One may argue that the presented averages just reflect idiosyncratic noise and not clear motives behind preferences for redistribution. This is, however, unlikely to be the case. Figure H17 in the Appendix shows the share of subjects that weakly rank a given motive first for the same question; the distribution strongly resembles that in Figure 6. One can also replicate Figure 6 while only including the sub-populations of subjects that rank a given motive first. This is presented in Figure H18. This decomposition shows both that subjects have consistent views – the positive externality answer is at the bottom for the negative externality group and vice versa, for example – and that respondents can be described as having one primary motive and other secondary motives. Results across Survey 1 and Survey 2 remain very similar; we show the data restricted to Survey 2 respondents in Figure H16.

II.G.2. The predictive power of externality beliefs

In the second method we explore the predictive power of each type of belief on redistributive preferences. We run descriptive pre-specified regressions that include either fairness views, externality beliefs, political preferences, or "economist determinants"[45] as regressors. We then compare the explanatory power of these models using the adjusted R^2.

Table 2 displays the results of these regressions.[46] Fairness variables have the most predictive

[43] In Survey 1 (for which only the control group is included), respondents have been shown one pre-treatment externality question ("*Do more equal countries function worse?*"). It is the first time the concept is mentioned in Survey 2 (about 2/3 of the respondents in Figure 6). Results are very similar across samples, see Figure H16 for only Survey 2 data.

[44] If including "Taxation is theft" as a fairness motive, this falls to 60%.

[45] Respondents' trust in government and belief that higher taxes lead to efficiency losses.

[46] Column (1) characterizes a regression that only includes demographic controls; Column (2) includes our two main fairness variables, the belief that society is unfair because some get much more than they are entitled to and some get too little, and the belief that one gets rich due to luck rather than hard work; Column (3) includes our two main externality variables, the belief that unequal countries generally function worse, and the belief that inequality generally affects society in a negative way; Column (4) includes the strict political variables of whether the respondent leans Republican and whether the respondent supports Kamala Harris and Bernie Sanders (rather than Mitt Romney or Donald Trump); Column (5) includes two variables economists often consider as potential

power in this specification; demographic controls and two fairness survey questions explain 38% of the variance in respondents' redistributive preferences. This is followed by the externality beliefs and political views, which are equally predictive at $\sim$ 30%. The "economist" regression has a relatively low predictive power of 15%, only slightly above the only-controls regression (10%).[47]

We can also explore whether externality views provide any *additional* predictive power to a fairness-based model of preferences for redistribution. Model (5) indicates that it may; when including all variables into a single regression, all fairness- and externality-variables remain strong predictors of redistributive preferences with somewhat lower point estimates. This indicates that while these views are to some extent correlated to each other, each still captures *independent* correlation with redistributive preferences.[48] We note that the externality questions perform much better in this exercise than the "economist" determinants; the individual's opinion on whether taxation reduces work effort is no longer significant in the combined regression, for example.

This method allows us to make two separate conclusions. First, the predictive power of inequality externality beliefs on redistributive preferences is strong; the total predictive power is only somewhat less than that of fairness views, and roughly the same as that of political affiliation. Second, externality beliefs capture information on redistributive preferences that is not included in individuals' fairness views.

Gelbach decomposition: Motives behind partisan redistributive split One might also ask how much of the partisan split in redistributive beliefs is explained by variation in externality and fairness views respectively. To explore this question we employ a Gelbach decomposition [Gelbach, 2016]. We use the decomposition to illustrate which portion of the partisan gap in the redistributive preference index goes through either the two main externality variables, the two main fairness variables, governmental trust, or efficiency concerns.[49] In total, 54% of Republicans' lower support for redistribution can be explained by these variables or a list of standard controls. About half of this can be accounted for by fairness views (27% of the partisan gap). Then comes externality beliefs (12%), demographic controls (10%), and governmental trust (5%). Efficiency concerns are not a relevant factor ($\sim$0%).

This analysis indicates that externality beliefs explain part of the partisan split in redistributive preferences, although a somewhat larger portion is driven by fairness views. This is consistent with the stronger partisan split across fairness views than externality beliefs shown in Figure 9. At the same time, it is also notable that externality beliefs explain more of the divide than trust in government and efficiency concerns.

determinants for redistributive preferences, namely whether the respondent generally trusts the government to do the right thing and whether the respondent agrees that higher taxes make people work much less; Column (6) displays the results of a regression that includes all variables from regressions (1) through (5). Note that all regressions only include observations from the baseline control group.

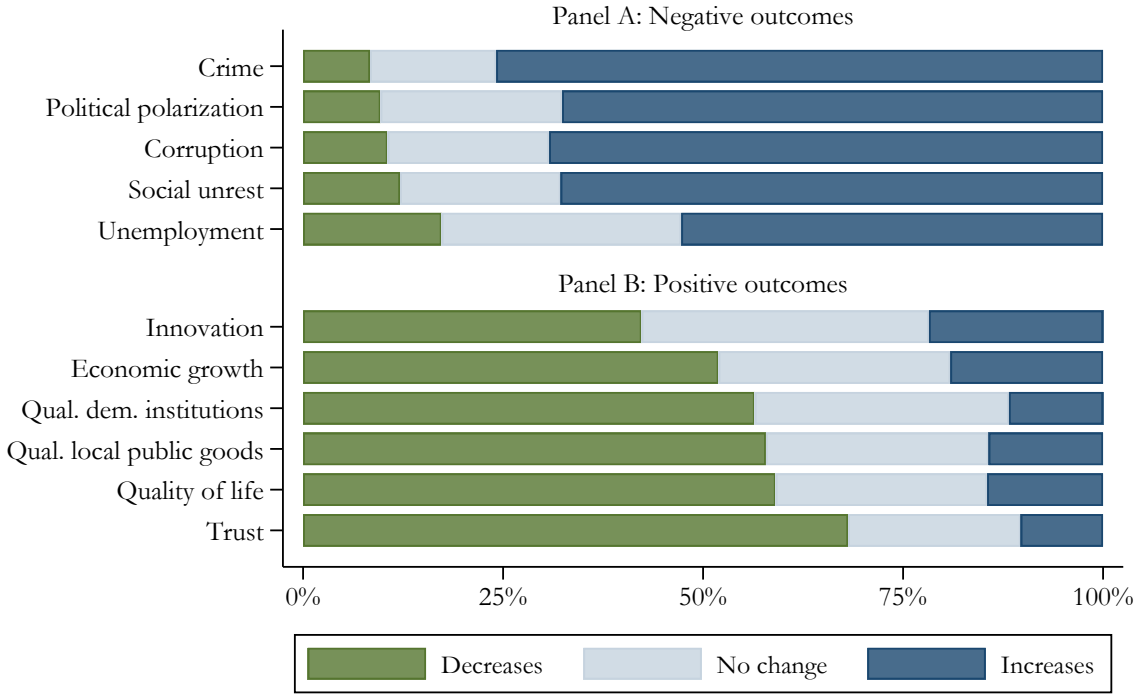
[47] Note that these results are replicated in Survey 2, see Table I32.

[48] Similar results hold when exploring three-variable versions of the fairness and externality modules.

[49] The variables are the same as in Table 2.

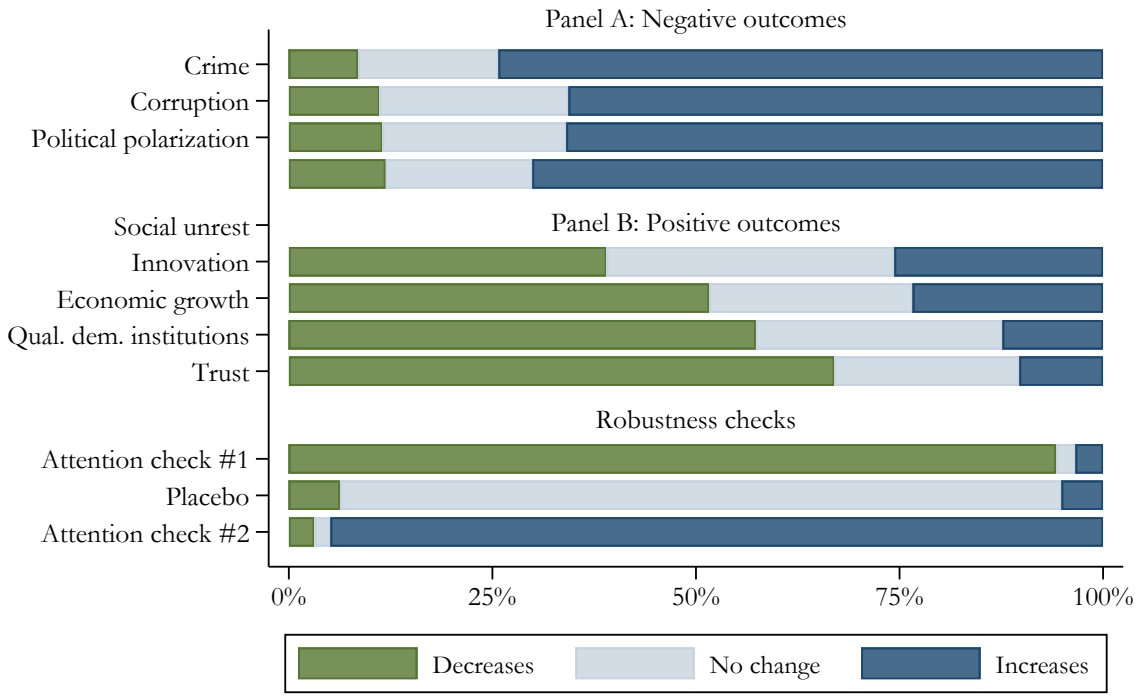
II.H. Figures

Figure H1: Distribution of Externality Beliefs in Survey 1 (Control Group)



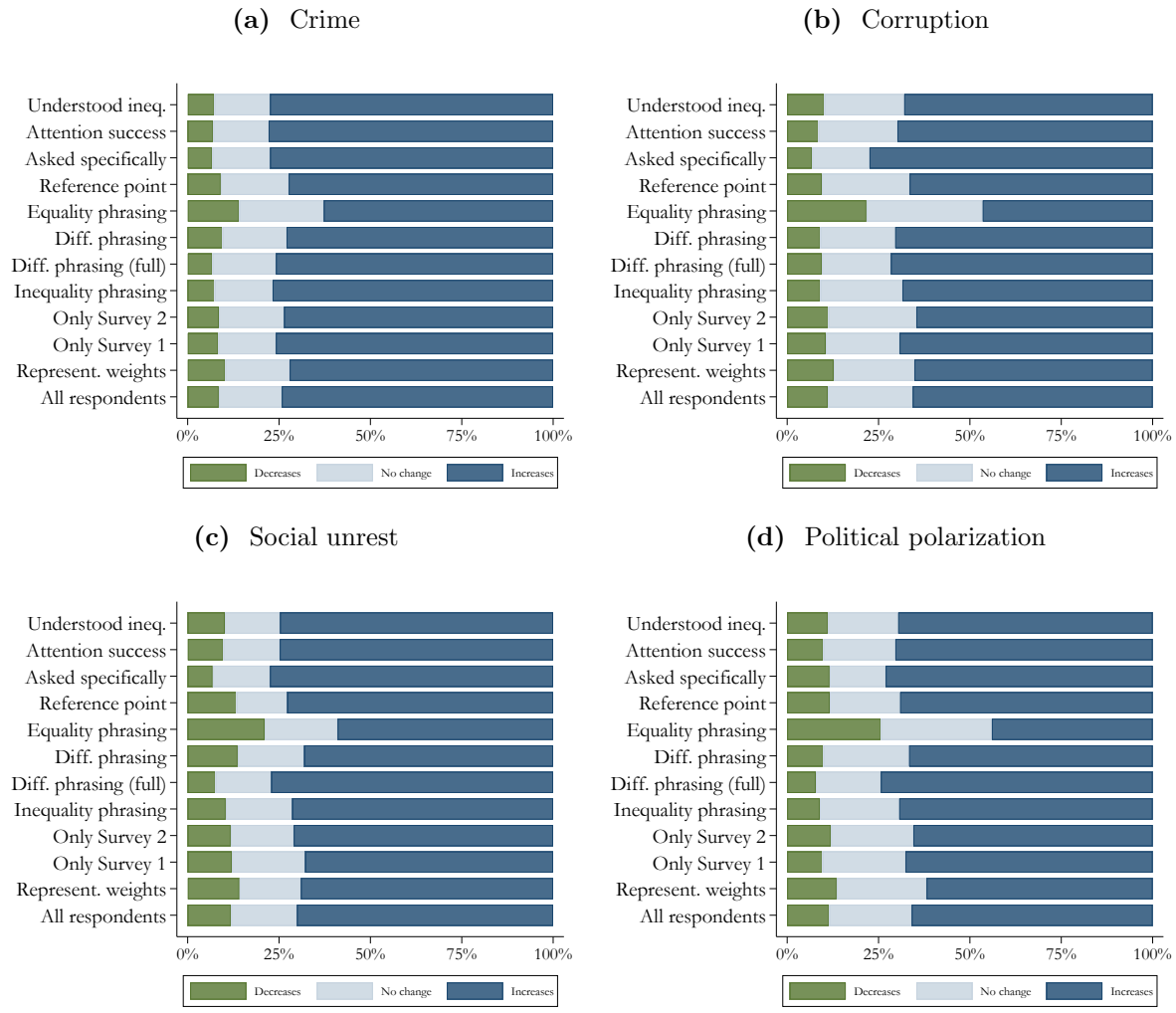
Note. Specific externality beliefs for control group in Survey 1. Questions are ordered according to which portion of respondents believe that inequality decreases the variable. Full question example: "How does more economic inequality change the amount of crime in a country?" Answer option example: "More inequality → a lot more crime". $N \in \{628, 3, 292\}$. For the equivalent figure using pooled data or only data from Survey 2 respectively, see Figures 1 and H2.

Figure H2: Distribution of Externality Beliefs in Survey 2



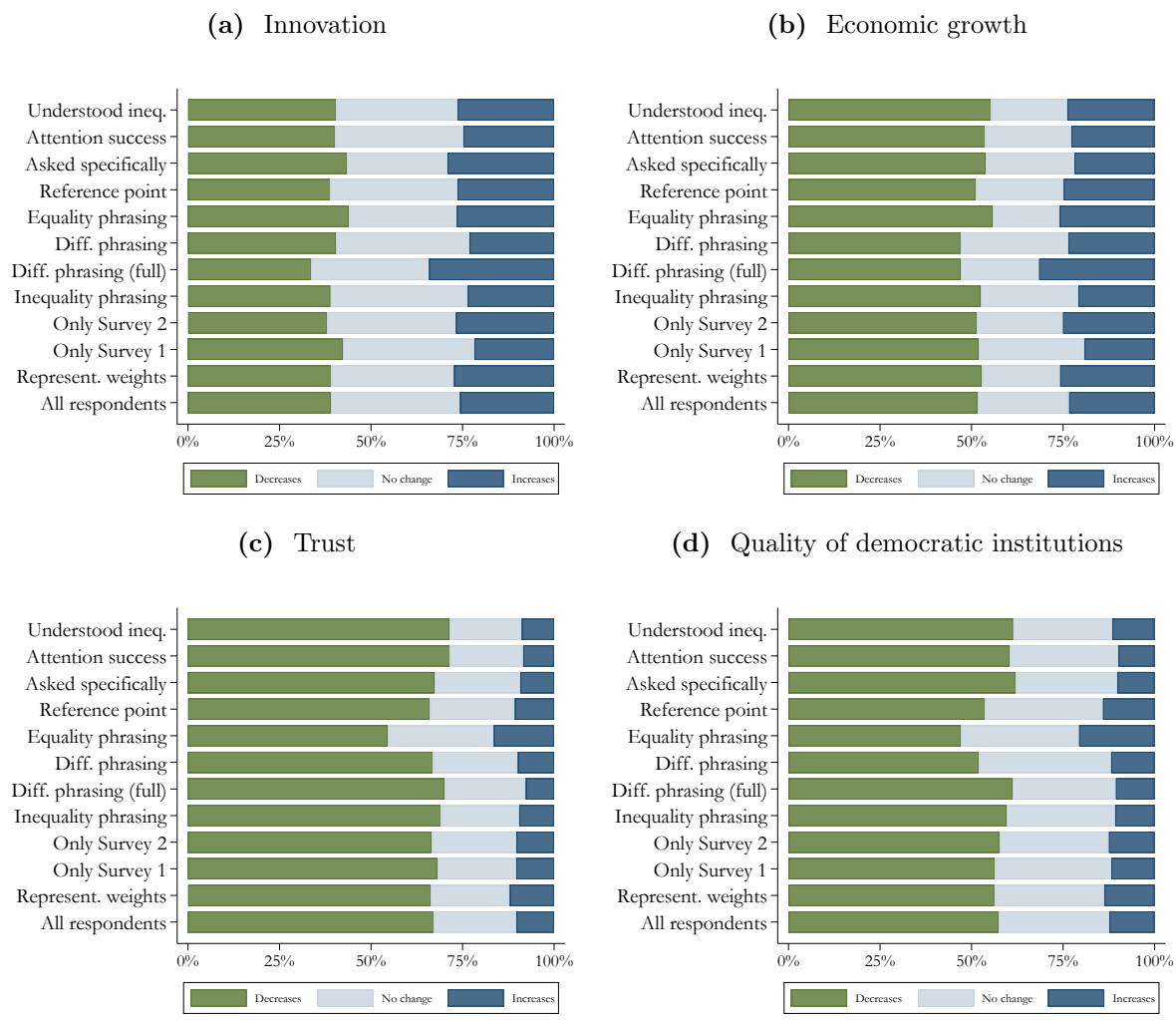
Note. Questions are ordered according to the net share of respondents who believe that inequality decreases the variable (except robustness checks). Full question example: "How does more economic inequality change the amount of crime in a country?" Answer option example: "More inequality → a lot more crime". The placebo question asks respondents how they think more economic inequality would change the amount of daylight hours in a country. The two attention check questions ask the respondents explicitly to answer either "Decreases a lot" (Attention check #1) or "Increases a lot" (Attention check #2). The high share of respondents who correctly answer these questions is partly mechanical, as individuals who incorrectly answered at least two attention checks were removed from the sample. $N = 2,360$. Order was fully randomized. For 20% of respondents, any mention of "more inequality" was substituted with "larger differences in income and wealth" throughout the survey. For 20% of respondents, any mention of "more inequality" was substituted with "more equality" throughout the survey. A respondent answering "decreases" to an equality-based question is coded as "increases" in the graph and vice versa. For the equivalent figures using pooled data or only data from Survey 1, see Figures 1 and H1 respectively.

Figure H3: Robustness of externality beliefs I



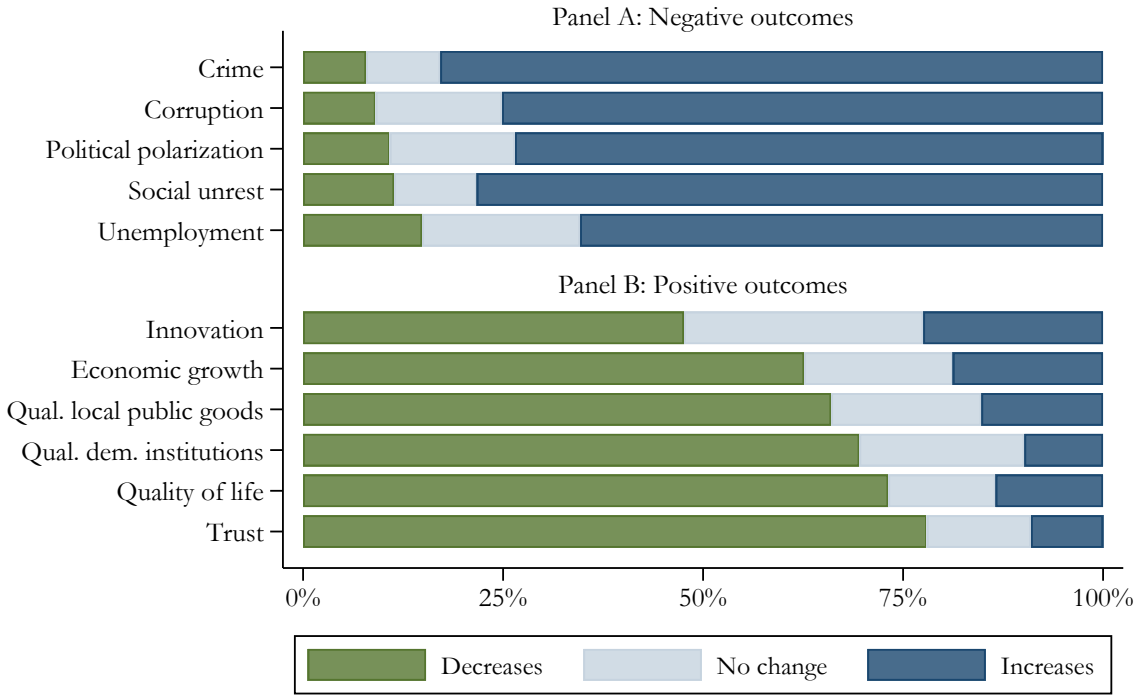
Note. Answers to *How does more economic inequality change the amount of [outcome] in a country?*, indicating inequality externality beliefs, across various specifications. The different specifications are, from bottom to top; (i) All respondents and all phrasings in both surveys, $N \in \{2990, 3292\}$, (ii) All respondents weighted for full representativity on age, gender, race, college attendance, income, region, and party affiliation, $N \in \{2990, 3292\}$, (iii) Only respondents from Survey 1 $N \in \{630, 932\}$, (iv) Only respondents from the Survey 2 $N = 2360$, (v) Only respondents who saw an "inequality" phrasing in either survey, $N \in \{2043, 2345\}$, (vi) Only respondents who saw a "differences in income and wealth" phrasing for the full survey (only in Survey 2), $N = 472$, (vii) Only respondents who saw a "differences in income and wealth" phrasing for this question, but were generally asked about inequality otherwise (only in Survey 1), $N \in \{219, 332\}$, (viii) Only respondents who saw an "equality" phrasing for the full survey (only in Survey 2), $N = 475$, (ix) Only respondents who were explicitly told the reference point of inequality and the magnitude of inequality change (only in Survey 2), $N =\in \{1748, 1777\}$, (x) Only respondents who were explicitly asked to think through their answer and were given 15 seconds to do so, then asked to confirm their answer or change it, if they wished (only in Survey 2, and at the end of the survey), $N \in \{292, 298\}$ (xi) All respondents restricted to those who succeeded on every attention check, $N \in \{1677, 1873\}$, (xii) All respondents from Survey 2 who correctly answered a simple question on distributional concepts $N = 1571$. Treatment groups from Survey 1 are always excluded.

Figure H4: Robustness of externality beliefs II



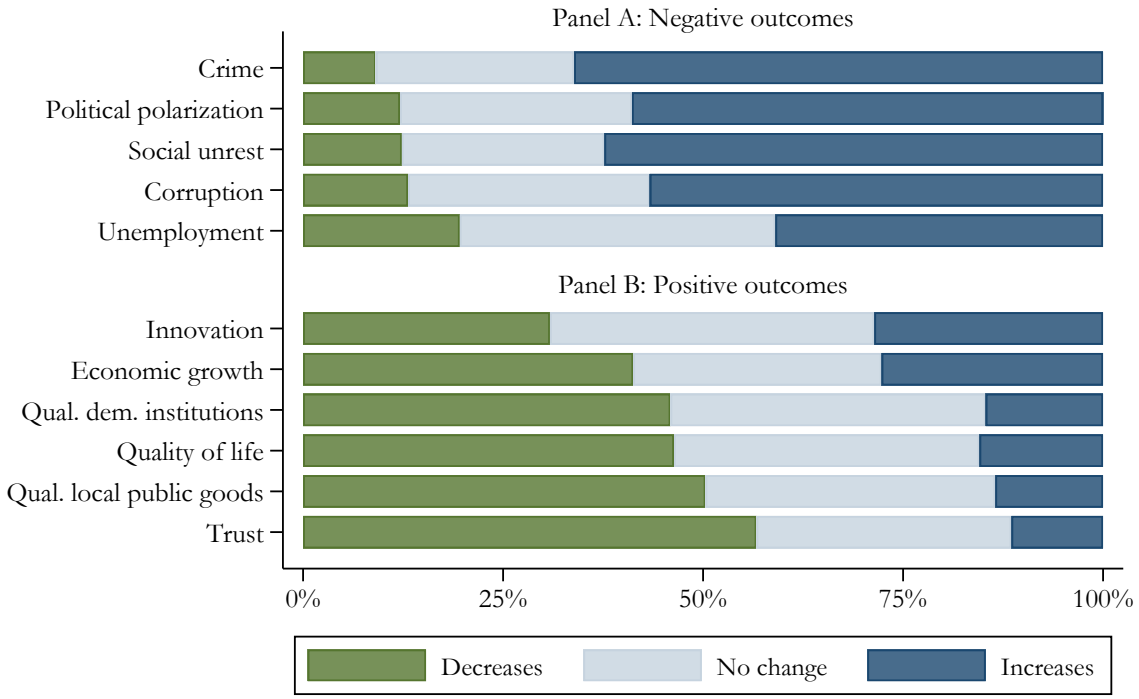
Note. Answers to *How does more economic inequality change the amount of [outcome] in a country?*, indicating inequality externality beliefs, across various specifications. The different specifications are, from bottom to top; (i) All respondents and all phrasings in both surveys, $N \in \{2990, 3292\}$, (ii) All respondents weighted for full representativity on age, gender, race, college attendance, income, region, and party affiliation, $N \in \{2990, 3292\}$, (iii) Only respondents from Survey 1 $N \in \{630, 932\}$, (iv) Only respondents from the Survey 2 $N = 2360$, (v) Only respondents who saw an "inequality" phrasing in either survey, $N \in \{2043, 2345\}$, (vi) Only respondents who saw a "differences in income and wealth" phrasing for the full survey (only in Survey 2), $N = 472$, (vii) Only respondents who saw a "differences in income and wealth" phrasing for this question, but were generally asked about inequality otherwise (only in Survey 1), $N \in \{219, 332\}$, (viii) Only respondents who saw an "equality" phrasing for the full survey (only in Survey 2), $N = 475$, (ix) Only respondents who were explicitly told the reference point of inequality and the magnitude of inequality change (only in Survey 2), $N =\in \{1748, 1777\}$, (x) Only respondents who were explicitly asked to think through their answer and were given 15 seconds to do so, then asked to confirm their answer or change it, if they wished (only in Survey 2, and at the end of the survey), $N \in \{292, 298\}$ (xi) All respondents restricted to those who succeeded on every attention check, $N \in \{1677, 1873\}$, (xii) All respondents from Survey 2 who correctly answered a simple question on distributional concepts $N = 1571$. Treatment groups from Survey 1 are always excluded.

Figure H5: Externality Beliefs for Democratic-leaning Respondents



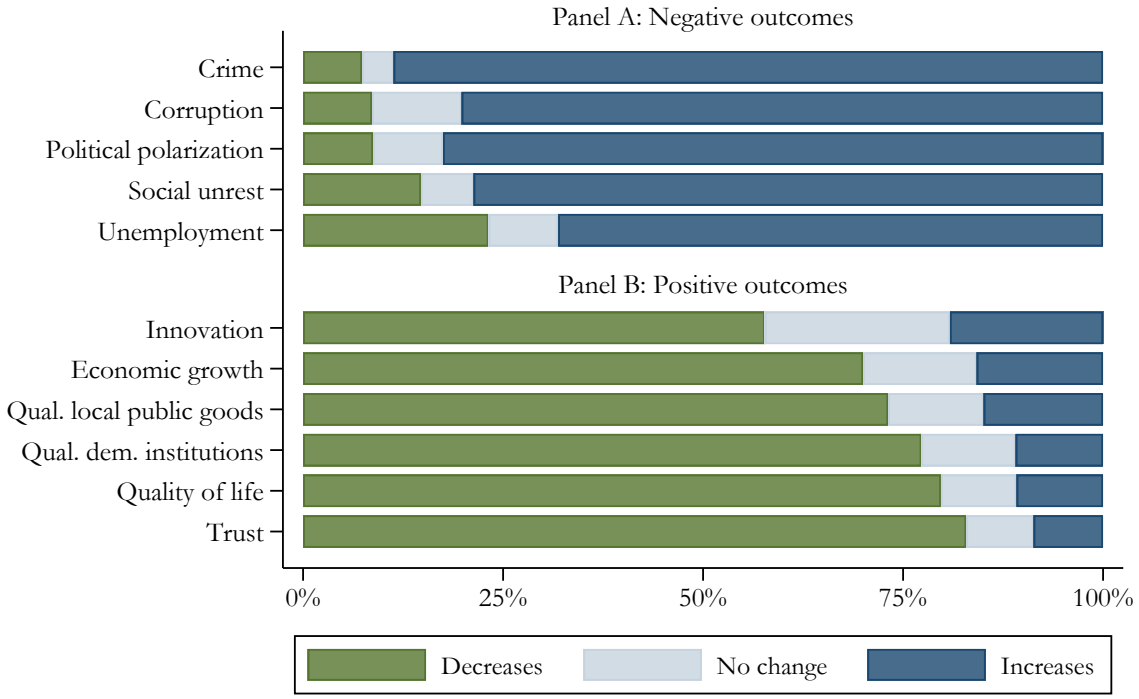
Note. Methodology as in Figure 1. Sample restricted to individuals who identify as Democrat or leaning Democrat.

Figure H6: Externality Beliefs for Republican-leaning Respondents



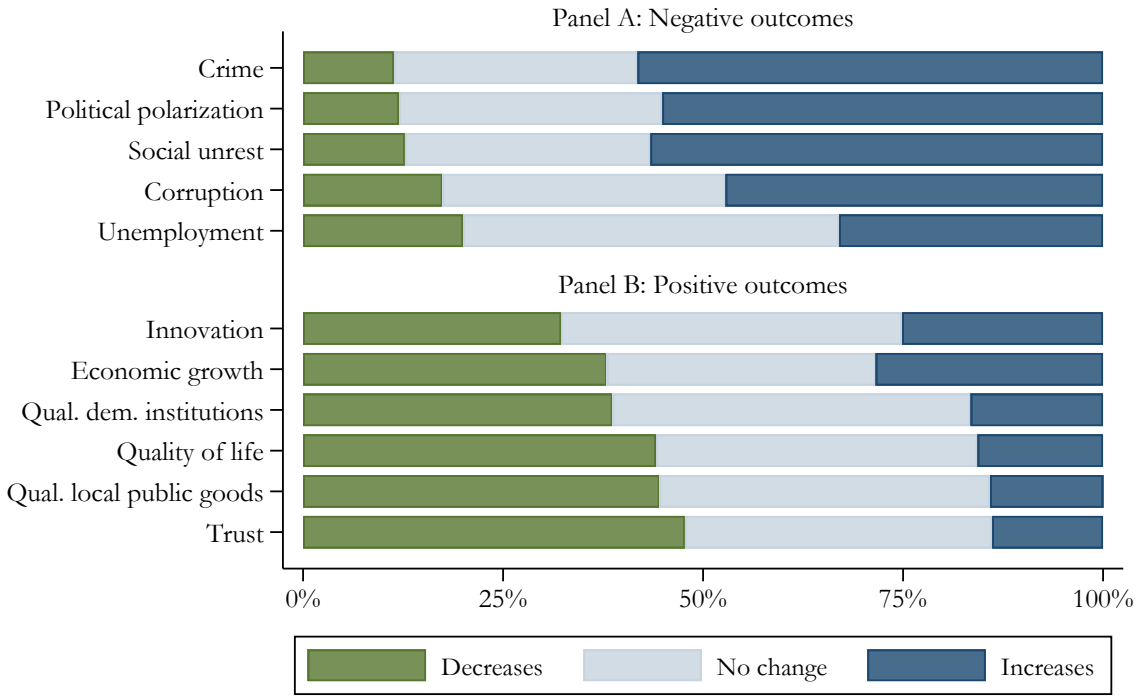
Note. Methodology as in Figure 1. Sample restricted to individuals who identify as Republican or leaning Republican.

Figure H7: Externality Beliefs for Very Liberal Respondents



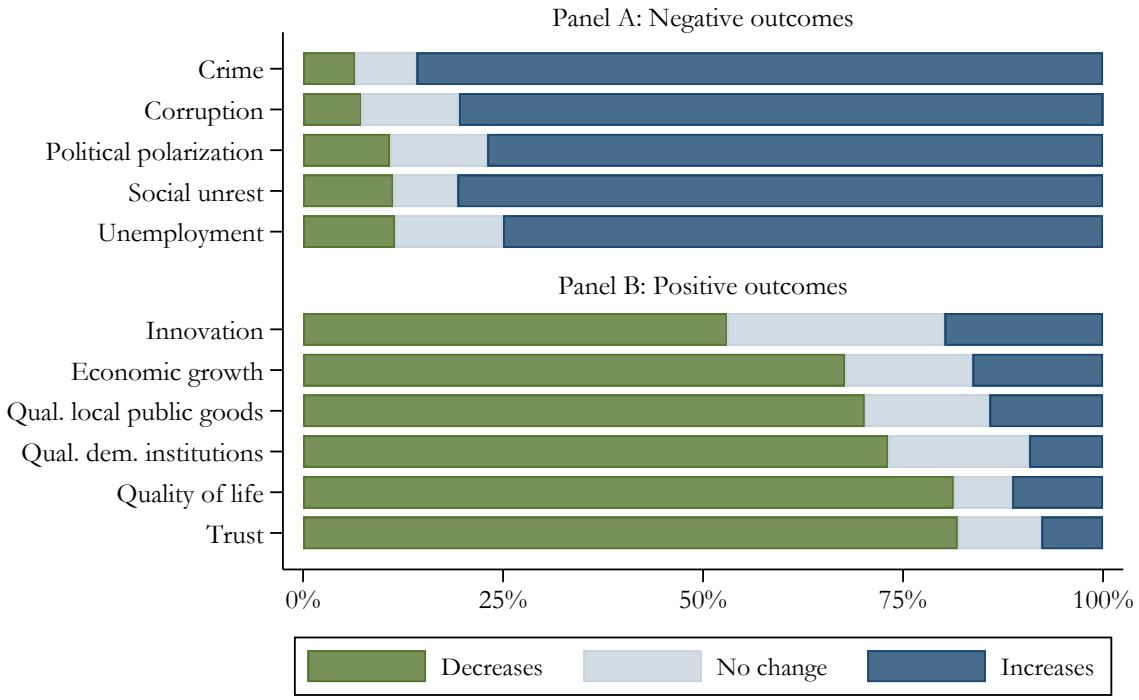
Note. Methodology as in Figure 1. Sample restricted to individuals who identify as very liberal.

Figure H8: Externality Beliefs for Very Conservative Respondents



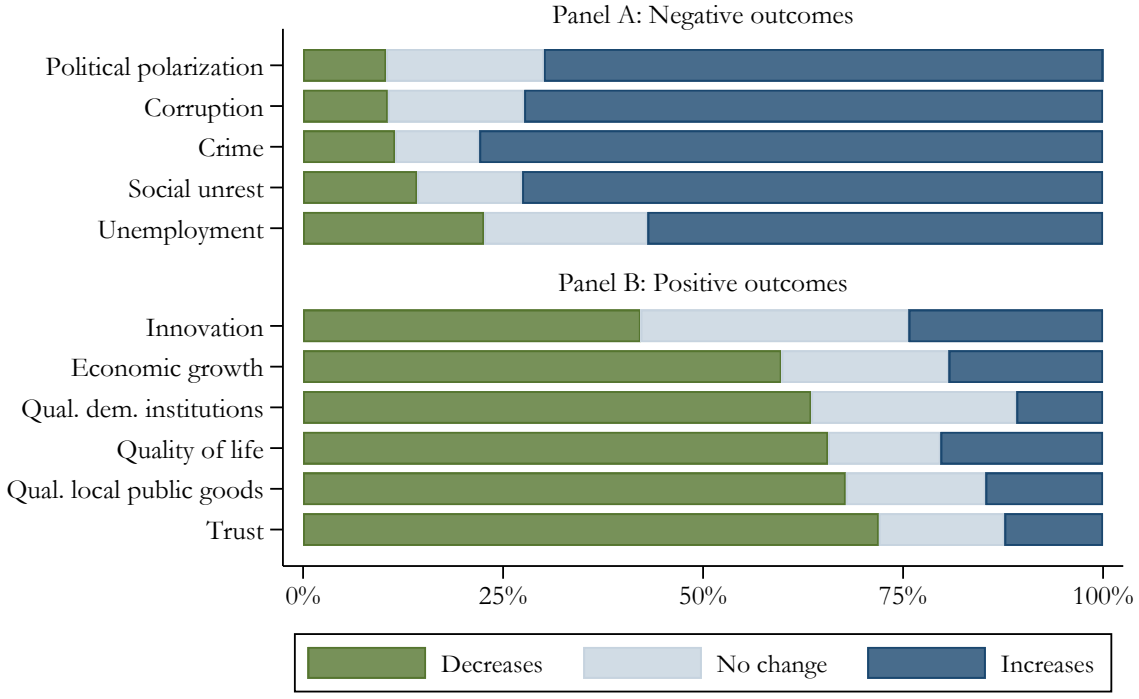
Note. Methodology as in Figure 1. Sample restricted to individuals who identify as very conservative.

Figure H9: Externality Beliefs for Sanders supporters



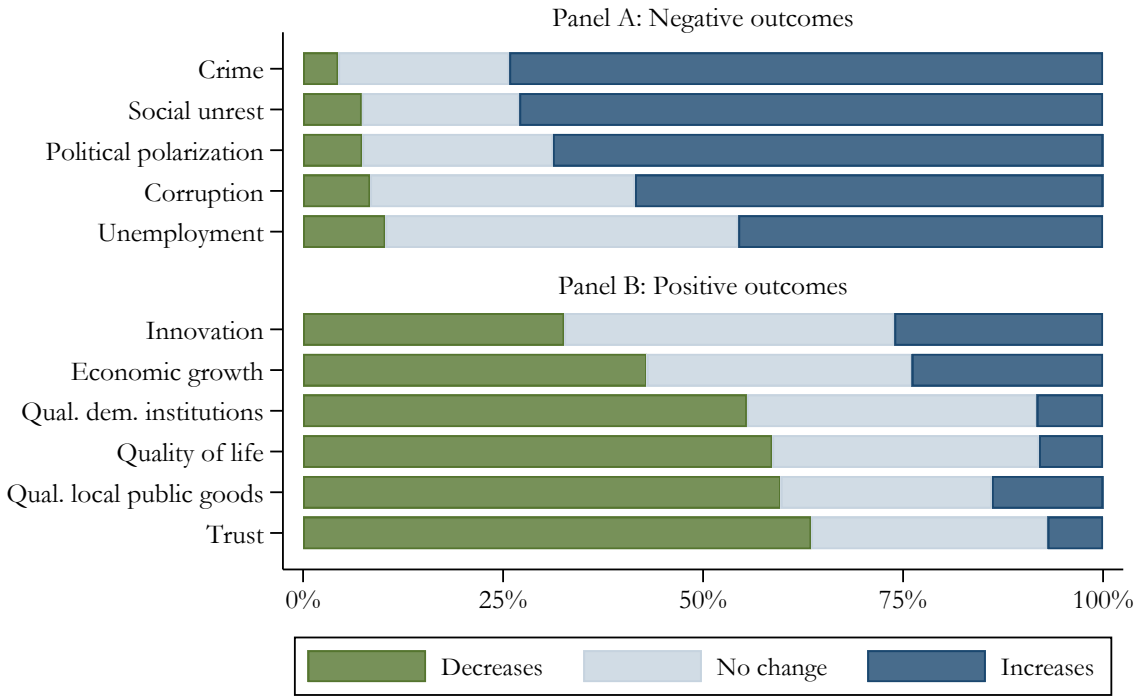
Note. Methodology as in Figure 1. Sample restricted to individuals who identify as closest to Bernie Sanders among four politicians (Donald Trump, Mitt Romney, Kamala Harris, and Bernie Sanders)

Figure H10: Externality Beliefs for Harris supporters



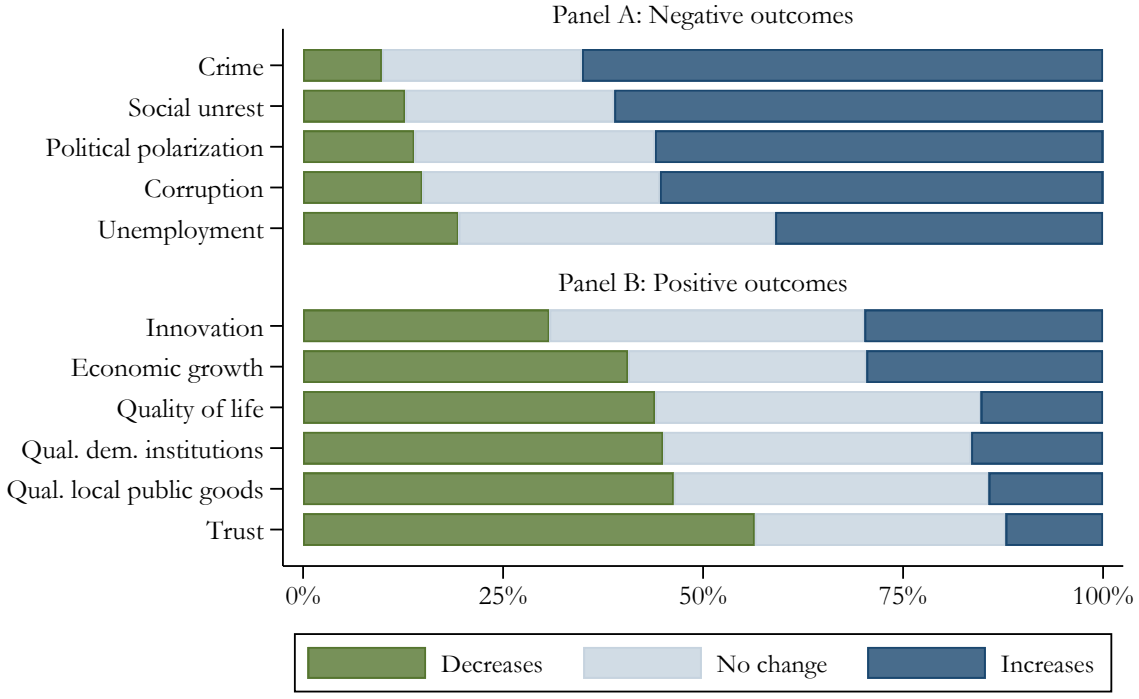
Note. Methodology as in Figure 1. Sample restricted to individuals who identify as closest to Kamala Harris among four politicians (Donald Trump, Mitt Romney, Kamala Harris, and Bernie Sanders)

Figure H11: Externality Beliefs for Romney supporters



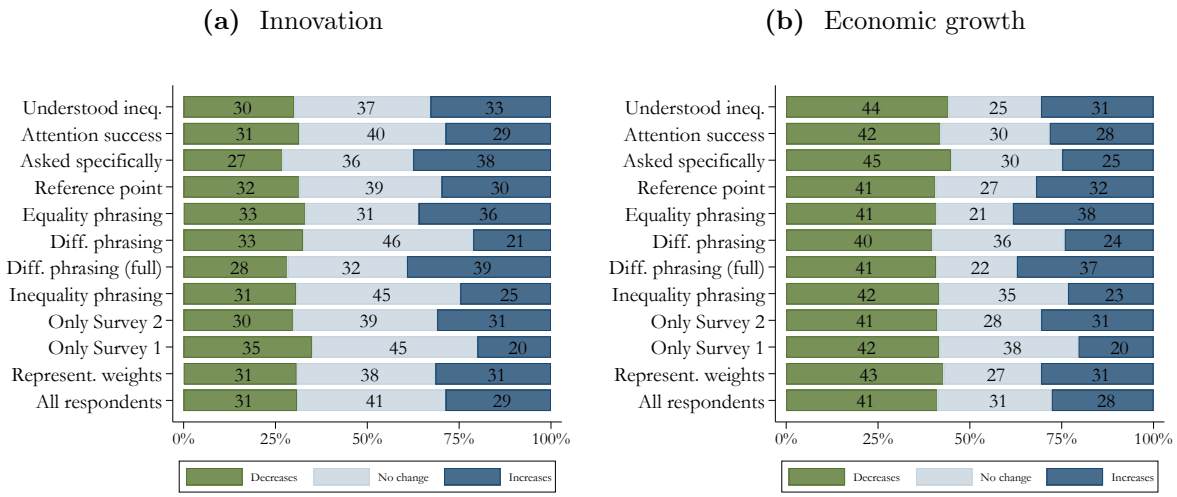
Note. Methodology as in Figure 1. Sample restricted to individuals who identify as closest to Mitt Romney among four politicians (Donald Trump, Mitt Romney, Kamala Harris, and Bernie Sanders)

Figure H12: Externality Beliefs for Trump supporters



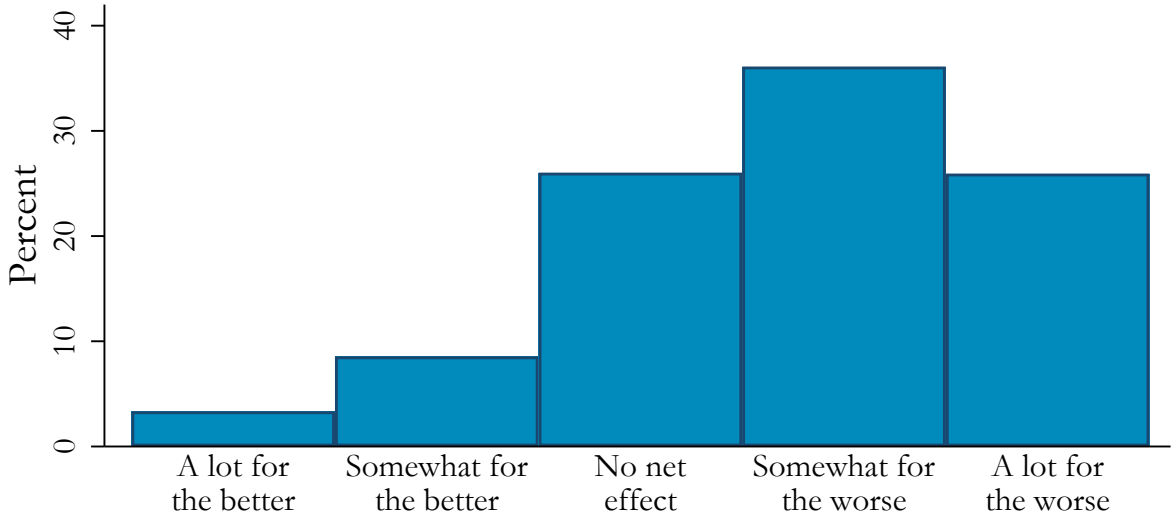
Note. Methodology as in Figure 1. Sample restricted to individuals who identify as closest to Donald Trump among four politicians (Donald Trump, Mitt Romney, Kamala Harris, and Bernie Sanders)

Figure H13: Robustness: Innovation and economic growth for Republican-leaning respondents



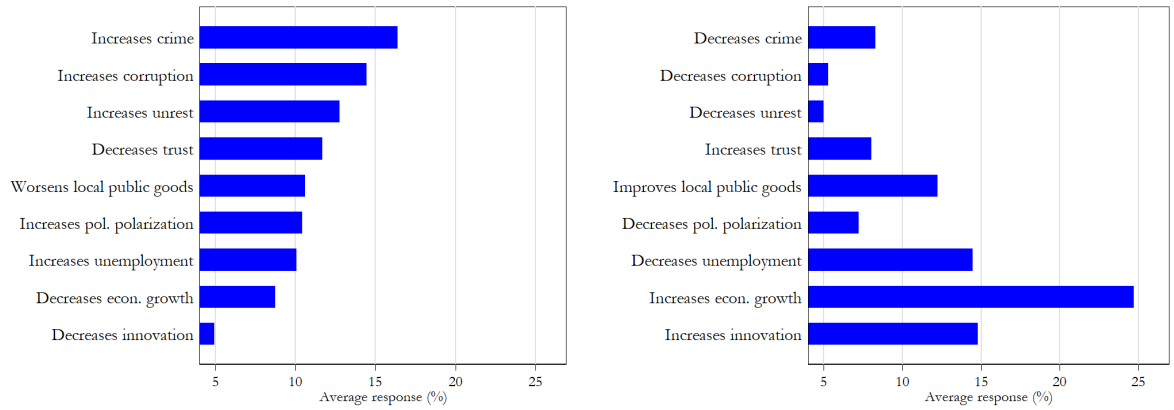
Note. Answers to *How does more economic inequality change the amount of [outcome] in a country?* for innovation and economic growth among Republican-leaning respondents across various specifications. The result that negative externality beliefs are more commonly held than positive externality beliefs among Republican-leaning respondents holds for economic growth in every robustness test. In the case of innovation, the same result does not hold for those who answer a question on understanding of distributional concerns correctly, those who are specifically asked to think through their answer and given time to do so (sample size $N < 125$), the "equality" phrasing, and the "differences in income and wealth" phrasing. Note that "No change" is often the highest-scoring value for innovation. The different specifications are, from bottom to top, only including Republican-identified respondents; (i) All respondents and all phrasings in both surveys, (ii) All respondents weighted such that the full sample to have full representativity on age, gender, race, college attendance, income, region, and party affiliation, (iii) Only respondents from Survey 1, (iv) Only respondents from the Survey 2, (v) Only respondents who saw an "inequality" phrasing in either survey, (vi) Only respondents who saw a "differences in income and wealth" phrasing for the full survey (only in Survey 2), (vii) Only respondents who saw a "differences in income and wealth" phrasing for this question, but were generally asked about inequality otherwise (only in Survey 1), (viii) Only respondents who saw an "equality" phrasing for the full survey (only in Survey 2), (ix) Only respondents who were explicitly told the reference point of inequality and the magnitude of inequality change (only in Survey 2), (x) Only respondents who were explicitly asked to think through their answer and were given 15 seconds to do so, then asked to confirm their answer or change it, if they wished (only in Survey 2, and at the end of the survey) (xi) All respondents restricted to those who succeeded on every attention check, (xii) All respondents from Survey 2 who correctly answered a simple question on distributional concepts. Sample sizes in all cases are slightly less than one-third of those in Figure H3-H4. Treatment groups from Survey 1 are always excluded.

Figure H14: *Do you think more economic inequality changes society for the better or for the worse?*



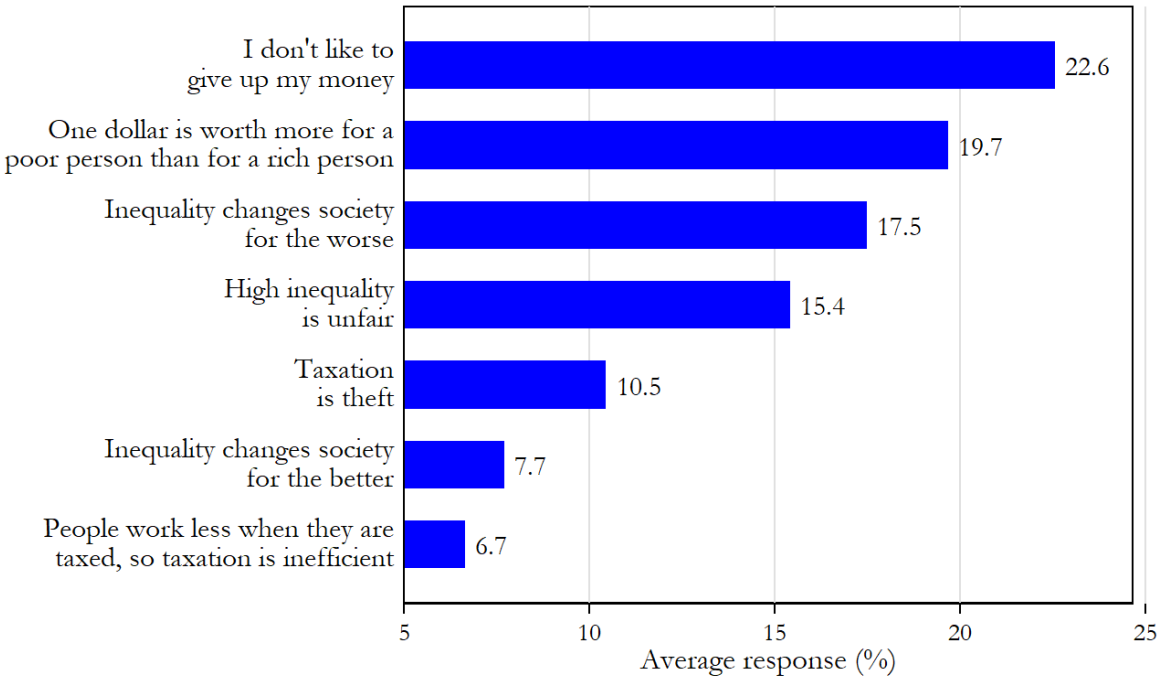
Note. Question text from Survey 1: "*Generally speaking, do you think more economic inequality changes society for the better or for the worse?*". $N = 3,292$ across Survey 1 and Survey 2 (pooled sample). Question text differs across surveys; see Appendix II.C.5 for more details. The accompanying data is shown in Table C2.

Figure H15: Comparative Magnitudes of Externality Channels



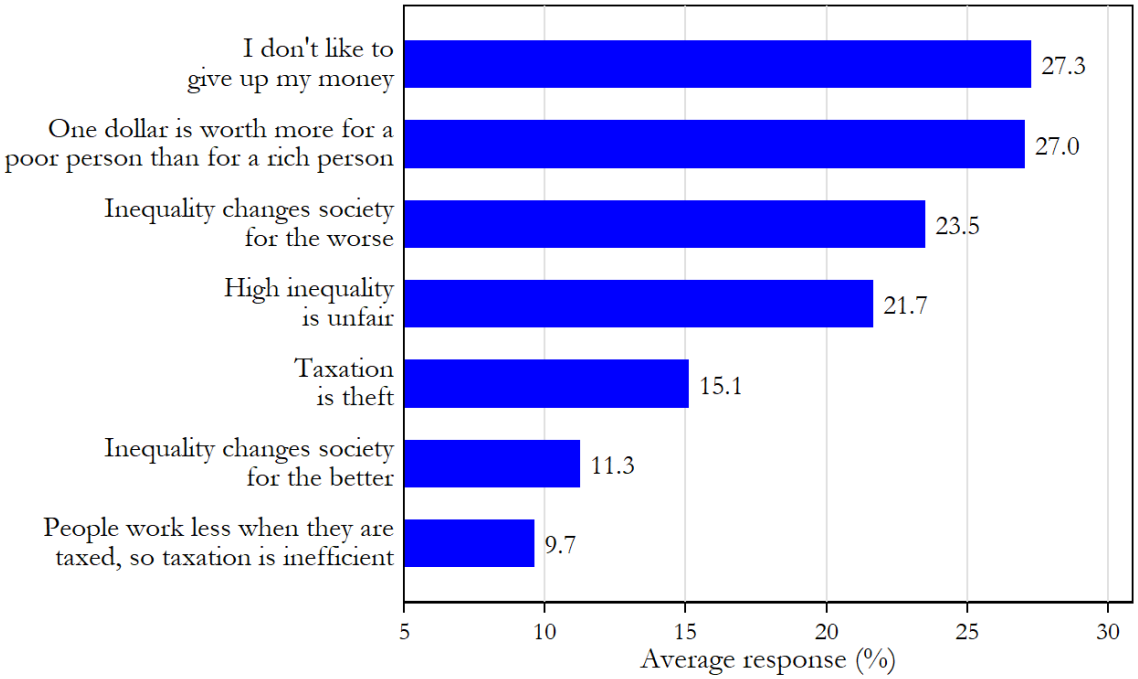
Note. These questions were only asked to those in the control groups who also (i) answered that inequality is a negative (left) or positive (right) externality, and (ii) did not answer that they changed their mind when posed this question. Sample size is $n = 472$ (left) and $n = 100$ (right).

Figure H16: Mean share for each motive behind preferences for redistribution in Survey 2



Question text: *When thinking about your preferred level of redistribution, what matters most to you? Please indicate what dimensions matter by giving scores below that add up to 100.* Answer option texts are identical to graph labels. Standard errors are approximately 0.5%.

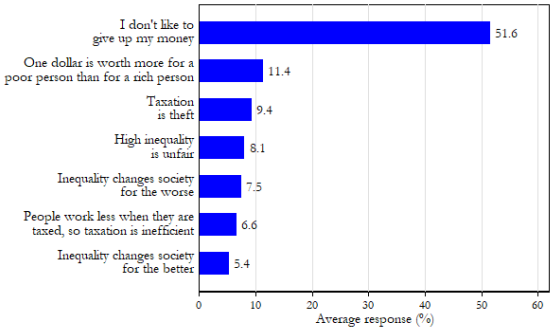
Figure H17: Share of subjects that rank a given motive first



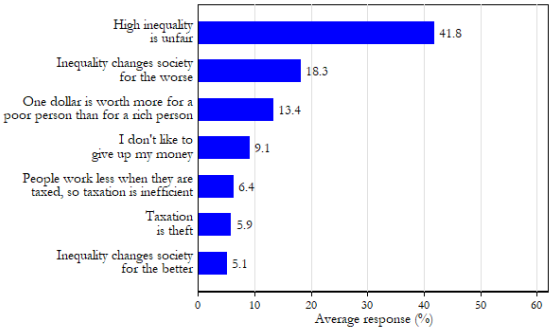
Note. Share of respondents who ranked the given motive weakly first. When a respondent ranked several motives equally, all are counted (which means the total percentage is above 100%). Question text: *When thinking about your preferred level of redistribution, what matters most to you? Please indicate what dimensions matter by giving scores below that add up to 100.* Answer option texts are identical to graph labels.

Figure H18: Share of points going to each motive conditional on the given motive attaining the highest share of points

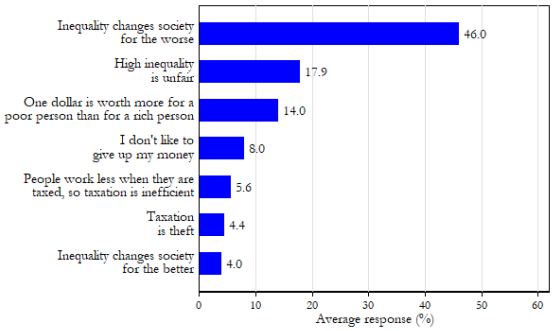
(a) Selfish ranked first



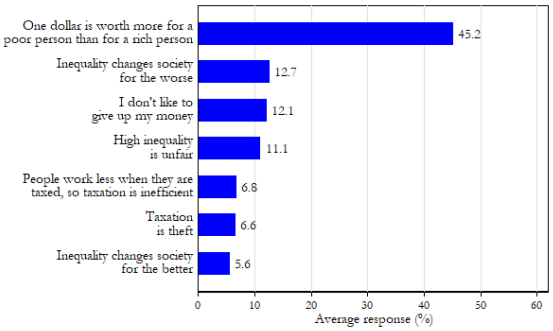
(b) Fairness concerns ranked first



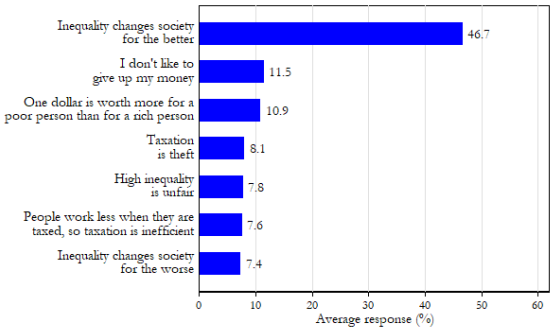
(c) Negative externality concerns ranked first



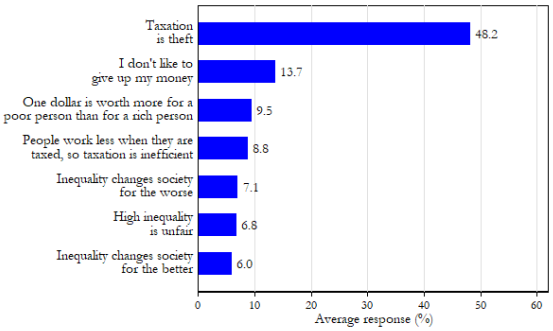
(d) DMU ranked first



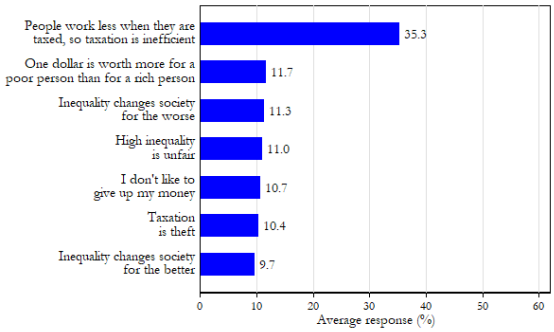
(e) Positive externality concerns ranked first



(f) Dislike taxes ranked first

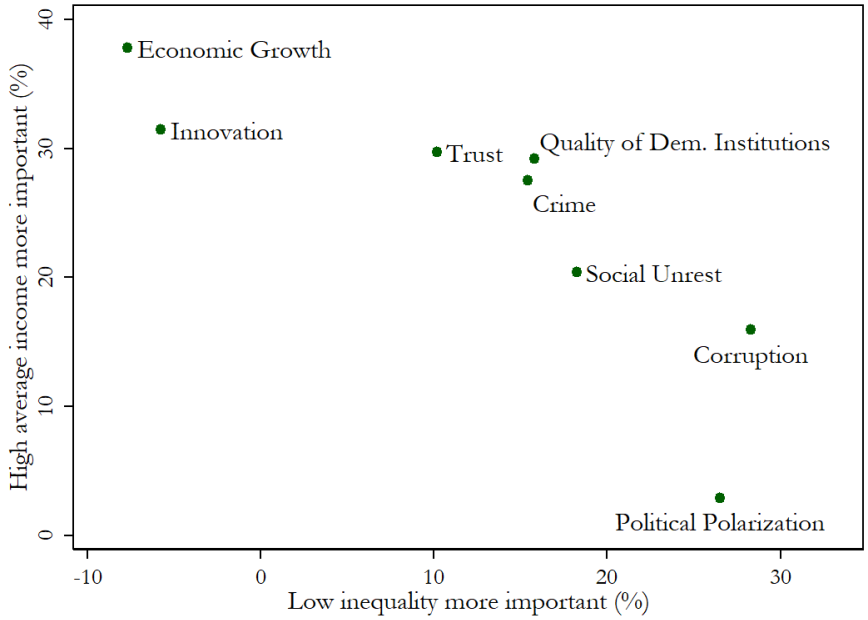


(g) Efficiency concerns ranked first



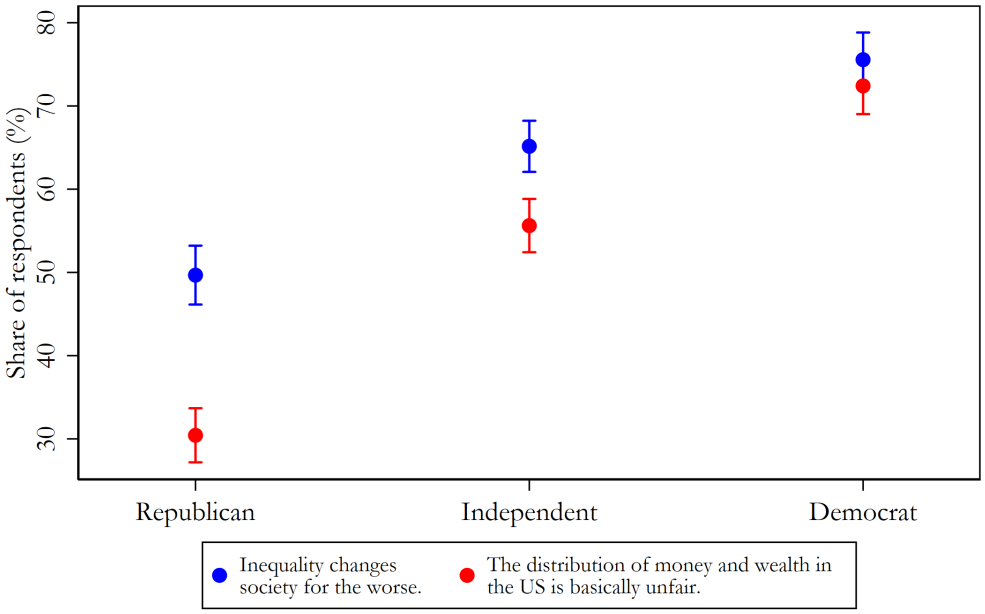
Question text: *When thinking about your preferred level of redistribution, what matters most to you? Please indicate what dimensions matter by giving scores below that add up to 100.* Answer option texts are identical to graph labels.

Figure H19: Predictive power of average income and economic inequality in explaining respondents' predicted outcomes



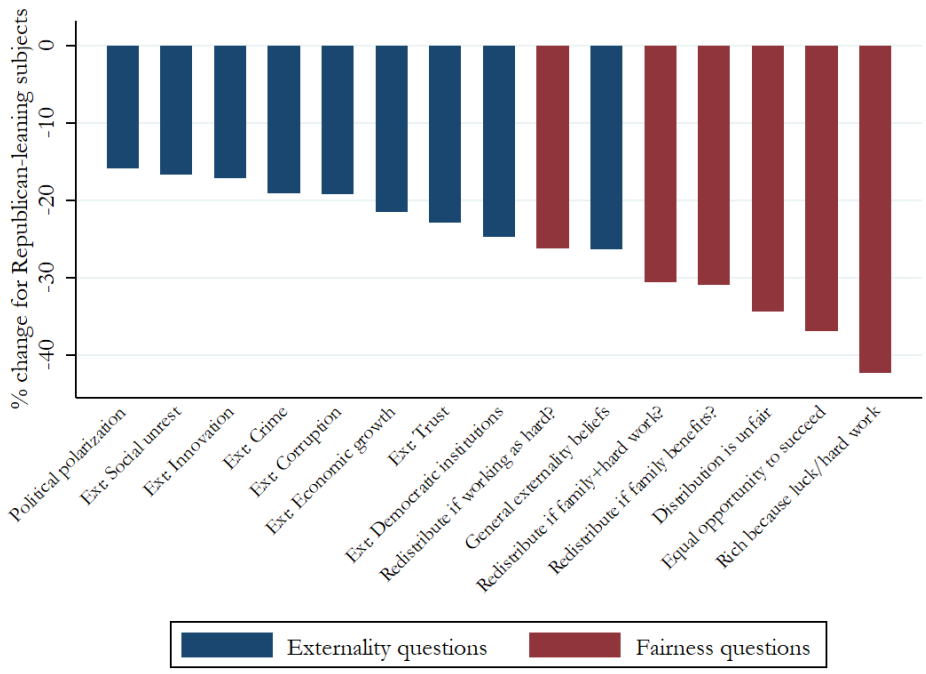
Note. This graph is based on a set of questions of the type; *Do you think an average society with a high level of average income and a low level of economic inequality would have (compared to other countries): A [very low/low/average/high/very high] level of [output].* Each respondent in Survey 2 was shown one of these questions for the output they were specifically asked to evaluate; the level of inequality and average income was randomized. The graph illustrates the increased share of respondents who answer a negative outcome (e.g. "a high" or "a very high" level of crime) when told that the average income is [low/high] (the y-axis) or that the level of income inequality is [low/high] (the x-axis). The graph thus shows how the level of average income or inequality affects the prediction of respondents for the outcome. Each data point is elicited from ~ 200 respondents. This means that approximately 50 respondents received each type of question (where a type of question indicates an outcome, an income level, and an inequality level – for example crime with low inequality and high average income). Standard errors roughly are 0.5%.

Figure H20: Externality Beliefs and Fairness Views over Party Affiliation (Survey 2)



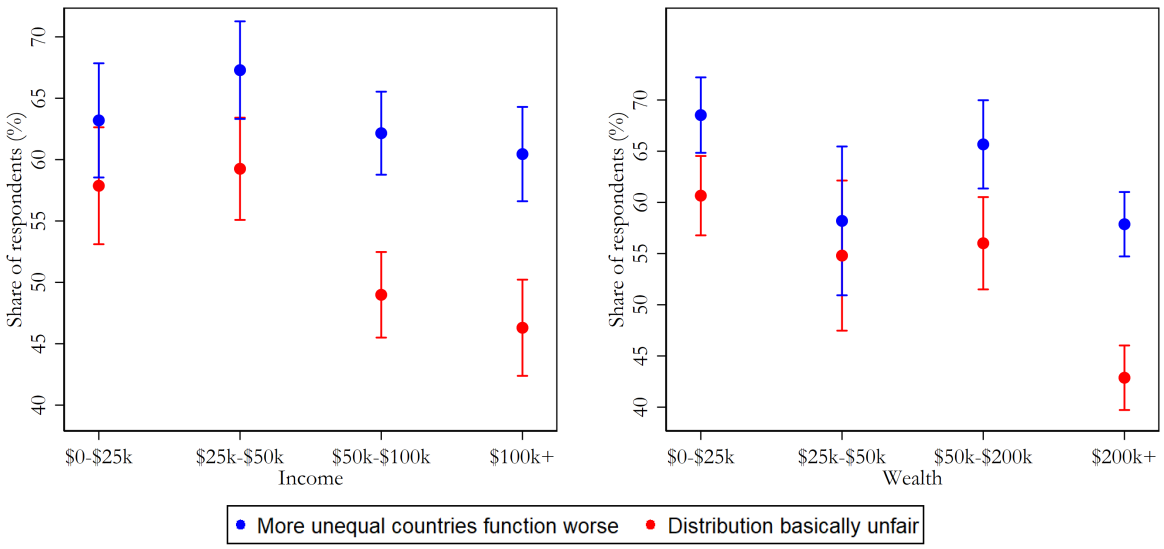
Note. This graph uses the pre-treatment externality and fairness question with Survey 2 respondents (n=2,360). Respondents are asked to agree or disagree with the following two statements: "*The distribution of money and wealth in the US is basically fair, because everybody has an equal opportunity to succeed*" and answer the question "*Does economic inequality change society for the better or the worse?*". The equivalent graph for Survey 1 respondents is Figure 9

Figure H21: Party affiliation polarization across questions (Survey 2)



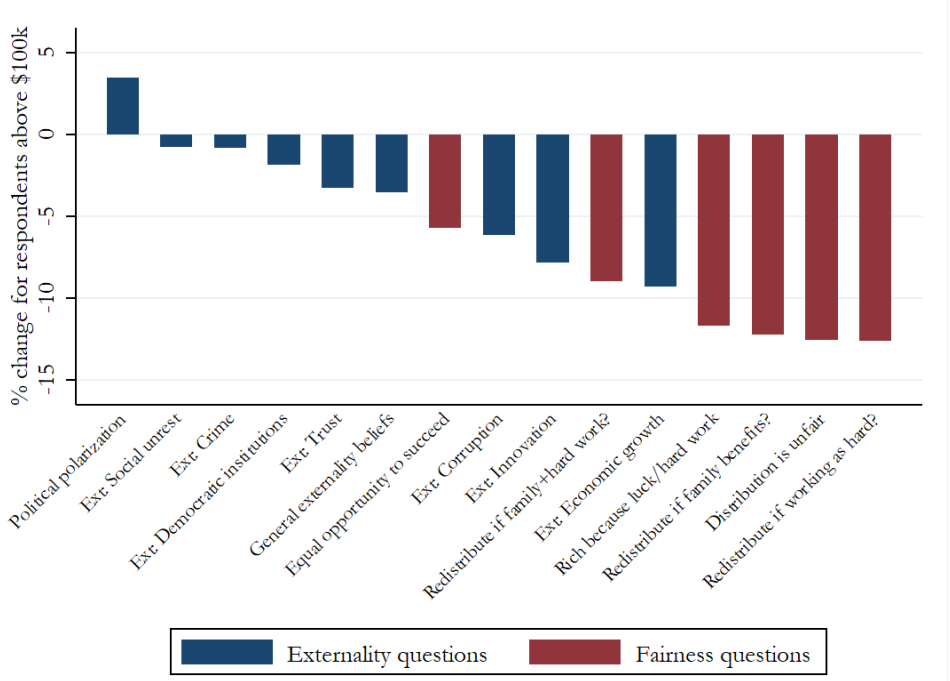
Note. Drop in anti-inequality sentiment for Republican-leaning respondents across every externality and fairness question in Survey 2 without any controls. With a standard set of controls the same relation holds (all fairness questions have larger polarization than any externality question). Questions are largely split on pre-specified criteria or natural binary points (e.g. agree/disagree), keeping total shares close to 50% where possible.

Figure H22: Externality Beliefs and Fairness Views over Income and Wealth (Survey 2)



Note. This graph uses the pre-treatment externality and fairness question with Survey 2 respondents (n=2,360). Respondents are asked to agree or disagree with the following two statements: "*The distribution of money and wealth in the US is basically fair, because everybody has an equal opportunity to succeed*" and answer the question "*Does economic inequality change society for the better or the worse?*". For the equivalent graph from Survey 1, see Figure 7.

Figure H23: Income polarization across questions (Survey 2)



Note. Drop in anti-inequality sentiment for respondents with incomes above $100,000 across every externality and fairness question in Survey 2 without any controls. With a standard set of controls there is more variation, although the average slope of the fairness questions stay significantly lower than that of the externality questions. Questions are largely split on pre-specified criteria or natural binary points (e.g. agree/disagree), keeping total shares close to 50% where possible.

II.I. Tables

Table I1: Distribution of Inequality Externality Beliefs

	Crime	Corr-uption	Pol. polar.	Social unrest	Unemp-loyment	Inno-vation	Econ. growth	Public goods	Quality of life	Dem. inst.	Trust
Increases	74%	66%	66%	70%	53%	26%	23%	14%	14%	12%	10%
No change	17%	23%	23%	18%	30%	35%	25%	28%	26%	30%	23%
Decreases	9%	11%	12%	12%	17%	40%	51%	58%	59%	57%	67%
Respondents	3,292	2,994	2,990	3,292	641	3,017	3,292	643	628	3,065	3,292

Note. The corresponding table to Figure 1. Shows the distribution of specific externality beliefs for the full sample (control group of Survey 1 and all of Survey 2). "Increase" is the share of respondents that state that inequality "increases a lot" or "increases somewhat" the outcome. "No change" is the share of respondents that state that inequality does not induce a change on the outcome. "Decrease" is the share of respondents that state that inequality "decreases a lot" or "decreases somewhat" the outcome. Passive control respondents were asked every question, while active control respondents were asked the crime, trust, social unrest, and economic growth questions along with a random subset of three additional questions. For the equivalent table using only data from Survey 1 or see Tables I2 and I3.

Table I2: Distribution of Externality Beliefs in Survey 1 (Control Group)

	Crime	Corr-uption	Pol. polar.	Social unrest	Unemp-loyment	Inno-vation	Econ. growth	Public goods	Quality of life	Dem. inst.	Trust
Increases	76%	69%	68%	68%	53%	22%	19%	14%	14%	12%	10%
No change	16%	20%	23%	20%	30%	36%	29%	28%	26%	32%	22%
Decreases	8%	11%	10%	12%	17%	42%	52%	58%	59%	56%	68%
Respondents	932	634	630	932	641	657	932	643	628	705	932

Note. This is the corresponding table to Table I1 for Survey 1 respondents, control group only. Shows the distribution of specific externality beliefs in Survey 1. "Increase" is the share of respondents that state that inequality "increases a lot" or "increases somewhat" the outcome. "No change" is the share of respondents that state that inequality does not induce a change on the outcome. "Decrease" is the share of respondents that state that inequality "decreases a lot" or "decreases somewhat" the societal outcome. Passive control respondents were asked every question, while active control respondents were asked the crime, trust, social unrest, and economic growth questions along with a random subset of three additional questions. For the equivalent table in Survey 2 see Table I3.

Table I3: Distribution of Externality Beliefs in Survey 2

	Crime	Corr-uption	Pol. polar.	Social unrest	Inno-vation	Econ. growth	Dem. inst.	Trust	Placebo	Attent. Incr.	Attent. Decr.
Increases	74%	65%	65%	71%	27%	25%	13%	10%	5%	95%	3%
No change	16%	24%	23%	17%	35%	23%	30%	23%	89%	2%	2%
Decreases	9%	11%	12%	12%	38%	52%	58%	67%	6%	3%	95%
Respondents	2,360	2,360	2,360	2,360	2,360	2,360	2,360	2,360	2,360	2,360	2,360

Note. This is the corresponding table to Table I1 for Survey 2 respondents. Shows the distribution of specific externality beliefs in Survey 2. "Increase" is the share of respondents that state that inequality "increases a lot" or "increases somewhat" the outcome. "No change" is the share of respondents that state that inequality does not induce a change on the outcome. "Decrease" is the share of respondents that state that inequality "decreases a lot" or "decreases somewhat" the outcome. For the equivalent table in Survey 1 see Table I2.

Table I4: Balance table for posterior externality beliefs

Variable	(1) Passive Control	(2) Active Control	(3) Difference
General neg. ext.	0.582	0.614	0.032
	(0.494)	(0.487)	(0.032)
Ineq. incr. crime	0.757	0.761	0.005
	(0.430)	(0.427)	(0.028)
Ineq. red. trust	0.669	0.698	0.029
	(0.471)	(0.460)	(0.031)
Ineq. incr. growth	0.190	0.193	0.003
	(0.392)	(0.395)	(0.026)
Society is unfair (post)	0.587	0.609	0.022
	(0.493)	(0.489)	(0.033)
Rich because of hard work	0.392	0.383	-0.009
	(0.489)	(0.487)	(0.032)
Observations	538	394	932

Note. This table represents mean (standard deviations) for posterior externality beliefs of respondents in the active (column 1) and passive (column 2) control groups. Column (3) characterizes the difference across the two. *Significance levels*: *10%, **5%, ***1%.

Table I5: Balance table for prior views and values

Variable	(1) Passive Control	(2) Active Control	(3) Difference
Prior belief fair	0.481	0.492	0.011
	(0.500)	(0.501)	(0.033)
Belief uneq countr. worse.	0.584	0.617	0.033
	(0.493)	(0.487)	(0.032)
Trusts the government	0.288	0.327	0.039
	(0.453)	(0.470)	(0.031)
Belief work less if tax	0.400	0.376	-0.024
	(0.490)	(0.485)	(0.032)
Observations	538	394	932

Note. This table represents mean (standard deviations) for posterior fairness views of respondents in the active (column 1) and passive (column 2) control groups. Column (3) characterizes the difference across the two. *Significance levels*: *10%, **5%, ***1%.

Table I6: Balance table for observable characteristics

Variable	(1) Passive Control	(2) Active Control	(3) Difference
Leans Republican	0.532	0.492	-0.039
	(0.499)	(0.501)	(0.033)
Prior belief unfair	0.519	0.508	-0.011
	(0.500)	(0.501)	(0.033)
Trusts the government	0.288	0.327	0.039
	(0.453)	(0.470)	(0.031)
Male	0.498	0.495	-0.003
	(0.500)	(0.501)	(0.033)
Black	0.087	0.081	-0.006
	(0.283)	(0.274)	(0.018)
Neither black or white	0.162	0.107	-0.055**
	(0.369)	(0.309)	(0.022)
Income: 0-25k	0.214	0.236	0.022
	(0.410)	(0.425)	(0.028)
Income: 25-50k	0.331	0.249	-0.082***
	(0.471)	(0.433)	(0.030)
Income: 50-100k	0.257	0.312	0.056*
	(0.437)	(0.464)	(0.030)
Income: 100k and more	0.199	0.203	0.004
	(0.400)	(0.403)	(0.027)
Age 30-39	0.164	0.188	0.024
	(0.370)	(0.391)	(0.025)
Age 40-49	0.182	0.150	-0.032
	(0.386)	(0.357)	(0.025)
Age 50-59	0.128	0.147	0.019
	(0.335)	(0.355)	(0.023)
Age 60-69	0.175	0.162	-0.012
	(0.380)	(0.369)	(0.025)
Age 70 and above	0.206	0.223	0.017
	(0.405)	(0.417)	(0.027)
4-year college degree or more	0.459	0.513	0.054
	(0.499)	(0.500)	(0.033)
Unemployed	0.099	0.107	0.008
	(0.298)	(0.309)	(0.020)
Outside the labor force	0.457	0.431	-0.026
	(0.499)	(0.496)	(0.033)
West	0.258	0.206	-0.053*
	(0.438)	(0.405)	(0.028)
North-East	0.138	0.190	0.053**
	(0.345)	(0.393)	(0.025)
Midwest	0.238	0.228	-0.009
	(0.426)	(0.420)	(0.028)
Observations	538	394	932

Note. This table represents mean (standard deviations) for sociodemographic variables of respondents in the active (column 1) and passive (column 2) control groups. Column (3) characterizes the difference across the two. *Significance levels*: *10%, **5%, ***1%.

Table I7: Balance table Crime vs. Control

	(1)	(2)	(3)
Variable	Control	Crime	Difference
Leans Republican	0.515	0.525	0.010
	(0.500)	(0.500)	(0.023)
Prior belief unfair	0.514	0.529	0.016
	(0.500)	(0.499)	(0.023)
Trusts the government	0.305	0.285	-0.020
	(0.461)	(0.452)	(0.021)
Male	0.497	0.466	-0.031
	(0.500)	(0.499)	(0.023)
Black	0.085	0.095	0.011
	(0.279)	(0.294)	(0.013)
Neither black or white	0.138	0.128	-0.011
	(0.346)	(0.334)	(0.016)
Income: 0-25k	0.223	0.235	0.012
	(0.417)	(0.424)	(0.019)
Income: 25-50k	0.296	0.267	-0.029
	(0.457)	(0.443)	(0.021)
Income: 50-100k	0.280	0.307	0.026
	(0.449)	(0.461)	(0.021)
Income: 100k and more	0.201	0.192	-0.009
	(0.401)	(0.394)	(0.018)
Age 30-39	0.174	0.158	-0.016
	(0.379)	(0.365)	(0.017)
Age 40-49	0.168	0.166	-0.002
	(0.374)	(0.372)	(0.017)
Age 50-59	0.136	0.144	0.007
	(0.343)	(0.351)	(0.016)
Age 60-69	0.170	0.182	0.013
	(0.375)	(0.386)	(0.018)
Age 70 and above	0.214	0.211	-0.002
	(0.410)	(0.408)	(0.019)
4-year college degree or more	0.482	0.498	0.017
	(0.500)	(0.500)	(0.023)
Unemployed	0.102	0.093	-0.009
	(0.303)	(0.291)	(0.014)
Outside the labor force	0.446	0.426	-0.021
	(0.497)	(0.495)	(0.023)
West	0.236	0.269	0.033
	(0.425)	(0.444)	(0.020)
North-East	0.160	0.166	0.006
	(0.367)	(0.372)	(0.017)
Midwest	0.234	0.175	-0.059***
	(0.424)	(0.380)	(0.019)
Prior belief unfair	0.514	0.529	0.016
	(0.500)	(0.499)	(0.023)
Belief work less if tax	0.389	0.372	-0.018
	(0.488)	(0.484)	(0.022)
Trusts the government	0.305	0.285	-0.020
	(0.461)	(0.452)	(0.021)
Belief pay less than prod.	0.734	0.741	0.007
	(0.442)	(0.439)	(0.020)
Belief uneq countr. worse.	0.598	0.643	0.045**
	(0.491)	(0.479)	(0.022)
Observations	932	933	1,865

Note. This table represents mean (standard deviations) for pre-treatment beliefs and characteristics in the Control (column 1) and Crime (column 2) groups. Column (3) characterizes the difference across the two. *Significance levels*: *10%, **5%, ***1%.

Table I8: Balance table Trust vs. Control

	(1)	(2)	(3)
Variable	Control	Trust	Difference
Leans Republican	0.515	0.527	0.012
	(0.500)	(0.500)	(0.024)
Prior belief unfair	0.514	0.526	0.012
	(0.500)	(0.500)	(0.024)
Trusts the government	0.305	0.325	0.020
	(0.461)	(0.469)	(0.022)
Male	0.497	0.476	-0.020
	(0.500)	(0.500)	(0.024)
Black	0.085	0.103	0.018
	(0.279)	(0.304)	(0.014)
Neither black or white	0.138	0.127	-0.011
	(0.346)	(0.333)	(0.016)
Income: 0-25k	0.223	0.227	0.003
	(0.417)	(0.419)	(0.020)
Income: 25-50k	0.296	0.320	0.024
	(0.457)	(0.467)	(0.022)
Income: 50-100k	0.280	0.282	0.002
	(0.449)	(0.450)	(0.022)
Income: 100k and more	0.201	0.171	-0.030
	(0.401)	(0.377)	(0.019)
Age 30-39	0.174	0.172	-0.002
	(0.379)	(0.378)	(0.018)
Age 40-49	0.168	0.166	-0.002
	(0.374)	(0.372)	(0.018)
Age 50-59	0.136	0.145	0.009
	(0.343)	(0.353)	(0.017)
Age 60-69	0.170	0.164	-0.006
	(0.375)	(0.370)	(0.018)
Age 70 and above	0.214	0.213	-0.000
	(0.410)	(0.410)	(0.020)
4-year college degree or more	0.482	0.468	-0.014
	(0.500)	(0.499)	(0.024)
Unemployed	0.102	0.099	-0.003
	(0.303)	(0.299)	(0.014)
Outside the labor force	0.446	0.455	0.008
	(0.497)	(0.498)	(0.024)
West	0.236	0.248	0.012
	(0.425)	(0.432)	(0.021)
North-East	0.160	0.162	0.003
	(0.367)	(0.369)	(0.018)
Midwest	0.234	0.215	-0.019
	(0.424)	(0.411)	(0.020)
Prior belief unfair	0.514	0.526	0.012
	(0.500)	(0.500)	(0.024)
Belief work less if tax	0.389	0.364	-0.026
	(0.488)	(0.481)	(0.023)
Trusts the government	0.305	0.325	0.020
	(0.461)	(0.469)	(0.022)
Belief pay less than prod.	0.734	0.772	0.038*
	(0.442)	(0.420)	(0.021)
Belief uneq countr. worse.	0.598	0.636	0.039*
	(0.491)	(0.481)	(0.023)
Observations	932	825	1,757

Note. This table represents mean (standard deviations) for pre-treatment beliefs and characteristics in the Control (column 1) and Trust (column 2) groups. Column (3) characterizes the difference across the two. *Significance levels*: *10%, **5%, ***1%.

Table I9: Balance table Full ext. vs. Control

Variable	(1) Control	(2) FullExt	(3) Difference
Leans Republican	0.515	0.507	-0.008
	(0.500)	(0.500)	(0.024)
Prior belief unfair	0.514	0.523	0.009
	(0.500)	(0.500)	(0.024)
Trusts the government	0.305	0.303	-0.002
	(0.461)	(0.460)	(0.022)
Male	0.497	0.497	0.000
	(0.500)	(0.500)	(0.024)
Black	0.085	0.091	0.007
	(0.279)	(0.288)	(0.014)
Neither black or white	0.138	0.158	0.020
	(0.346)	(0.365)	(0.017)
Income: 0-25k	0.223	0.216	-0.007
	(0.417)	(0.412)	(0.020)
Income: 25-50k	0.296	0.290	-0.006
	(0.457)	(0.454)	(0.022)
Income: 50-100k	0.280	0.335	0.055**
	(0.449)	(0.472)	(0.022)
Income: 100k and more	0.201	0.158	-0.042**
	(0.401)	(0.365)	(0.018)
Age 30-39	0.174	0.168	-0.006
	(0.379)	(0.374)	(0.018)
Age 40-49	0.168	0.180	0.012
	(0.374)	(0.385)	(0.018)
Age 50-59	0.136	0.133	-0.003
	(0.343)	(0.340)	(0.016)
Age 60-69	0.170	0.177	0.007
	(0.375)	(0.382)	(0.018)
Age 70 and above	0.214	0.188	-0.026
	(0.410)	(0.391)	(0.019)
4-year college degree or more	0.482	0.533	0.051**
	(0.500)	(0.499)	(0.024)
Unemployed	0.102	0.083	-0.019
	(0.303)	(0.276)	(0.014)
Outside the labor force	0.446	0.403	-0.043*
	(0.497)	(0.491)	(0.024)
West	0.236	0.245	0.009
	(0.425)	(0.430)	(0.021)
North-East	0.160	0.153	-0.007
	(0.367)	(0.360)	(0.017)
Midwest	0.234	0.227	-0.006
	(0.424)	(0.419)	(0.020)
Prior belief unfair	0.514	0.523	0.009
	(0.500)	(0.500)	(0.024)
Belief work less if tax	0.389	0.350	-0.040*
	(0.488)	(0.477)	(0.023)
Trusts the government	0.305	0.303	-0.002
	(0.461)	(0.460)	(0.022)
Belief pay less than prod.	0.734	0.776	0.042**
	(0.442)	(0.417)	(0.021)
Belief uneq countr. worse.	0.598	0.616	0.018
	(0.491)	(0.487)	(0.023)
Observations	932	809	1,741

Note. This table represents mean (standard deviations) for pre-treatment beliefs and characteristics in the Control (column 1) and Full Externality (column 2) groups. Column (3) characterizes the difference across the two. *Significance levels*: *10%, **5%, ***1%.

Table I10: Balance table Fairness vs. Control

Variable	(1) Control	(2) Fairness	(3) Difference
Leans Republican	0.515	0.526	0.011
	(0.500)	(0.500)	(0.024)
Prior belief unfair	0.514	0.500	-0.014
	(0.500)	(0.500)	(0.024)
Trusts the government	0.305	0.275	-0.029
	(0.461)	(0.447)	(0.021)
Male	0.497	0.540	0.043*
	(0.500)	(0.499)	(0.024)
Black	0.085	0.096	0.012
	(0.279)	(0.295)	(0.014)
Neither black or white	0.138	0.148	0.010
	(0.346)	(0.355)	(0.017)
Income: 0-25k	0.223	0.208	-0.016
	(0.417)	(0.406)	(0.019)
Income: 25-50k	0.296	0.271	-0.025
	(0.457)	(0.445)	(0.021)
Income: 50-100k	0.280	0.321	0.041*
	(0.449)	(0.467)	(0.022)
Income: 100k and more	0.201	0.201	0.000
	(0.401)	(0.401)	(0.019)
Age 30-39	0.174	0.159	-0.014
	(0.379)	(0.366)	(0.018)
Age 40-49	0.168	0.175	0.007
	(0.374)	(0.381)	(0.018)
Age 50-59	0.136	0.151	0.015
	(0.343)	(0.359)	(0.017)
Age 60-69	0.170	0.178	0.008
	(0.375)	(0.383)	(0.018)
Age 70 and above	0.214	0.206	-0.007
	(0.410)	(0.405)	(0.019)
4-year college degree or more	0.482	0.514	0.032
	(0.500)	(0.500)	(0.024)
Unemployed	0.102	0.094	-0.008
	(0.303)	(0.292)	(0.014)
Outside the labor force	0.446	0.436	-0.011
	(0.497)	(0.496)	(0.023)
West	0.236	0.221	-0.015
	(0.425)	(0.415)	(0.020)
North-East	0.160	0.156	-0.004
	(0.367)	(0.363)	(0.017)
Midwest	0.234	0.212	-0.022
	(0.424)	(0.409)	(0.020)
Prior belief unfair	0.514	0.500	-0.014
	(0.500)	(0.500)	(0.024)
Belief work less if tax	0.389	0.354	-0.035
	(0.488)	(0.479)	(0.023)
Trusts the government	0.305	0.275	-0.029
	(0.461)	(0.447)	(0.021)
Belief pay less than prod.	0.734	0.740	0.006
	(0.442)	(0.439)	(0.021)
Belief uneq countr. worse.	0.598	0.576	-0.022
	(0.491)	(0.495)	(0.023)
Observations	932	872	1,804

Note. This table represents mean (standard deviations) for pre-treatment beliefs and characteristics in the Control (column 1) and Fairness (column 2) groups. Column (3) characterizes the difference across the two. *Significance levels*: *10%, **5%, ***1%.

Table I11: Definitional text for externality questions

Externality	Additional definition
The amount of crime	*Note: When we say the amount of crime we mean the overall crime rate, including homicides, robberies, property crime and more.*
The overall level of trust	*Note: When we say the total level of trust we mean the strength of a country's social fabric. Some examples are whether most people trust others, whether people cooperate with each other, how many people return lost wallets, and so on.*
The amount of social unrest	None
The rate of economic growth	None
The amount of corruption	None
The overall amount of unemployment	None
The overall amount of innovation	None
The overall quality of life	*Note: Here we want you to compare between people with the same incomes living in more or less unequal societies.*
The overall amount of political polarization	*Note: When we say political polarization we mean to what extent people's and politicians' opinions are divided on political issues, as well as how strong these divisions are.*
The quality of democratic institutions	*Note: When we say the quality of democratic institutions we mean the capable and equitable functioning of the political system, the avoidance of abuses of power, the equality of the rule of law, whether civil liberties are respected, and so on.*
The quality of local public goods	*Note: When we say the quality of local public goods we mean the quality of things like schools, local government services, parks, youth centers and more.*

Table I12: Definitional text for externality questions, secondary study

Externality	Additional definition
The amount of crime	*Note: When we say the amount of crime we mean the overall crime rate, including homicides, robberies, property crime and more.*
The overall level of trust	*Note: When we say the total level of trust we mean the strength of a country's social fabric. Some examples are whether most people trust others, whether people cooperate with each other, and so on.*
The amount of social unrest	*Note: By social unrest we mean unconventional and sometimes violent forms of collective behavior that disrupt the typical social order in society.*
The rate of economic growth	*Note: By economic growth we mean the increase in the production of goods and services in the society.*
The amount of corruption	*Note: By corruption we mean dishonest or fraudulent acts committed by those in power, usually in the form of accepting bribes.*
The overall amount of innovation	*Note: By innovation we mean how many new technologies and products that are developed in the society.*
The overall amount of political polarization	*Note: When we say political polarization we mean the extent to which opinions are divided on political issues, both among most people and politicians, in addition to how strong these differences are and whether people with different views speak together. Increasing polarization means that there is generally less agreement in society.*
The quality of democratic institutions	*Note: When we say the quality of democratic institutions we mean the capable and equitable functioning of the political system, the avoidance of abuses of power, the equality of the rule of law, whether civil liberties are respected, and so on.*
Daylight hours (placebo)	*Note: By the number of daylight hours we mean the number of hours when the sun is visible within a country on an average day.*
Attention check #1	*Note: Here we just want you to choose the top option to show that you are reading the questions. Thank you.*
Attention check #2	*Note: Here we just want you to choose the bottom option to show that you are reading the questions. Thank you.*

Table I13: Main correlations of sociodemographic and externality beliefs

	(1)	(2)	(3)	(4)
	General neg. ext.	Ineq. incr. crime	Ineq. red. trust	Ineq. red. growth
	b/se	b/se	b/se	b/se
Leans Republican	-0.134***	-0.091***	-0.118***	-0.120***
	(0.018)	(0.017)	(0.018)	(0.020)
Prior belief unfair	0.299***	0.193***	0.239***	0.210***
	(0.018)	(0.016)	(0.017)	(0.019)
Trusts the government	0.037**	0.051***	0.040**	0.051**
	(0.018)	(0.017)	(0.018)	(0.020)
Male	-0.032*	0.002	0.007	-0.044**
	(0.017)	(0.016)	(0.017)	(0.018)
Black	-0.072**	-0.073**	-0.016	-0.022
	(0.031)	(0.029)	(0.031)	(0.034)
Neither black nor white	-0.047**	-0.044**	-0.016	0.027
	(0.022)	(0.021)	(0.022)	(0.024)
Income: 25-50k	0.041	0.036	0.020	0.003
	(0.025)	(0.023)	(0.025)	(0.027)
Income: 50-100k	0.034	0.031	0.015	-0.006
	(0.025)	(0.023)	(0.024)	(0.027)
Income: 100k and more	-0.005	0.007	-0.030	-0.079***
	(0.027)	(0.026)	(0.027)	(0.029)
Age 30-39	-0.013	-0.009	0.014	-0.017
	(0.029)	(0.028)	(0.030)	(0.032)
Age 40-49	-0.026	-0.002	0.038	0.035
	(0.031)	(0.029)	(0.030)	(0.033)
Age 50-59	0.016	0.003	0.043	0.058*
	(0.031)	(0.029)	(0.031)	(0.033)
Age 60-69	-0.053*	-0.021	-0.004	-0.006
	(0.031)	(0.030)	(0.031)	(0.034)
Age 70 and above	-0.052	-0.036	0.027	-0.003
	(0.035)	(0.032)	(0.035)	(0.037)
4-year college degree or more	0.045**	0.050***	0.042**	0.010
	(0.018)	(0.017)	(0.018)	(0.019)
Unemployed	-0.039	-0.040	0.002	-0.066*
	(0.031)	(0.031)	(0.031)	(0.034)
Outside the labor force	-0.000	0.000	-0.013	-0.019
	(0.020)	(0.019)	(0.020)	(0.021)
West	0.041**	0.043**	0.046**	0.015
	(0.020)	(0.019)	(0.020)	(0.021)
North-East	0.032	0.030	-0.000	0.007
	(0.024)	(0.023)	(0.025)	(0.026)
Midwest	-0.023	0.014	0.022	-0.036
	(0.022)	(0.021)	(0.022)	(0.023)
Constant	0.527***	0.634***	0.539***	0.497***
	(0.038)	(0.035)	(0.038)	(0.040)
Adjusted R2	0.160	0.087	0.111	0.094
Observations	3292	3292	3292	3292

Note. This table reports results from regressions that regress externality beliefs on sociodemographic variables. Sample is composed of Survey 1 control group respondents and Survey 2 respondents. Standard errors are in parentheses. *Significance levels*: *10%, **5%, ***1%.

Table I14: Correlations of sociodemographic and externality beliefs, 2

	(1)	(2)	(3)	(4)
	Ineq. red. inno.	Ineq. incr. unrest	Ineq. worsens dem. inst.	Ineq. worsens public goods
	b/se	b/se	b/se	b/se
Leans Republican	-0.093***	-0.082***	-0.159***	-0.098**
	(0.020)	(0.018)	(0.020)	(0.044)
Prior belief unfair	0.177***	0.211***	0.224***	0.200***
	(0.019)	(0.017)	(0.019)	(0.041)
Trusts the government	0.028	0.062***	0.026	-0.035
	(0.022)	(0.018)	(0.020)	(0.044)
Male	-0.007	-0.010	0.022	-0.059
	(0.019)	(0.016)	(0.018)	(0.042)
Black	-0.081**	-0.082***	-0.071**	0.045
	(0.035)	(0.031)	(0.034)	(0.074)
Neither black nor white	0.014	-0.061***	-0.034	0.073
	(0.025)	(0.022)	(0.024)	(0.055)
Income: 25-50k	-0.021	0.008	0.009	-0.004
	(0.028)	(0.024)	(0.027)	(0.053)
Income: 50-100k	-0.016	0.034	0.006	-0.012
	(0.028)	(0.024)	(0.027)	(0.057)
Income: 100k and more	-0.086***	-0.007	-0.027	-0.088
	(0.030)	(0.026)	(0.030)	(0.066)
Age 30-39	-0.026	-0.063**	0.024	-0.066
	(0.034)	(0.029)	(0.033)	(0.070)
Age 40-49	0.019	0.009	0.032	0.015
	(0.035)	(0.030)	(0.033)	(0.069)
Age 50-59	-0.053	0.006	0.063*	0.004
	(0.035)	(0.030)	(0.034)	(0.078)
Age 60-69	-0.051	-0.002	0.040	0.004
	(0.035)	(0.030)	(0.034)	(0.074)
Age 70 and above	-0.050	0.042	0.007	0.028
	(0.039)	(0.033)	(0.038)	(0.076)
4-year college degree or more	0.021	0.068***	0.055***	0.094**
	(0.019)	(0.017)	(0.019)	(0.042)
Unemployed	0.028	-0.045	0.007	-0.130*
	(0.036)	(0.032)	(0.035)	(0.070)
Outside the labor force	-0.018	-0.020	0.002	0.022
	(0.022)	(0.020)	(0.022)	(0.049)
West	0.010	0.064***	0.020	0.013
	(0.022)	(0.020)	(0.022)	(0.050)
North-East	-0.001	0.039	-0.006	-0.028
	(0.027)	(0.024)	(0.027)	(0.060)
Midwest	-0.038	0.030	-0.040*	-0.057
	(0.024)	(0.022)	(0.024)	(0.052)
Constant	0.407***	0.571***	0.478***	0.548***
	(0.043)	(0.037)	(0.041)	(0.084)
Adjusted R2	0.067	0.095	0.105	0.070
Observations	3017	3292	3065	643

Note. This table reports results from regressions that regress externality beliefs on sociodemographic variables. Sample is composed of Survey 1 control group respondents and Survey 2 respondents. Standard errors are in parentheses. *Significance levels*: *10%, **5%, ***1%.

Table I15: Correlations of sociodemographic and externality beliefs, 3

	(1)	(2)	(3)	(4)
	Ineq. incr. corruption	Ineq. incr. pol. pol.	Ineq. incr. unemp.	Ineq. decr. QoL
	b/se	b/se	b/se	b/se
Leans Republican	-0.097***	-0.103***	-0.172***	-0.202***
	(0.019)	(0.020)	(0.044)	(0.044)
Prior belief unfair	0.216***	0.159***	0.213***	0.242***
	(0.019)	(0.019)	(0.041)	(0.041)
Trusts the government	0.023	0.032*	0.020	0.063
	(0.020)	(0.020)	(0.046)	(0.043)
Male	-0.019	0.032*	-0.013	0.041
	(0.018)	(0.018)	(0.041)	(0.040)
Black	-0.057*	-0.130***	0.032	-0.140*
	(0.032)	(0.034)	(0.078)	(0.075)
Neither black nor white	0.012	-0.019	-0.029	-0.018
	(0.023)	(0.023)	(0.055)	(0.056)
Income: 25-50k	-0.017	-0.005	0.007	0.024
	(0.026)	(0.027)	(0.054)	(0.054)
Income: 50-100k	-0.034	0.041	-0.009	0.010
	(0.026)	(0.027)	(0.060)	(0.058)
Income: 100k and more	-0.085***	0.024	-0.039	0.004
	(0.029)	(0.029)	(0.068)	(0.067)
Age 30-39	-0.041	-0.035	0.008	0.103
	(0.032)	(0.033)	(0.070)	(0.072)
Age 40-49	-0.013	-0.008	0.027	0.113
	(0.032)	(0.033)	(0.068)	(0.071)
Age 50-59	0.027	0.035	0.005	0.037
	(0.032)	(0.034)	(0.073)	(0.082)
Age 60-69	-0.038	0.022	-0.074	0.097
	(0.034)	(0.034)	(0.072)	(0.074)
Age 70 and above	-0.026	0.027	-0.005	0.048
	(0.037)	(0.037)	(0.075)	(0.080)
4-year college degree or more	0.040**	0.075***	-0.028	0.034
	(0.019)	(0.019)	(0.043)	(0.042)
Unemployed	-0.075**	0.003	0.029	0.089
	(0.034)	(0.035)	(0.070)	(0.072)
Outside the labor force	-0.045**	-0.003	0.027	0.070
	(0.022)	(0.021)	(0.050)	(0.049)
West	0.025	0.064***	0.111**	-0.004
	(0.021)	(0.022)	(0.049)	(0.050)
North-East	0.017	-0.010	-0.020	-0.059
	(0.026)	(0.027)	(0.064)	(0.057)
Midwest	-0.039*	0.013	0.051	0.001
	(0.024)	(0.024)	(0.051)	(0.050)
Constant	0.656***	0.532***	0.485***	0.419***
	(0.040)	(0.041)	(0.084)	(0.088)
Adjusted R2	0.092	0.069	0.095	0.128
Observations	2994	2990	641	628

Note. This table reports results from regressions that regress externality beliefs on sociodemographic variables. Sample is composed of Survey 1 control group respondents and Survey 2 respondents. Standard errors are in parentheses. *Significance levels*: *10%, **5%, ***1%.

Table I16: Non-monotonic beliefs

	Crime	Trust	Social unrest	Pol. polar.	Corr uption	Dem. inst.	Inno- vation	Econ. growth
Yes (*monotonic*)	87%	82%	82%	83%	80%	79%	75%	76%
No (*non-monotonic*)	13%	18%	18%	17%	20%	21%	25%	24%
Sample size	244	242	239	236	243	236	228	222

Note. The share of respondents who think their expressed inequality externality is non-monotonic in inequality level. Full question text: *In the earlier question you answered that "[Answer]". Do you think this is true in any kind of country – no matter whether the country is initially very equal, very unequal, or anything else?*

Table I17: Treatment effects without controls

	(1)	(2)	(3)	(4)	(5)
	RP Index	Wants redistribution	Increase top taxes	Gov. reduce ineq.	Ineq. is serious issue
	b/se	b/se	b/se	b/se	b/se
Crime Ext. Tr.	0.036	0.031	-0.006	0.005	0.022
	(0.046)	(0.022)	(0.023)	(0.023)	(0.023)
Trust Ext. Tr.	0.055	0.010	0.005	0.041*	0.023
	(0.047)	(0.023)	(0.024)	(0.024)	(0.024)
Full Ext. Tr.	0.124***	0.059**	-0.014	0.056**	0.078***
	(0.048)	(0.023)	(0.024)	(0.024)	(0.024)
Fairness Tr.	0.173***	0.042*	0.053**	0.052**	0.102***
	(0.047)	(0.023)	(0.023)	(0.024)	(0.023)
Controls	No	No	No	No	No
R2	0.004	0.002	0.002	0.002	0.006
Observations	4371.000	4371.000	4371.000	4371.000	4371.000

Note. This table reports results from regressions that regress preferences for redistribution on treatment variables *without* controlling for other factors. Standard errors are in parentheses. *Significance levels*: *10%, **5%, ***1%.

Table I18: Main Treatment Effects with Fairness Video

	(1) RP Index (st. dev)	(2) Wants redistribution (0-1)	(3) Increase top taxes (0-1)	(4) Gov. reduce ineq. (0-1)	(5) Ineq. is serious issue (0-1)
Crime Ext. Tr.	0.037	0.031	-0.005	0.007	0.020
	(0.036)	(0.020)	(0.021)	(0.020)	(0.019)
Trust Ext. Tr.	0.043	0.006	0.004	0.036*	0.017
	(0.037)	(0.021)	(0.022)	(0.020)	(0.020)
Full Ext. Tr.	0.107***	0.050**	-0.012	0.048**	0.069***
	(0.037)	(0.021)	(0.022)	(0.020)	(0.020)
Fairness Tr.	0.208***	0.052**	0.065***	0.067***	0.115***
	(0.037)	(0.021)	(0.021)	(0.020)	(0.019)
Leans Republican	-0.635***	-0.190***	-0.210***	-0.264***	-0.249***
	(0.030)	(0.017)	(0.016)	(0.016)	(0.016)
Prior belief unfair	0.707***	0.146***	0.260***	0.260***	0.350***
	(0.027)	(0.015)	(0.015)	(0.014)	(0.015)
Male	-0.138***	-0.056***	-0.061***	-0.036***	-0.046***
	(0.026)	(0.015)	(0.015)	(0.014)	(0.013)
Controls	Yes	Yes	Yes	Yes	Yes
R2	0.391	0.169	0.170	0.293	0.313
Observations	4371	4371	4371	4371	4371

Note. This table reports results from a regression of different redistributive preference outcomes and the treatment dummies, as well as socio-economic control variables. The RP index is normalized on the sample and has units of the number of standard deviations. The remaining variables are binary (0-1). Controls not listed in the table include trust in government, race, income-group, age-group, education, employment status, geographic region. Standard errors are in parentheses. *Significance levels*: *10%, **5%, ***1%.

Table I19: Treatment effects with controls

	(1)	(2)	(3)	(4)	(5)
	RP Index	Wants redistribution	Increase top taxes	Gov. reduce ineq.	Ineq. is serious issue
	b/se	b/se	b/se	b/se	b/se
Crime Ext. Tr.	0.037	0.031	-0.005	0.007	0.020
	(0.036)	(0.020)	(0.021)	(0.020)	(0.019)
Trust Ext. Tr.	0.043	0.006	0.004	0.036*	0.017
	(0.037)	(0.021)	(0.022)	(0.020)	(0.020)
Full Ext. Tr.	0.107***	0.050**	-0.012	0.048**	0.069***
	(0.037)	(0.021)	(0.022)	(0.020)	(0.020)
Fairness Tr.	0.208***	0.052**	0.065***	0.067***	0.115***
	(0.037)	(0.021)	(0.021)	(0.020)	(0.019)
Leans Republican	-0.635***	-0.190***	-0.210***	-0.264***	-0.249***
	(0.030)	(0.017)	(0.016)	(0.016)	(0.016)
Prior belief unfair	0.707***	0.146***	0.260***	0.260***	0.350***
	(0.027)	(0.015)	(0.015)	(0.014)	(0.015)
Trusts the government	0.174***	0.070***	0.016	0.115***	0.050***
	(0.028)	(0.017)	(0.016)	(0.015)	(0.015)
Male	-0.138***	-0.056***	-0.061***	-0.036***	-0.046***
	(0.026)	(0.015)	(0.015)	(0.014)	(0.013)
Black	0.016	0.081***	-0.124***	0.000	0.066***
	(0.045)	(0.028)	(0.026)	(0.026)	(0.023)
Neither black or white	0.077**	0.060***	-0.009	0.038*	0.022
	(0.037)	(0.021)	(0.021)	(0.020)	(0.019)
Income: 25-50k	0.018	-0.011	0.039*	0.009	-0.012
	(0.036)	(0.020)	(0.020)	(0.019)	(0.018)
Income: 50-100k	-0.084**	-0.038*	0.008	-0.038**	-0.052***
	(0.036)	(0.020)	(0.020)	(0.019)	(0.019)
Income: 100k and more	-0.131***	-0.055**	-0.004	-0.048**	-0.082***
	(0.042)	(0.024)	(0.024)	(0.022)	(0.022)
Age 30-39	0.103**	0.021	0.050*	0.060**	0.018
	(0.046)	(0.027)	(0.026)	(0.025)	(0.024)
Age 40-49	0.024	-0.014	0.091***	-0.029	-0.013
	(0.046)	(0.027)	(0.026)	(0.025)	(0.024)
Age 50-59	-0.046	-0.090***	0.114***	-0.055**	-0.036
	(0.049)	(0.028)	(0.027)	(0.027)	(0.026)
Age 60-69	-0.170***	-0.147***	0.119***	-0.132***	-0.084***
	(0.048)	(0.028)	(0.027)	(0.026)	(0.025)
Age 70 and above	-0.274***	-0.183***	0.112***	-0.225***	-0.098***
	(0.050)	(0.028)	(0.027)	(0.027)	(0.026)
4-year college degree or more	-0.041	-0.001	-0.012	-0.029**	-0.018
	(0.027)	(0.015)	(0.015)	(0.014)	(0.014)
Unemployed	0.029	-0.003	0.032	0.000	0.012
	(0.047)	(0.026)	(0.026)	(0.025)	(0.024)
Outside the labor force	-0.029	-0.024	0.046***	-0.021	-0.042***
	(0.030)	(0.017)	(0.017)	(0.016)	(0.016)
West	-0.018	-0.016	0.006	0.000	-0.016
	(0.032)	(0.018)	(0.018)	(0.017)	(0.017)
North-East	0.113***	0.033	0.057***	0.051***	0.022
	(0.036)	(0.021)	(0.020)	(0.019)	(0.019)
Midwest	0.010	-0.017	0.044**	-0.010	-0.003
	(0.032)	(0.018)	(0.018)	(0.017)	(0.017)
Controls	Yes	Yes	Yes	Yes	Yes
R2	0.391	0.169	0.170	0.293	0.313
Observations	4371.000	4371.000	4371.000	4371.000	4371.000

Note. This table reports results from regressions that regress preferences for redistribution on treatment variables and reporting all controls. Standard errors are in parentheses. *Significance levels*: *10%, **5%, ***1%.

Table I20: First-stage effects of treatments

	(1) General neg. ext. b/se	(2) Ineq. incr. crime b/se	(3) Ineq. red. trust b/se	(4) Ineq. red. growth b/se	(5) Society unfair (post) b/se	(6) Rich b/c hard work b/se
Crime Ext. Tr.	0.088***	0.093***	0.059***	0.086***	0.012	-0.018
	(0.021)	(0.018)	(0.020)	(0.022)	(0.020)	(0.020)
Trust Ext. Tr.	0.050**	0.048**	0.096***	0.076***	0.025	-0.028
	(0.021)	(0.019)	(0.020)	(0.023)	(0.020)	(0.020)
Full Ext. Tr.	0.085***	0.084***	0.097***	0.062***	0.016	-0.030
	(0.021)	(0.019)	(0.020)	(0.023)	(0.020)	(0.020)
Fairness Tr.	0.075***	0.017	0.037*	0.033	0.079***	-0.079***
	(0.021)	(0.019)	(0.021)	(0.022)	(0.020)	(0.020)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
R2	0.159	0.084	0.093	0.102	0.239	0.241
Observations	4371.000	4371.000	4371.000	4371.000	4371.000	4371.000

Note. This table reports results from a regression of different externality beliefs and fairness views on the treatment dummies, as well as socio-economic control variables. Controls not listed in the table include political leaning, gender, trust in government, race, income-group, age-group, education, employment status, geographic region. Standard errors are in parentheses. *Significance levels*: *10%, **5%, ***1%.

Table I21: Treatment effects with controls and population weights.

	(1) RP Index b/se	(2) Wants redistribution b/se	(3) Gov. reduce ineq. b/se	(4) Ineq. is serious issue b/se	(5) Increase top taxes b/se
Crime Ext. Tr.	-0.007	0.021	-0.023	0.006	-0.015
	(0.051)	(0.029)	(0.027)	(0.026)	(0.028)
Trust Ext. Tr.	-0.008	-0.007	0.015	-0.005	-0.015
	(0.056)	(0.031)	(0.029)	(0.029)	(0.030)
Full Ext. Tr.	0.091*	0.064**	0.035	0.052**	-0.019
	(0.051)	(0.030)	(0.028)	(0.027)	(0.030)
Fairness Tr.	0.148***	0.023	0.055**	0.089***	0.045
	(0.050)	(0.030)	(0.027)	(0.026)	(0.029)
Controls	Yes	Yes	Yes	Yes	Yes
R2	0.386	0.184	0.273	0.309	0.181
Observations	4363.000	4363.000	4363.000	4363.000	4363.000

Note. This table reports results from a regression of different redistributive preference outcomes and the treatment dummies, as well as socio-economic control variables. Controls not listed in the table include trust in government, race, income-group, age-group, education, employment status, geographic region. Standard errors are in parentheses. Observations are reweighted to match representativity by gender, race, age, political affiliation, college degree, income-group, and geographic region. *Significance levels*: *10%, **5%, ***1%.

Table I22: Emotional reactions to treatments

	(1) Active control	(2) Crime	(3) Trust	(4) Full externality	(5) Fairness
Anger	2.8%	6.2%	2.9%	7.8%	11.8%
Concern	19.5%	37.2%	28.2%	32.0%	32.9%
Surprise	10.8%	13.9%	12.5%	13.0%	12.9%
Interest	41.5%	37.1%	42.2%	37.8%	34.0%
Indifference	17.7%	17.7%	19.2%	17.5%	17.9%
Confusion	16.9%	4.2%	6.0%	5.8%	4.5%
Observations	390	927	822	806	867

Table I23: First-stage effects of treatments with population weights

	(1) General neg. ext. b/se	(2) Ineq. incr. crime b/se	(3) Ineq. red. trust b/se	(4) Ineq. red. growth b/se	(5) Society unfair (post) b/se	(6) Rich b/c hard work b/se
Crime Ext. Tr.	0.061**	0.074***	0.031	0.049*	0.014	0.006
	(0.028)	(0.025)	(0.027)	(0.029)	(0.026)	(0.027)
Trust Ext. Tr.	0.003	0.059**	0.090***	0.058*	0.033	0.036
	(0.031)	(0.027)	(0.028)	(0.031)	(0.028)	(0.028)
Full Ext. Tr.	0.088***	0.101***	0.106***	0.057*	0.015	0.037
	(0.028)	(0.026)	(0.027)	(0.031)	(0.027)	(0.028)
Fairness Tr.	0.073***	0.019	0.033	0.003	0.062**	0.086***
	(0.028)	(0.027)	(0.029)	(0.030)	(0.027)	(0.026)
Controls						
R2	0.170	0.091	0.097	0.103	0.246	0.233
Observations	4363.000	4363.000	4363.000	4363.000	4363.000	4363.000

Note. This table reports results from a regression of different externality beliefs and fairness views on the treatment dummies, as well as socio-economic control variables. Controls not listed in the table include political leaning, gender, trust in government, race, income-group, age-group, education, employment status, geographic region. Standard errors are in parentheses. Observations are reweighted to match representativity by gender, race, age, political affiliation, college degree, income-group, and geographic region. *Significance levels*: *10%, **5%, ***1%.

Table I24: Mediation analysis: Treatment effects including beliefs as regressors

	(1) RP Index b/se	(2) RP Index b/se	(3) RP Index b/se	(4) RP Index b/se
Crime Ext. Tr.	0.037	-0.018	0.024	-0.008
	(0.036)	(0.035)	(0.034)	(0.033)
Trust Ext. Tr.	0.043	0.013	0.019	0.004
	(0.037)	(0.036)	(0.034)	(0.034)
Full Ext. Tr.	0.107***	0.055	0.087**	0.058*
	(0.037)	(0.036)	(0.034)	(0.034)
Fairness Tr.	0.208***	0.170***	0.135***	0.122***
	(0.037)	(0.035)	(0.034)	(0.033)
General neg. ext.		0.468***		0.301***
		(0.029)		(0.028)
Ineq. incr. crime		0.149***		0.076**
		(0.032)		(0.030)
Society is unfair (post)			0.508***	0.414***
			(0.029)	(0.030)
Rich because of hard luck			0.406***	0.372***
			(0.029)	(0.029)
Controls	Yes	Yes	Yes	Yes
R2	0.391	0.442	0.489	0.507
Observations	4371	4371	4371	4371

Note. This table reports results from a regression of different redistributive preference outcomes on the treatment indicators and post-treatment inequality beliefs and fairness views, as well as socio-economic control variables. Controls not listed include pre-treatment fairness views, race, income-group, age-group, gender, race, income-group, age-group, education, employment status, geographic region. Standard errors are in parentheses. *Significance levels*: *10%, **5%, ***1%.

Table I25: Treatment effects interacted with prior externality belief

	(1)	(2)	(3)	(4)	(5)
	RP Index	Wants redistribution	Increase top taxes	Gov. reduce ineq.	Ineq. is serious issue
	b/se	b/se	b/se	b/se	b/se
Crime Ext. Tr.	0.121**	0.062*	0.021	0.023	0.068**
	(0.059)	(0.032)	(0.035)	(0.032)	(0.031)
Trust Ext. Tr.	0.090	0.017	-0.018	0.059*	0.072**
	(0.059)	(0.033)	(0.035)	(0.032)	(0.032)
Full Ext. Tr.	0.137**	0.074**	0.001	0.030	0.091***
	(0.058)	(0.033)	(0.035)	(0.032)	(0.032)
Fairness Tr.	0.220***	0.045	0.069**	0.063**	0.139***
	(0.056)	(0.031)	(0.034)	(0.031)	(0.030)
Crime*Unequal countries function worse	-0.153**	-0.054	-0.049	-0.031	-0.086**
	(0.074)	(0.042)	(0.044)	(0.041)	(0.040)
Trust*Unequal countries function worse	-0.091	-0.022	0.028	-0.042	-0.096**
	(0.075)	(0.043)	(0.045)	(0.041)	(0.041)
Full Ext*Unequal countries function worse	-0.056	-0.041	-0.025	0.026	-0.040
	(0.075)	(0.043)	(0.044)	(0.041)	(0.041)
Fairness*Unequal countries function worse	-0.013	0.014	-0.004	0.010	-0.038
	(0.073)	(0.043)	(0.043)	(0.040)	(0.039)
Belief uneq countr. worse.	0.314***	0.081***	0.116***	0.102***	0.153***
	(0.051)	(0.030)	(0.031)	(0.028)	(0.028)
Controls	Yes	Yes	Yes	Yes	Yes
R2	0.405	0.173	0.181	0.302	0.323
Observations	4371.000	4371.000	4371.000	4371.000	4371.000

Note. This table reports results from a regression of different redistributive preference outcomes on the treatment indicators and their interaction with pre-treatment externality view. Controls not listed include pre-treatment fairness views, race, income-group, age-group, race, income-group, age-group, education, employment status, geographic region. Standard errors are in parentheses. *Significance levels*: *10%, **5%, ***1%.

Table I26: Treatment effects interacted with those that say they learned something new in the video

	(1)	(2)	(3)	(4)	(5)
	RP Index	Wants redistribution	Increase top taxes	Gov. reduce ineq.	Ineq. is serious issue
	b/se	b/se	b/se	b/se	b/se
Crime Ext. Tr.	-0.152***	-0.049*	-0.052	-0.064**	-0.054*
	(0.055)	(0.029)	(0.032)	(0.029)	(0.029)
Trust Ext. Tr.	-0.046	-0.053*	0.033	-0.024	-0.022
	(0.056)	(0.031)	(0.034)	(0.030)	(0.030)
Full Ext. Tr.	-0.057	-0.060*	-0.009	-0.025	0.011
	(0.059)	(0.032)	(0.036)	(0.032)	(0.032)
Fairness Tr.	0.012	0.006	0.028	-0.053*	0.036
	(0.057)	(0.031)	(0.033)	(0.030)	(0.030)
Learned something new	0.097*	-0.010	0.088***	0.033	0.029
	(0.053)	(0.031)	(0.031)	(0.029)	(0.030)
Crime*Learned something new	0.220***	0.121***	0.021	0.085**	0.090**
	(0.077)	(0.043)	(0.044)	(0.041)	(0.041)
Trust*Learned something new	0.077	0.090**	-0.087*	0.069	0.039
	(0.079)	(0.045)	(0.046)	(0.042)	(0.043)
FullExt*Learned something new	0.174**	0.156***	-0.053	0.082*	0.064
	(0.080)	(0.045)	(0.047)	(0.043)	(0.044)
Fairness*Learned something new	0.231***	0.071	0.008	0.156***	0.098**
	(0.078)	(0.045)	(0.045)	(0.042)	(0.041)
Controls	Yes	Yes	Yes	Yes	Yes
R2	0.403	0.176	0.175	0.305	0.320
Observations	4371.000	4371.000	4371.000	4371.000	4371.000

Note. This table reports results from a regression of different redistributive preference outcomes on the treatment indicators and their interaction with self-reported indicator to have learned something new. Controls not listed include pre-treatment fairness views, race, income-group, age-group, race, income-group, age-group, education, employment status, geographic region. Standard errors are in parentheses. *Significance levels*: *10%, **5%, ***1%.

Table I27: Treatment effects interacted with male dummy

	(1)	(2)	(3)	(4)	(5)
	RP Index	Wants redistribution	Increase top taxes	Gov. reduce ineq.	Ineq. is serious issue
	b/se	b/se	b/se	b/se	b/se
Crime Ext. Tr.	0.024	0.049*	-0.019	0.014	-0.010
	(0.051)	(0.030)	(0.029)	(0.028)	(0.027)
Trust Ext. Tr.	0.012	0.005	-0.012	0.036	-0.011
	(0.051)	(0.031)	(0.030)	(0.029)	(0.028)
Full Ext. Tr.	0.010	0.048	-0.088***	0.016	0.039
	(0.052)	(0.032)	(0.030)	(0.029)	(0.028)
Fairness Tr.	0.194***	0.072**	0.052*	0.071**	0.085***
	(0.053)	(0.032)	(0.030)	(0.029)	(0.028)
Male	-0.198***	-0.041	-0.108***	-0.044	-0.092***
	(0.051)	(0.030)	(0.030)	(0.028)	(0.028)
CrimeXmale	0.024	-0.038	0.027	-0.015	0.060
	(0.073)	(0.041)	(0.042)	(0.040)	(0.039)
TrustXmale	0.062	0.003	0.032	-0.001	0.056
	(0.075)	(0.043)	(0.043)	(0.040)	(0.040)
FullExtXmale	0.196***	0.004	0.153***	0.065	0.060
	(0.075)	(0.043)	(0.043)	(0.040)	(0.040)
FairnessXmale	0.030	-0.038	0.029	-0.006	0.059
	(0.074)	(0.043)	(0.042)	(0.040)	(0.039)
Controls	Yes	Yes	Yes	Yes	Yes
R2	0.392	0.169	0.173	0.294	0.314
Observations	4371.000	4371.000	4371.000	4371.000	4371.000

Note. This table reports results from a regression of different redistributive preference outcomes on the treatment indicators and their interaction with a male dummy. Controls not listed include pre-treatment fairness views, race, income-group, age-group, race, income-group, age-group, education, employment status, geographic region. Standard errors are in parentheses. *Significance levels*: *10%, **5%, ***1%.

Table I28: Treatment effects interacted with having a yearly income above $\$50,000$

	(1)	(2)	(3)	(4)	(5)
	RP Index	Wants redistribution	Increase top taxes	Gov. reduce ineq.	Ineq. is serious issue
	b/se	b/se	b/se	b/se	b/se
Crime Ext. Tr.	0.041	0.016	0.013	-0.006	0.035
	(0.052)	(0.030)	(0.029)	(0.029)	(0.027)
CrimeIncome	-0.009	0.030	-0.037	0.027	-0.033
	(0.073)	(0.041)	(0.042)	(0.040)	(0.039)
Trust Ext. Tr.	0.067	-0.001	0.030	0.024	0.043
	(0.053)	(0.030)	(0.030)	(0.029)	(0.028)
TrustIncome	-0.049	0.015	-0.055	0.026	-0.055
	(0.074)	(0.043)	(0.043)	(0.040)	(0.040)
Full Ext. Tr.	0.099*	0.041	-0.009	0.019	0.091***
	(0.053)	(0.031)	(0.031)	(0.029)	(0.029)
FullIncome	0.017	0.019	-0.007	0.058	-0.046
	(0.074)	(0.043)	(0.043)	(0.040)	(0.040)
Fairness Tr.	0.308***	0.091***	0.106***	0.080***	0.168***
	(0.054)	(0.032)	(0.030)	(0.029)	(0.028)
FairnessIncome	-0.197***	-0.076*	-0.080*	-0.023	-0.105***
	(0.073)	(0.043)	(0.042)	(0.040)	(0.039)
Controls	Yes	Yes	Yes	Yes	Yes
R2	0.392	0.170	0.171	0.294	0.314
Observations	4371	4371	4371	4371	4371

Note. This table reports results from a regression of different redistributive preference outcomes on the treatment indicators and their interaction with having a yearly income above $\$50,000$. Controls not listed include pre-treatment fairness views, race, income-group, age-group, race, income-group, age-group, education, employment status, geographic region. Standard errors are in parentheses. *Significance levels*: *10%, **5%, ***1%.

Table I29: Treatment effects interacted with having a yearly income above $\$100,000$

	(1)	(2)	(3)	(4)	(5)
	RP Index	Wants redistribution	Increase top taxes	Gov. reduce ineq.	Ineq. is serious issue
	b/se	b/se	b/se	b/se	b/se
Crime Ext. Tr.	0.038	0.027	0.006	0.007	0.015
	(0.041)	(0.023)	(0.023)	(0.023)	(0.022)
CrimeIncome	-0.005	0.022	-0.052	0.001	0.022
	(0.087)	(0.049)	(0.053)	(0.047)	(0.048)
Trust Ext. Tr.	0.056	0.003	0.033	0.033	0.012
	(0.042)	(0.024)	(0.024)	(0.023)	(0.022)
TrustIncome	-0.072	0.018	-0.157***	0.013	0.023
	(0.092)	(0.052)	(0.056)	(0.049)	(0.051)
Full Ext. Tr.	0.093**	0.047*	-0.012	0.040*	0.059***
	(0.042)	(0.024)	(0.024)	(0.023)	(0.022)
FullIncome	0.095	0.020	0.013	0.045	0.058
	(0.094)	(0.055)	(0.055)	(0.050)	(0.051)
Fairness Tr.	0.227***	0.061**	0.074***	0.071***	0.121***
	(0.041)	(0.024)	(0.024)	(0.022)	(0.022)
FairnessIncome	-0.098	-0.048	-0.044	-0.019	-0.030
	(0.089)	(0.050)	(0.053)	(0.048)	(0.047)
Controls	Yes	Yes	Yes	Yes	Yes
R2	0.391	0.169	0.172	0.294	0.314
Observations	4371	4371	4371	4371	4371

Note. This table reports results from a regression of different redistributive preference outcomes on the treatment indicators and their interaction with having a yearly income above $\$100,000$. Controls not listed include pre-treatment fairness views, race, income-group, age-group, race, income-group, age-group, education, employment status, geographic region. Standard errors are in parentheses. *Significance levels*: *10%, **5%, ***1%.

Table I30: Treatment effects interacted with Republican leaning dummy

	(1)	(2)	(3)	(4)	(5)
	RP Index	Wants redistribution	Increase top taxes	Gov. reduce ineq.	Ineq. is serious issue
	b/se	b/se	b/se	b/se	b/se
Crime Ext. Tr.	0.039	0.068**	0.009	-0.026	0.006
	(0.051)	(0.032)	(0.029)	(0.030)	(0.028)
Trust Ext. Tr.	0.063	0.033	0.023	0.033	0.002
	(0.052)	(0.034)	(0.029)	(0.030)	(0.029)
Full Ext. Tr.	0.192***	0.116***	0.020	0.061**	0.079***
	(0.051)	(0.033)	(0.029)	(0.029)	(0.028)
Fairness Tr.	0.218***	0.069**	0.061**	0.069**	0.115***
	(0.051)	(0.033)	(0.028)	(0.029)	(0.027)
Leans Republican	-0.592***	-0.135***	-0.188***	-0.273***	-0.256***
	(0.053)	(0.031)	(0.031)	(0.029)	(0.029)
CrimeXRepublicanLeaning	-0.005	-0.071*	-0.028	0.065	0.027
	(0.073)	(0.041)	(0.042)	(0.040)	(0.039)
TrustXRepublicanLeaning	-0.039	-0.054	-0.035	0.005	0.028
	(0.074)	(0.043)	(0.043)	(0.040)	(0.040)
FullExtXRepublicanLeaning	-0.165**	-0.128***	-0.062	-0.027	-0.020
	(0.074)	(0.043)	(0.043)	(0.040)	(0.040)
FairnessXRepublicanLeaning	-0.020	-0.034	0.008	-0.003	0.000
	(0.073)	(0.043)	(0.042)	(0.040)	(0.039)
Controls	Yes	Yes	Yes	Yes	Yes
R2	0.391	0.171	0.171	0.294	0.313
Observations	4371.000	4371.000	4371.000	4371.000	4371.000

Note. This table reports results from a regression of different redistributive preference outcomes on the treatment indicators and their interaction with an indicator that the respondent leans republican. Controls not listed include pre-treatment fairness views, race, income-group, age-group, education, employment status, geographic region. Standard errors are in parentheses. *Significance levels*: *10%, **5%, ***1%.

Table I31: Treatment effects interacted with dummy indicating that the subject believes that the current economic system in the US is unfair

	(1)	(2)	(3)	(4)	(5)
	RP Index	Wants redistribution	Increase top taxes	Gov. reduce ineq.	Ineq. is serious issue
	b/se	b/se	b/se	b/se	b/se
Crime Ext. Tr.	0.022	-0.007	0.006	0.024	0.009
	(0.051)	(0.027)	(0.032)	(0.028)	(0.028)
Trust Ext. Tr.	0.039	-0.011	0.036	0.036	-0.004
	(0.052)	(0.028)	(0.033)	(0.028)	(0.029)
Full Ext. Tr.	0.091*	0.038	-0.011	0.041	0.064**
	(0.053)	(0.028)	(0.033)	(0.029)	(0.030)
Fairness Tr.	0.147***	0.009	0.088***	0.035	0.080***
	(0.052)	(0.028)	(0.032)	(0.028)	(0.029)
Prior belief unfair	0.669***	0.103***	0.286***	0.251***	0.322***
	(0.051)	(0.030)	(0.030)	(0.028)	(0.028)
CrimeXdPriorUnfair	0.030	0.073*	-0.021	-0.031	0.022
	(0.073)	(0.041)	(0.042)	(0.040)	(0.039)
TrustXdPriorUnfair	0.008	0.033	-0.062	0.001	0.040
	(0.074)	(0.042)	(0.044)	(0.040)	(0.040)
FullExtXdPriorUnfair	0.031	0.024	-0.002	0.013	0.010
	(0.075)	(0.043)	(0.043)	(0.041)	(0.040)
FairnessXdPriorUnfair	0.119	0.084**	-0.046	0.065*	0.068*
	(0.073)	(0.042)	(0.042)	(0.039)	(0.039)
Controls	Yes	Yes	Yes	Yes	Yes
R2	0.391	0.170	0.171	0.294	0.314
Observations	4371.000	4371.000	4371.000	4371.000	4371.000

Note. This table reports results from a regression of different redistributive preference outcomes on the treatment indicators and their interaction with pre-treatment fairness views. Controls not listed include, political leaning, pre-treatment fairness views, race, income-group, age-group, education, employment status, geographic region. Standard errors are in parentheses. *Significance levels*: *10%, **5%, ***1%.

Table I34: Treatment effects with controls using all completed responses

	(1)	(2)	(3)	(4)	(5)
	RP Index	Wants redistribution	Gov. reduce ineq.	Ineq. is serious issue	Increase top taxes
	b/se	b/se	b/se	b/se	b/se
Crime Ext. Tr.	0.039	0.034*	0.013	0.023	-0.015
	(0.035)	(0.020)	(0.019)	(0.019)	(0.020)
Trust Ext. Tr.	0.055	0.012	0.049**	0.014	0.004
	(0.036)	(0.021)	(0.019)	(0.020)	(0.021)
Full Ext. Tr.	0.098***	0.046**	0.047**	0.060***	-0.013
	(0.036)	(0.020)	(0.019)	(0.019)	(0.021)
Fairness Tr.	0.202***	0.056***	0.070***	0.106***	0.056***
	(0.036)	(0.021)	(0.019)	(0.019)	(0.020)
Controls	Yes	Yes	Yes	Yes	Yes
R2	0.360	0.169	0.272	0.277	0.159
Observations	4865.000	4865.000	4865.000	4865.000	4865.000

Note. This table reports results from a regression of different redistributive preference outcomes and the treatment dummies, as well as socio-economic control variables. Controls not listed in the table include trust in government, race, income-group, age-group, education, employment status, geographic region. Standard errors are in parentheses. *Significance levels*: *10%, **5%, ***1%.

Table I32: Predictive power of various beliefs in Survey 2

	(1) RP Index b/se	(2) RP Index b/se	(3) RP Index b/se	(4) RP Index b/se	(5) RP Index b/se	(6) RP Index b/se
Rich because of luck		0.681***				0.394***
		(0.044)				(0.045)
Society is unfair		0.648***				0.500***
		(0.044)				(0.043)
Ineq. incr. crime			0.324***			0.102**
			(0.046)			(0.040)
Neg. externality belief			0.515***			0.151***
			(0.044)			(0.039)
Leans Republican				-0.502***		-0.261***
				(0.053)		(0.052)
Sanders/Harris supporter				0.618***		0.415***
				(0.055)		(0.052)
Trusts the government					0.370***	0.090**
					(0.047)	(0.039)
Taxation reduces work					-0.181***	0.028
					(0.043)	(0.035)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Only Control Group	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R2	0.133	0.404	0.234	0.358	0.165	0.486
Observations	2360.000	2360.000	2360.000	2360.000	2360.000	2360.000

Note. This table reports results from a regression of different redistributive preference outcomes on fairness views, political views, externality beliefs and attitudes towards the government, as well as socio-economic control variables. Controls not listed include gender, race, income-group, age-group, education, employment status, geographic region. Standard errors are in parentheses. Observations are reweighted to match representativity by gender, race, age, political affiliation, college degree, income-group, and geographic region. Data is from Survey 2 only. Standard errors are in parentheses. *Significance levels*: *10%, **5%, ***1%.

Table I33: Predictive power of various beliefs with population weights

	(1) RP Index b/se	(2) RP Index b/se	(3) RP Index b/se	(4) RP Index b/se	(5) RP Index b/se	(6) RP Index b/se
Rich because of hard work		-0.612***				-0.398***
		(0.084)				(0.081)
Society is unfair (post)		0.546***				0.360***
		(0.083)				(0.077)
Belief uneq countr. worse.			0.510***			0.344***
			(0.080)			(0.072)
General neg. ext.			0.555***			0.256***
			(0.082)			(0.081)
Leans Republican				-0.335**		-0.215*
				(0.137)		(0.110)
SandersKamala				0.573***		0.268**
				(0.138)		(0.111)
govtrust					0.228***	0.064
					(0.054)	(0.048)
Agrees/disagrees that people work much less if taxed more					-0.054	-0.001
					(0.038)	(0.032)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Only Control Group	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R2	0.092	0.328	0.277	0.263	0.131	0.448
Observations	929.000	929.000	929.000	929.000	929.000	929.000

Note. This table reports results from a regression of different redistributive preference outcomes on fairness views, political views, externality beliefs and attitudes towards the government, as well as socio-economic control variables. Controls not listed include gender, race, income-group, age-group, education, employment status, geographic region. Standard errors are in parentheses. Observations are reweighted to match representativity by gender, race, age, political affiliation, college degree, income-group, and geographic region. *Significance levels*: *10%, **5%, ***1%.

Table I35: First-stage effects of treatments using all completed responses

	(1)	(2)	(3)	(4)	(5)	(6)
	General neg. ext.	Ineq. incr. crime	Ineq. red. trust	Ineq. red. growth	Society unfair (post)	Rich b/c hard work
	b/se	b/se	b/se	b/se	b/se	b/se
Crime Ext. Tr.	0.083***	0.092***	0.060***	0.087***	0.014	0.017
	(0.020)	(0.018)	(0.019)	(0.021)	(0.019)	(0.019)
Trust Ext. Tr.	0.042**	0.042**	0.090***	0.081***	0.020	0.021
	(0.020)	(0.019)	(0.020)	(0.022)	(0.020)	(0.020)
Full Ext. Tr.	0.073***	0.074***	0.087***	0.058***	0.009	0.027
	(0.020)	(0.018)	(0.020)	(0.022)	(0.019)	(0.019)
Fairness Tr.	0.066***	0.018	0.038*	0.035*	0.067***	0.071***
	(0.020)	(0.019)	(0.020)	(0.021)	(0.019)	(0.019)
Controls						
R2	0.162	0.092	0.100	0.097	0.234	0.219
Observations	4865.000	4865.000	4865.000	4865.000	4865.000	4865.000

Note. This table reports results from a regression of different externality beliefs and fairness views on the treatment dummies, as well as socio-economic control variables. Controls not listed in the table include political leaning, gender, trust in government, race, income-group, age-group, education, employment status, geographic region. Standard errors are in parentheses. *Significance levels*: *10%, **5%, ***1%.

Table I36: Predictive power of various beliefs using all completed responses

	(1)	(2)	(3)	(4)	(5)	(6)
	RP Index	RP Index	RP Index	RP Index	RP Index	RP Index
	b/se	b/se	b/se	b/se	b/se	b/se
Rich because of luck		-0.550***				-0.337***
		(0.059)				(0.056)
Society unfair (post)		0.628***				0.440***
		(0.058)				(0.056)
Belief uneq. countr. worse			0.457***			0.298***
			(0.056)			(0.050)
Neg. externality belief			0.600***			0.224***
			(0.055)			(0.052)
Leans Republican				-0.361***		-0.194***
				(0.084)		(0.072)
Sanders/Harris supporter				0.592***		0.331***
				(0.085)		(0.076)
Trusts the government					0.220***	0.081**
					(0.036)	(0.033)
Taxation reduces work					-0.097***	-0.012
					(0.025)	(0.020)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Only Control Group	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R2	0.090	0.346	0.274	0.279	0.138	0.465
Observations	1026.000	1026.000	1026.000	1026.000	1026.000	1026.000

Note. This table reports results from a regression of different redistributive preference outcomes on fairness views, political views, externality beliefs and attitudes towards the government, as well as socio-economic control variables. Controls not listed include gender, race, income-group, age-group, education, employment status, geographic region. Standard errors are in parentheses. *Significance levels*: *10%, **5%, ***1%.

Table I37: Treatment effects with controls using only respondents that passed all attention checks

	(1)	(2)	(3)	(4)	(5)
	RP Index	Wants redistribution	Gov. reduce ineq.	Ineq. is serious issue	Increase top taxes
	b/se	b/se	b/se	b/se	b/se
Crime Ext. Tr.	0.027	0.038	0.011	0.007	-0.016
	(0.043)	(0.025)	(0.024)	(0.023)	(0.025)
Trust Ext. Tr.	0.021	0.008	0.039	0.009	-0.026
	(0.045)	(0.026)	(0.024)	(0.024)	(0.026)
Full Ext. Tr.	0.075*	0.045*	0.056**	0.050**	-0.040
	(0.044)	(0.026)	(0.024)	(0.024)	(0.026)
Fairness Tr.	0.185***	0.050*	0.081***	0.094***	0.047*
	(0.043)	(0.026)	(0.024)	(0.023)	(0.026)
Controls	Yes	Yes	Yes	Yes	Yes
R2	0.436	0.201	0.335	0.360	0.192
Observations	2892.000	2892.000	2892.000	2892.000	2892.000

Note. This table reports results from a regression of different redistributive preference outcomes and the treatment dummies, as well as socio-economic control variables. Controls not listed in the table include trust in government, race, income-group, age-group, education, employment status, geographic region. Standard errors are in parentheses. *Significance levels*: *10%, **5%, ***1%.

Table I38: First-stage effects of treatments using only respondents that passed all attention checks

	(1)	(2)	(3)	(4)	(5)	(6)
	General neg. ext.	Ineq. incr. crime	Ineq. red. trust	Ineq. red. growth	Society unfair (post)	Rich b/c hard work
	b/se	b/se	b/se	b/se	b/se	b/se
Crime Ext. Tr.	0.096***	0.099***	0.058**	0.107***	-0.000	0.011
	(0.025)	(0.021)	(0.024)	(0.027)	(0.024)	(0.024)
Trust Ext. Tr.	0.046*	0.047**	0.096***	0.062**	0.036	0.000
	(0.025)	(0.023)	(0.024)	(0.028)	(0.024)	(0.024)
Full Ext. Tr.	0.075***	0.081***	0.102***	0.054*	0.016	0.020
	(0.025)	(0.022)	(0.024)	(0.028)	(0.024)	(0.024)
Fairness Tr.	0.084***	0.035	0.044*	0.037	0.082***	0.054**
	(0.025)	(0.023)	(0.025)	(0.027)	(0.024)	(0.024)
Controls						
R2	0.169	0.086	0.095	0.121	0.278	0.284
Observations	2892.000	2892.000	2892.000	2892.000	2892.000	2892.000

Note. This table reports results from a regression of different externality beliefs and fairness views on the treatment dummies, as well as socio-economic control variables. Controls not listed in the table include political leaning, gender, trust in government, race, income-group, age-group, education, employment status, geographic region. Standard errors are in parentheses. *Significance levels*: *10%, **5%, ***1%.

Table I39: Predictive power of various beliefs using only respondents that passed all attention checks

	(1) RP Index b/se	(2) RP Index b/se	(3) RP Index b/se	(4) RP Index b/se	(5) RP Index b/se	(6) RP Index b/se
Rich because of luck		-0.642***				-0.383***
		(0.080)				(0.075)
Society unfair (post)		0.624***				0.396***
		(0.076)				(0.072)
Belief uneq. countr. worse			0.448***			0.281***
			(0.073)			(0.067)
Neg. externality belief			0.681***			0.298***
			(0.073)			(0.068)
Leans Republican				-0.366***		-0.195**
				(0.114)		(0.099)
Sanders/Harris supporter				0.629***		0.328***
				(0.114)		(0.104)
Trusts the government					0.256***	0.043
					(0.047)	(0.045)
Taxation reduces work					-0.090***	-0.003
					(0.033)	(0.025)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Only Control Group	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R2	0.130	0.411	0.344	0.337	0.188	0.525
Observations	597.000	597.000	597.000	597.000	597.000	597.000

Note. This table reports results from a regression of different redistributive preference outcomes on fairness views, political views, externality beliefs and attitudes towards the government, as well as socio-economic control variables. Controls not listed include gender, race, income-group, age-group, education, employment status, geographic region. Standard errors are in parentheses. *Significance levels*: *10%, **5%, ***1%.

Table I40: Treatment effects with controls and controlling for passing attention checks

	(1) General neg. ext. b/se	(2) Ineq. incr. crime b/se	(3) Ineq. red. trust b/se	(4) Ineq. red. growth b/se	(5) Society unfair (post) b/se	(6) Rich b/c hard work b/se
Crime Ext. Tr.	0.086***	0.091***	0.056***	0.084***	0.010	0.018
	(0.020)	(0.018)	(0.020)	(0.022)	(0.020)	(0.020)
Trust Ext. Tr.	0.046**	0.044**	0.090***	0.072***	0.021	0.026
	(0.021)	(0.019)	(0.020)	(0.023)	(0.020)	(0.020)
Full Ext. Tr.	0.078***	0.079***	0.090***	0.057**	0.011	0.028
	(0.021)	(0.019)	(0.020)	(0.023)	(0.020)	(0.020)
Fairness Tr.	0.070***	0.013	0.033	0.029	0.076***	0.078***
	(0.021)	(0.019)	(0.021)	(0.022)	(0.020)	(0.020)
Controls						
R2	0.167	0.091	0.101	0.106	0.244	0.242
Observations	4371.000	4371.000	4371.000	4371.000	4371.000	4371.000

Note. This table reports results from a regression of different redistributive preference outcomes and the treatment dummies, as well as socio-economic control variables. Controls not listed in the table include trust in government, race, income-group, age-group, education, employment status, geographic region, failing or passing any attention check. Standard errors are in parentheses. *Significance levels*: *10%, **5%, ***1%.

Table I41: Treatment effects, dropping active control group

	(1)	(2)	(3)	(4)	(5)
	RP Index	Wants redistribution	Gov. reduce ineq.	Ineq. is serious issue	Increase top taxes
	b/se	b/se	b/se	b/se	b/se
Crime Ext. Tr.	0.058	0.024	0.016	0.014	0.029
	(0.042)	(0.024)	(0.023)	(0.023)	(0.025)
Trust Ext. Tr.	0.064	-0.001	0.044*	0.012	0.037
	(0.043)	(0.025)	(0.023)	(0.023)	(0.026)
Full Ext. Tr.	0.127***	0.043*	0.056**	0.064***	0.021
	(0.043)	(0.025)	(0.024)	(0.023)	(0.026)
Fairness Tr.	0.228***	0.045*	0.076***	0.110***	0.099***
	(0.042)	(0.025)	(0.023)	(0.022)	(0.025)
Controls	Yes	Yes	Yes	Yes	Yes
R2	0.392	0.172	0.292	0.317	0.167
Observations	3977.000	3977.000	3977.000	3977.000	3977.000

Note. This table reports results from a regression of different redistributive preference outcomes and the treatment dummies, as well as socio-economic control variables. Controls not listed in the table include trust in government, race, income-group, age-group, education, employment status, geographic region. Standard errors are in parentheses. *Significance levels*: *10%, **5%, ***1%.

Table I42: First-stage effects, dropping active control group

	(1)	(2)	(3)	(4)	(5)	(6)
	General neg. ext.	Ineq. incr. crime	Ineq. red. trust	Ineq. red. growth	Society unfair (post)	Rich b/c hard work
	b/se	b/se	b/se	b/se	b/se	b/se
Crime Ext. Tr.	0.098***	0.092***	0.067***	0.095***	0.020	0.022
	(0.024)	(0.021)	(0.024)	(0.025)	(0.023)	(0.023)
Trust Ext. Tr.	0.060**	0.048**	0.105***	0.084***	0.033	0.031
	(0.025)	(0.022)	(0.024)	(0.026)	(0.024)	(0.023)
Full Ext. Tr.	0.095***	0.084***	0.106***	0.071***	0.024	0.032
	(0.025)	(0.022)	(0.024)	(0.026)	(0.024)	(0.024)
Fairness Tr.	0.084***	0.016	0.045*	0.042	0.087***	0.082***
	(0.024)	(0.022)	(0.024)	(0.026)	(0.023)	(0.023)
Controls						
R2	0.158	0.084	0.093	0.103	0.237	0.245
Observations	3977.000	3977.000	3977.000	3977.000	3977.000	3977.000

Note. This table reports results from a regression of different externality beliefs and fairness views on the treatment dummies, as well as socio-economic control variables. Controls not listed in the table include political leaning, gender, trust in government, race, income-group, age-group, education, employment status, geographic region. Standard errors are in parentheses. *Significance levels*: *10%, **5%, ***1%.

Table I43: Treatment effects, dropping passive control group

	(1)	(2)	(3)	(4)	(5)
	RP Index	Wants redistribution	Gov. reduce ineq.	Ineq. is serious issue	Increase top taxes
	b/se	b/se	b/se	b/se	b/se
Crime Ext. Tr.	0.010	0.039	-0.004	0.028	-0.049*
	(0.047)	(0.027)	(0.025)	(0.026)	(0.026)
Trust Ext. Tr.	0.016	0.014	0.024	0.025	-0.040
	(0.048)	(0.028)	(0.026)	(0.026)	(0.027)
Full Ext. Tr.	0.081*	0.058**	0.036	0.079***	-0.056**
	(0.048)	(0.028)	(0.026)	(0.026)	(0.027)
Fairness Tr.	0.180***	0.060**	0.055**	0.123***	0.020
	(0.048)	(0.028)	(0.025)	(0.025)	(0.027)
Controls	Yes	Yes	Yes	Yes	Yes
R2	0.390	0.170	0.294	0.314	0.176
Observations	3833.000	3833.000	3833.000	3833.000	3833.000

Note. This table reports results from a regression of different redistributive preference outcomes and the treatment dummies, as well as socio-economic control variables. Controls not listed in the table include trust in government, race, income-group, age-group, education, employment status, geographic region. Standard errors are in parentheses. *Significance levels*: *10%, **5%, ***1%.

Table I44: First-stage effect, dropping passive control group

	(1)	(2)	(3)	(4)	(5)	(6)
	General neg. ext.	Ineq. incr. crime	Ineq. red. trust	Ineq. red. growth	Society unfair (post)	Rich b/c hard work
	b/se	b/se	b/se	b/se	b/se	b/se
Crime Ext. Tr.	0.075***	0.093***	0.047*	0.077***	-0.000	0.015
	(0.027)	(0.024)	(0.026)	(0.028)	(0.026)	(0.026)
Trust Ext. Tr.	0.037	0.048**	0.083***	0.066**	0.013	0.024
	(0.027)	(0.024)	(0.026)	(0.029)	(0.026)	(0.026)
Full Ext. Tr.	0.072***	0.084***	0.086***	0.053*	0.004	0.027
	(0.027)	(0.024)	(0.026)	(0.029)	(0.026)	(0.026)
Fairness Tr.	0.061**	0.017	0.025	0.023	0.066***	0.075***
	(0.027)	(0.025)	(0.027)	(0.029)	(0.026)	(0.026)
Controls						
R2	0.155	0.085	0.090	0.098	0.234	0.243
Observations	3833.000	3833.000	3833.000	3833.000	3833.000	3833.000

Note. This table reports results from a regression of different externality beliefs and fairness views on the treatment dummies, as well as socio-economic control variables. Controls not listed in the table include political leaning, gender, trust in government, race, income-group, age-group, education, employment status, geographic region. Standard errors are in parentheses. *Significance levels*: *10%, **5%, ***1%.

Table I45: Treatment effects without controls

	(1)	(2)	(3)	(4)	(5)
	RP Index	Wants redistribution	Gov. reduce ineq.	Ineq. is serious issue	Increase top taxes
	b/se	b/se	b/se	b/se	b/se
Crime Ext. Tr.	0.036	0.031	0.005	0.022	-0.006
	(0.046)	(0.022)	(0.023)	(0.023)	(0.023)
Trust Ext. Tr.	0.055	0.010	0.041*	0.023	0.005
	(0.047)	(0.023)	(0.024)	(0.024)	(0.024)
Full Ext. Tr.	0.124***	0.059**	0.056**	0.078***	-0.014
	(0.048)	(0.023)	(0.024)	(0.024)	(0.024)
Fairness Tr.	0.173***	0.042*	0.052**	0.102***	0.053**
	(0.047)	(0.023)	(0.024)	(0.023)	(0.023)
Controls	No	No	No	No	No
R2	0.004	0.002	0.002	0.006	0.002
Observations	4371.000	4371.000	4371.000	4371.000	4371.000

Note. This table reports results from a regression of different redistributive preference outcomes and the treatment dummies, as well as socio-economic control variables. Standard errors are in parentheses. *Significance levels*: *10%, **5%, ***1%.

Table I46: First-stage effects of treatments without controls

	(1)	(2)	(3)	(4)	(5)	(6)
	General neg. ext.	Ineq. incr. crime	Ineq. red. trust	Ineq. red. growth	Society unfair (post)	Rich b/c hard work
	b/se	b/se	b/se	b/se	b/se	b/se
Crime Ext. Tr.	0.092***	0.094***	0.059***	0.089***	0.015	0.018
	(0.022)	(0.018)	(0.021)	(0.023)	(0.023)	(0.022)
Trust Ext. Tr.	0.052**	0.049**	0.096***	0.079***	0.030	0.033
	(0.023)	(0.020)	(0.021)	(0.024)	(0.023)	(0.023)
Full Ext. Tr.	0.088***	0.084***	0.101***	0.069***	0.019	0.032
	(0.023)	(0.019)	(0.021)	(0.024)	(0.023)	(0.023)
Fairness Tr.	0.068***	0.012	0.033	0.028	0.067***	0.066***
	(0.023)	(0.020)	(0.022)	(0.024)	(0.023)	(0.022)
Controls						
R2	0.005	0.009	0.007	0.005	0.002	0.002
Observations	4371.000	4371.000	4371.000	4371.000	4371.000	4371.000

Note. This table reports results from a regression of different externality beliefs and fairness views on the treatment dummies, as well as socio-economic control variables. Standard errors are in parentheses. *Significance levels*: *10%, **5%, ***1%.

Table I47: Treatment effects with controls using non-dichotomized variables

	(1)	(2)	(3)	(4)	(5)
	RP Index	Wants redistribution	Gov. reduce ineq.	Ineq. serious issue	Increase top taxes
	b/se	b/se	b/se	b/se	b/se
Crime Ext. Tr.	0.026	0.107	0.024	0.039	-0.043
	(0.036)	(0.069)	(0.048)	(0.044)	(0.059)
Trust Ext. Tr.	0.033	-0.006	0.066	0.040	0.031
	(0.036)	(0.071)	(0.049)	(0.045)	(0.061)
Full Ext. Tr.	0.098***	0.179**	0.106**	0.140***	0.007
	(0.036)	(0.071)	(0.049)	(0.045)	(0.062)
Fairness Tr.	0.209***	0.288***	0.180***	0.263***	0.182***
	(0.035)	(0.071)	(0.048)	(0.044)	(0.058)
Controls	Yes	Yes	Yes	Yes	Yes
R2	0.422	0.318	0.386	0.357	0.142
Observations	4371.000	4371.000	4371.000	4371.000	4371.000

Note. This table reports results from a regression of different redistributive preference outcomes and the treatment dummies, as well as socio-economic control variables. Controls not listed in the table include trust in government, race, income-group, age-group, education, employment status, geographic region. Standard errors are in parentheses. *Significance levels*: *10%, **5%, ***1%.

Table I48: First-stage effects of treatments using non-dichotomized variables

	(1)	(2)	(3)	(4)	(5)	(6)
	General neg. ext.	Ineq. incr. crime	Ineq. red. trust	Ineq. red. growth	Society unfair (post)	Rich b/c hard work
	b/se	b/se	b/se	b/se	b/se	b/se
Crime Ext. Tr.	0.194***	0.238***	-0.162***	-0.173***	0.047	-0.018
	(0.046)	(0.044)	(0.046)	(0.050)	(0.050)	(0.020)
Trust Ext. Tr.	0.119**	0.138***	-0.252***	-0.118**	0.017	-0.028
	(0.048)	(0.045)	(0.047)	(0.051)	(0.052)	(0.020)
Full Ext. Tr.	0.201***	0.223***	-0.265***	-0.127**	0.013	-0.030
	(0.048)	(0.046)	(0.048)	(0.052)	(0.053)	(0.020)
Fairness Tr.	0.170***	0.079*	-0.104**	-0.049	0.204***	-0.079***
	(0.047)	(0.047)	(0.048)	(0.051)	(0.051)	(0.020)
Controls						
R2	0.163	0.096	0.108	0.091	0.269	0.241
Observations	4371.000	4371.000	4371.000	4371.000	4371.000	4371.000

Note. This table reports results from a regression of different externality beliefs and fairness views on the treatment dummies, as well as socio-economic control variables. Controls not listed in the table include political leaning, gender, trust in government, race, income-group, age-group, education, employment status, geographic region. Standard errors are in parentheses. *Significance levels*: *10%, **5%, ***1%.

Table I49: Predictive power of various beliefs without controls

	(1)	(2)	(3)	(4)	(5)	(6)
	RP Index	RP Index	RP Index	RP Index	RP Index	RP Index
	b/se	b/se	b/se	b/se	b/se	b/se
Rich because of luck		-0.655***				-0.418***
		(0.063)				(0.060)
Society unfair (post)		0.646***				0.445***
		(0.063)				(0.060)
Belief uneq. countr. worse			0.422***			0.249***
			(0.063)			(0.054)
Neg. externality belief			0.622***			0.217***
			(0.063)			(0.057)
Leans Republican				-0.458***		-0.292***
				(0.089)		(0.079)
Sanders/Harris supporter				0.581***		0.334***
				(0.089)		(0.081)
Trusts the government					0.216***	0.029
					(0.040)	(0.036)
Taxation reduces work					-0.092***	-0.004
					(0.028)	(0.021)
Controls	No	No	No	No	No	No
Only Control Group	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R2	0.104	0.317	0.183	0.249	0.046	0.443
Observations	932.000	932.000	932.000	932.000	932.000	932.000

Note. This table reports results from a regression of different redistributive preference outcomes on fairness views, political views, externality beliefs and attitudes towards the government, as well as socio-economic control variables. Standard errors are in parentheses. *Significance levels*: *10%, **5%, ***1%.

Table I50: Predictive power of various beliefs using RP-index based on non-dichotomized variables

	(1)	(2)	(3)	(4)	(5)	(6)
	RP Index	RP Index	RP Index	RP Index	RP Index	RP Index
	b/se	b/se	b/se	b/se	b/se	b/se
Rich b/c hard work		-0.657***				-0.388***
		(0.063)				(0.055)
Society unfair (post)		0.280***				0.214***
		(0.025)				(0.022)
Belief uneq. countr. worse			0.254***			0.142***
			(0.031)			(0.024)
General neg. ext.			0.265***			0.077***
			(0.031)			(0.025)
Leans Republican				-0.374***		-0.149**
				(0.084)		(0.065)
Sanders/Harris supporter				0.690***		0.359***
				(0.085)		(0.068)
Trusts the government					0.322***	0.156***
					(0.039)	(0.032)
Taxation reduces work					-0.120***	-0.036*
					(0.028)	(0.020)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Only Control Group	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R2	0.114	0.455	0.323	0.350	0.207	0.593
Observations	932.000	932.000	897.000	932.000	932.000	897.000

Note. This table reports results from a regression of different redistributive preference outcomes on fairness views, political views, externality beliefs and attitudes towards the government, as well as socio-economic control variables. Controls not listed include gender, race, income-group, age-group, education, employment status, geographic region. Standard errors are in parentheses. *Significance levels*: *10%, **5%, ***1%.

Table I51: Treatment effects with FDR sharpened q-values

	(1)	(2)	(3)	(4)	(5)
	RP Index	Wants redistribution	Increase top taxes	Gov. reduce ineq.	Ineq. is serious issue
Crime Ext. Tr.	0.037	0.031	-0.005	0.007	0.020
p-value	(.308)	(.127)	(.817)	(.705)	(.313)
q-value	(.288)	(.147)	(.610)	(.597)	(.288)
Trust Ext. Tr.	0.043	0.006	0.004	0.036*	0.017
p-value	(.244)	(.800)	(.842)	(.075)	(.407)
q-value	(.256)	(.610)	(.610)	(.091)	(.324)
Full Ext. Tr.	0.107***	0.050**	-0.012	0.048**	0.069***
p-value	(.004)	(.019)	(.572)	(.018)	(.001)
q-value	(.011)	(.032)	(.475)	(.032)	(.004)
Fairness Tr.	0.208***	0.052**	0.065***	0.067***	0.115***
p-value	(.000)	(.015)	(.002)	(.001)	(.000)
q-value	(.001)	(.032)	(.007)	(.004)	(.001)
Controls	Yes	Yes	Yes	Yes	Yes
Observations	4371.000	4371.000	4371.000	4371.000	4371.000

Note. This table reports FDR sharpened q-values from the regression in Table 1. p-values and q-values are in parentheses. *Significance levels*: *10%, **5%, ***1%.

Table I52: Respondents' belief about the survey bias by treatment group

	Right-Wing Bias (%)	No Bias (%)	Left-Wing Bias (%)
Crime tr.	5.68	71.49	22.83
Trust tr.	5.21	73.45	21.33
Full ext tr.	7.66	70.33	22.00
Fairness tr.	6.19	70.87	22.94
Control (passive)	7.81	73.98	18.22
Control (active)	6.85	72.84	20.30

Appendix III.

Appendix to Chapter Three

III.A. Question phrasing

III.A.1. Survey 1 (Elicitation)

III.A.1.1. Externality argument elicitation

Question text: Pro-redistribution externality elicitation

Imagine you want to convince a friend to support **more** economic redistribution with an argument about how economic inequality has **negative consequences** for society. Please write a **brief** (3 sentences maximum) argument below.

Please do not discuss economic fairness issues, but instead focus your argument on how inequality affects societies in other ways. You can for example make arguments for redistribution about how economic inequality affects the amount of [two of crime, economic growth, corruption, innovation, social unrest, trust, political polarization], or society overall – but please use your own words and ideas.

Remember that convincing arguments will be rewarded – if your arguments are found to be convincing, your survey payout will be doubled.

So, why should we redistribute more?

Question text: Anti-redistribution externality elicitation

Imagine you want to convince a friend to support **less** economic redistribution with an argument about how economic inequality has **positive consequences** for society. Please write a **brief** (3 sentences maximum) argument below.

Please do not discuss economic fairness issues, but instead focus your argument on how inequality affects societies in other ways. You can for example make arguments against redistribution about how economic inequality affects the amount of economic growth, innovation, or society overall – but please use your own words and ideas.

Remember that convincing arguments will be rewarded – if your arguments are found to be convincing, your survey payout will be doubled.

So, why should we redistribute less?

III.A.1.2. Fairness argument elicitation

Question text: Pro-redistribution fairness elicitation

Imagine you want to convince a friend to support **more** economic redistribution with an argument about how this would **be fair**. Please write a **brief** (3 sentences maximum) argument below.

You can make any argument you want as long as it relates to economic fairness issues (high incomes, low incomes, which people deserve income increases, and so on). You don't need to explicitly use the word "fair" unless you want to, but the argument should be about fairness.

Remember that convincing arguments will be rewarded – if your arguments are found to be convincing, your survey payout will be doubled.

So, why should we redistribute more?

Question text: Anti-redistribution fairness elicitation

Imagine you want to convince a friend to support **less** economic redistribution with an argument about how this would **be fair**. Please write a **brief** (3 sentences maximum) argument below.

You can make any argument you want as long as it relates to economic fairness issues (high incomes, low incomes, which people deserve income increases, and so on). You don't need to explicitly use the word "fair" unless you want to, but the argument should be about fairness.

Remember that convincing arguments will be rewarded – if your arguments are found to be convincing, your survey payout will be doubled.

So, why should we redistribute less?

III.A.2. Survey 2 (Quality check)

III.A.2.1. Introduction

In this survey we want you to tell us whether some arguments are **sensible** and **on-topic**.

You will see 16 arguments written by other survey respondents.

These arguments should all be about either **increasing** or **decreasing** the economic differences between people. We want you to tell us:

1. Whether the argument is on this topic and makes sense, and

2. Whether the argument is about **fairness**, **any other consequences of inequality on society**, or **neither**.

"Fairness" arguments could for example be about who deserves more or less income, whether taxation is fair, whether every person deserves a living wage, and so on.

"Other consequences of inequality on society" arguments could for example be about how more economic inequality affects the amount of crime, economic growth, social unrest, and so on. (Note that even though statements such as "inequality increases crime" has some fairness aspect to it, you should consider this as a consequence-argument.)

III.A.2.2. Argument-specific text: Sensibility

This argument should be arguing for [**less/more**] economic redistribution:

[*Argument text*]

We want to make sure that the argument is on the right topic and is possible to understand. Please be lenient and **ignore whether you agree with the argument**.

Does this argument make sense at all, given the topic?

· Yes

· No

III.A.2.3. Argument-specific text: Topic

Which describes this argument better:

· This is an argument about fairness ideas (whether people deserve the incomes they receive)

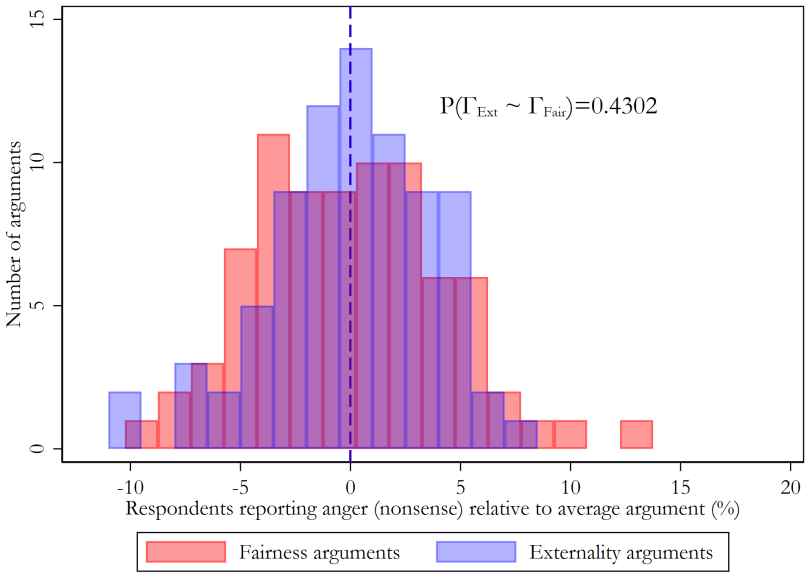
· This is an argument about how economic inequality changes something in society (for example crime, economic growth, or the political process)

· Neither of the two options above fit at all

III.B. Graphs

III.B.1. Anger (nonsense)

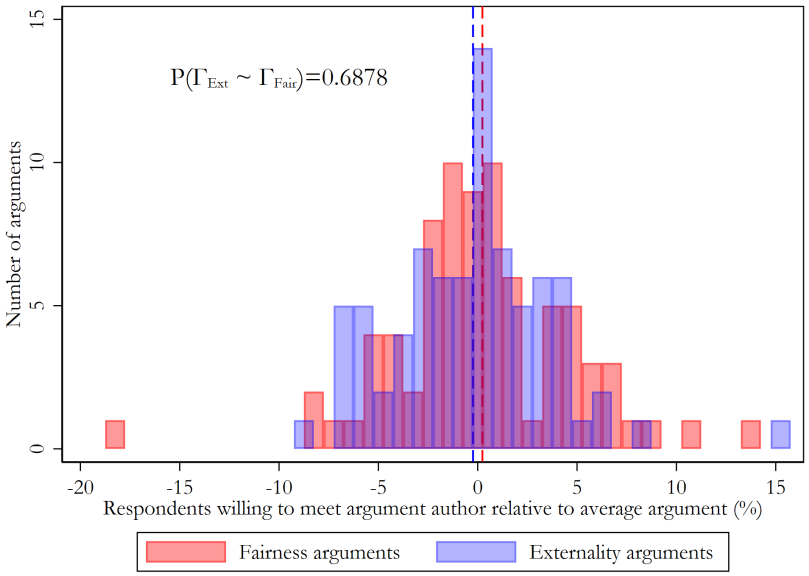
Figure B1: Self-reported "anger or agitation" due to agreement across argument type



Note. The percentage of individuals responding "Yes, because I think the argument is nonsense" or "Partly, because I think the argument is nonsense" to a question about whether "*a discussion about this argument could provoke an emotional reaction like anger or agitation in you*". There are 160 arguments, and each argument was viewed by an average of 202 respondents. In total there are 32,300 observations.

III.B.2. Having a longer conversation

Figure B2: Willingness to meet the person who wrote the argument



Note. The percentage of individuals responding "Yes" to a question about whether they would "*be willing to have a longer conversation with this person about these ideas?*". There are 160 arguments, and each argument was viewed by an average of 202 respondents. In total there are 32,300 observations.

III.C. Tables

Table C1: Anger because respondent agreed with the argument: Regression results

	(1) dAngerAgree b/se	(2) dAngerAgree b/se	(3) dAngerAgree b/se	(4) dAngerAgree b/se	(5) dAngerAgree b/se
ExtArg	-0.019***	-0.019***	0.002	-0.025***	-0.001
	(0.005)	(0.005)	(0.008)	(0.006)	(0.008)
Emotions			0.027***		0.033***
			(0.008)		(0.008)
Factual				0.025***	0.030***
				(0.007)	(0.007)
Controls	No	Yes	Yes	Yes	Yes
R2	0.00	0.04	0.04	0.04	0.04
Observations	32300	32300	32300	32300	32300

Note. This table represents the regression coefficients for the pre-specified "anger because agree" regression, with additional regressions including dummies for whether the argument was positive or emotional. Note that this outcome is a subset of the outcome in Table 2. Controls are binary variables for leaning Republican over Democrat, gender, self-identifying as black, self-identifying as non-white, four income groups ($0-$25,000, $25,000-$50,000, $50,000-$100,000, $100,000+), six age groups (20-29, 30-39, 40-49, 50-59, 60-69, 70+), having a college education, being unemployed, not being in the work force (e.g. students or seniors), and region (South, West, Northeast, Midwest). *Significance levels*: *10%, **5%, ***1%.

Table C2: Anger because respondent thinks the argument is nonsense: Regression results

	(1) dAngerNonsense b/se	(2) dAngerNonsense b/se	(3) dAngerNonsense b/se	(4) dAngerNonsense b/se	(5) dAngerNonsense b/se
ExtArg	0.000	0.000	-0.003	0.004	-0.002
	(0.005)	(0.005)	(0.007)	(0.005)	(0.007)
Emotions			-0.004		-0.008
			(0.007)		(0.008)
Factual				-0.015**	-0.016**
				(0.006)	(0.006)
Controls	No	Yes	Yes	Yes	Yes
R2	0.00	0.03	0.03	0.03	0.03
Observations	32300	32300	32300	32300	32300

Note. This table represents the regression coefficients for the pre-specified "anger because nonsense" regression, with additional regressions including dummies for whether the argument was positive or emotional. Note that this outcome is a subset of the outcome in Table 2. Controls are binary variables for leaning Republican over Democrat, gender, self-identifying as black, self-identifying as non-white, four income groups ($0-$25,000, $25,000-$50,000, $50,000-$100,000, $100,000+), six age groups (20-29, 30-39, 40-49, 50-59, 60-69, 70+), having a college education, being unemployed, not being in the work force (e.g. students or seniors), and region (South, West, Northeast, Midwest). *Significance levels*: *10%, **5%, ***1%.

Bibliography

Rolf Aaberge. Characterizations of Lorenz Curves and Income Distributions. *Social Choice and Welfare*, 17(4):639–653, 2000.

Alberto Alesina and Paola Giuliano. Chapter 4 - Preferences for Redistribution. In Jess Benhabib, Alberto Bisin, and Matthew O. Jackson, editors, *Handbook of Social Economics*, volume 1, pages 93–131. North-Holland, 2011. doi: 10.1016/B978-0-444-53187-2.00004-8.

Alberto Alesina, Edward L Glaeser, and Bruce Sacerdote. Why Doesn't the United States Have a European-Style Welfare State? *Brookings Papers on Economic Activity*, 2001(2):187–277, 2001.

Alberto Alesina, Elie Murard, and Hillel Rapoport. Immigration and Attitudes toward Redistribution in Europe. *Mimeo*, pages 1–85, 2018a.

Alberto Alesina, Stefanie Stantcheva, and Edoardo Teso. Intergenerational Mobility and Preferences for Redistribution. *American Economic Review*, 108(2):521–554, 2018b. ISSN 0002-8282. doi: 10.1257/aer.20162015.

Alberto Alesina, Armando Miano, and Stefanie Stantcheva. Immigration and Redistribution. *The Review of Economic Studies*, 90(1):1–39, 2023.

Ingvild Almås, 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. ISSN 0022-3808, 1537-534X. doi: 10.1086/705551.

Ingvild Almås, Alexander W Cappelen, Erik Ø Sørensen, and Bertil Tungodden. Global Evidence on the Selfish Rich Inequality Hypothesis. *Proceedings of the National Academy of Sciences*, 119(3), 2022.

Michael L Anderson. Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of the American Statistical Association*, 103(484):1481–1495, 2008.

Peter Andre, Carlo Pizzinelli, Christopher Roth, and Johannes Wohlfart. Subjective Models of the Macroeconomy: Evidence From Experts and Representative Samples. *The Review of Economic Studies*, 2022. ISSN 0034-6527. doi: 10.1093/restud/rdac008.

Aristotle. *Politics, Book IV*. Clarendon Press, 1885.

Thomas Aronsson and Olof Johansson-Stenman. When the Joneses' Consumption Hurts: Optimal Public Good Provision and Nonlinear Income Taxation. *Journal of Public Economics*, 92(5-6):986–997, 2008.

Thomas Aronsson and Olof Johansson-Stenman. Keeping up with the Joneses, the Smiths and the Tanakas: On International Tax Coordination and Social Comparisons. *Journal of Public Economics*, 131:71–86, 2015.

Thomas Aronsson and Olof Johansson-Stenman. Inequality Aversion and Marginal Income Taxation. *Proceedings. Annual Conference on Taxation and Minutes of the Annual Meeting of the National Tax Association*, 111:1–32, 2018a.

Thomas Aronsson and Olof Johansson-Stenman. Paternalism against Veblen: Optimal Taxation and Non-Respected Preferences for Social Comparisons. *American Economic Journal: Economic Policy*, 10(1):39–76, 2018b.

Thomas Aronsson and Olof Johansson-Stenman. Inequality Aversion and Marginal Income Taxation. In *Proceedings. Annual Conference on Taxation and Minutes of the Annual Meeting of the National Tax Association*, volume 111, pages 1–32. National Tax Association, 2018c.

Thomas Aronsson and Olof Johansson-Stenman. Optimal Second-Best Taxation When Individuals Have Social Preferences. *Umeå Economic Studies*, 973, 2020.

Anthony B Atkinson. On the Measurement of Inequality. *Journal of Economic Theory*, 2(3): 244–263, 1970.

Anthony B Atkinson. *Inequality–What Can Be Done?* Harvard University Press, 2014.

Anthony Barnes Atkinson and Joseph E Stiglitz. The Design of Tax Structure: Direct versus Indirect Taxation. *Journal of Public Economics*, 6(1-2):55–75, 1976.

Adrien Auclert and Matthew Rognlie. Inequality and Aggregate Demand. *National Bureau of Economic Research*, 24280, 2018.

Gary S Becker. Crime and Punishment: An Economic Approach. In *The Economic Dimensions of Crime*, pages 13–68. Springer, 1968.

Luna Bellani, Nona Bledow, Marius R Busemeyer, and Guido Schwerdt. When Everyone Thinks they're Middle-Class:(Mis-) Perceptions of Inequality and why they Matter for Social Policy. *Policy Papers / Cluster of Excellence 'The Politics of Inequality'*, 6, 2021.

Roland Benabou. Inequality and Growth. *National Bureau of Economic Research Macroeconomics Annual*, 11:11–74, 1996.

Roland Bénabou and Efe A Ok. Social Mobility and the Demand for Redistribution: The POUM Hypothesis. *The Quarterly Journal of Economics*, 116(2):447–487, 2001.

Andreas Bergh, Therese Nilsson, and Daniel Waldenström. *Sick of Inequality?: An Introduction to the Relationship between Inequality and Health.* Edward Elgar Publishing, 2016. ISBN 978-1-78536-421-1.

Marcelo Bergolo, Gabriel Burdin, Santiago Burone, Mauricio De Rosa, Matias Giaccobasso, and Martin Leites. Dissecting Inequality-averse Preferences. *Journal of Economic Behavior & Organization*, 200:782–802, 2022.

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

Michael J Boskin and Eytan Sheshinski. Optimal Redistributive Taxation when Individual Welfare Depends upon Relative Income. *The Quarterly Journal of Economics*, pages 589–601, 1978.

François Bourguignon. Crime as a Social Cost of Poverty and Inequality: A Review Focusing on Developing Countries. *Revista Desarrollo y Sociedad*, 44:61–99, 1999.

François Bourguignon and Amedeo Spadaro. Tax-benefit Revealed Social Preferences. *The Journal of Economic Inequality*, 10:75–108, 2012.

Ary Lans Bovenberg and Frederick van der Ploeg. Environmental Policy, Public Finance and the Labour Market in a Second-Best World. *Journal of Public Economics*, 55(3):349–390, 1994.

Emily Breza, Supreet Kaur, and Yogita Shamdasani. The Morale Effects of Pay Inequality. *The Quarterly Journal of Economics*, 133(2):611–663, 2018.

James M Buchanan and William C Stubblebine. Externality. In *Classic Papers in Natural Resource Economics*, pages 138–154. Springer, 1962.

Thomas Buser, Gianluca Grimalda, Louis Putterman, and Joël van der Weele. Overconfidence and Gender Gaps in Redistributive Preferences: Cross-Country Experimental Evidence. *Journal of Economic Behavior & Organization*, 178:267–286, October 2020. ISSN 01672681. doi: 10.1016/j.jebo.2020.07.005.

Alexander Cappelen, John List, Anya Samek, and Bertil Tungodden. The Effect of Early-Childhood Education on Social Preferences. *Journal of Political Economy*, 128(7):2739–2758, July 2020. ISSN 0022-3808, 1537-534X. doi: 10.1086/706858.

Alexander W Cappelen, 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.

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

David Card, 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.

Fredrik Carlsson, Dinky Daruvala, and Olof Johansson-Stenman. Are People Inequality-Averse, or Just Risk-Averse? *Economica*, 72(287):375–396, 2005.

Mark A Cohen, Roland T Rust, Sara Steen, and Simon T Tidd. Willingness-To-Pay for Crime Control Programs. *Criminology : an interdisciplinary journal*, 42(1):89–110, 2004.

David J. Cooper and John H. Kagel. Other-Regarding Preferences: A Selective Survey of Experimental Results. In John H Kagel and Alvin E Roth, editors, *The Handbook of Experimental Economics, Vol. 2*, pages 2017–2282. Princeton University Press, Princeton, 2016.

Frank A Cowell. Measurement of Inequality. *Handbook of Income Distribution*, 1:87–166, 2000.

Helmuth Cremer, Firouz Gahvari, and Norbert Ladoux. Externalities and Optimal Taxation. *Journal of Public Economics*, 70(3):343–364, 1998.

Guillermo Cruces, 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.

Peter A Diamond. Optimal Income Taxation: An Example With a U-shaped Pattern of Optimal Marginal Tax Rates. *American Economic Review*, pages 83–95, 1998.

Peter A Diamond and James A Mirrlees. Optimal Taxation and Public Production II: Tax Rules. *The American Economic Review*, 61(3):261–278, 1971.

David Donaldson and John A Weymark. A Single-Parameter Generalization of the Gini Indices of Inequality. *Journal of Economic Theory*, 22(1):67–86, 1980.

Arindrajit Dube. Making the Case for a Higher Minimum Wage, 2019.

Ruben Durante, Louis Putterman, and Joël van der Weele. Preferences for Redistribution and Perception of Fairness: An Experimental Study. *Journal of the European Economic Association*, 12(4):1059–1086, 2014.

Thomas Epper, Ernst Fehr, and Julien Senn. Other-Regarding Preferences and Redistributive Politics. *Working Paper*, 2020. ISSN 1556-5068. doi: 10.2139/ssrn.3526809.

Pablo Fajnzylber, Daniel Lederman, and Norman Loayza. Inequality and Violent Crime. *The Journal of Law and Economics*, 45(1):1–39, April 2002. ISSN 0022-2186, 1537-5285. doi: 10.1086/338347.

Daniel Feenberg and Elisabeth Coutts. An Introduction to the TAXSIM Model. *Journal of Policy Analysis and management*, 12(1):189–194, 1993.

Dietmar Fehr, Daniel Müller, and Marcel Preuss. Social Mobility Perceptions and Inequality Acceptance. *Working Paper*, 2020a.

Dietmar Fehr, Hannes Rau, Stefan T. Trautmann, and Yilong Xu. Inequality, Fairness and Social Capital. *European Economic Review*, 129:103566, 2020b. ISSN 0014-2921. doi: 10.1016/j.euroecorev.2020.103566.

Ernst Fehr and Klaus M Schmidt. A Theory of Fairness, Competition, and Cooperation. *The Quarterly Journal of Economics*, 114(3):817–868, 1999.

Raymond Fisman, Pamela Jakiela, Shachar Kariv, and Daniel Markovits. The Distributional Preferences of an Elite. *Science (New York, N.Y.)*, 349(6254), 2015. ISSN 10959203. doi: 10.1126/science.aab0096.

Raymond Fisman, Ilyana Kuziemko, and Silvia Vannutelli. Distributional Preferences in Larger Groups: Keeping up with the Joneses and Keeping Track of the Tails. *Journal of the European Economic Association*, 19(2):1407–1438, 2021.

Aina Gallego. Inequality and the Erosion of Trust among the Poor: Experimental Evidence. *Socio-Economic Review*, 14(3):443–460, July 2016. ISSN 1475-1461, 1475-147X. doi: 10.1093/ser/mww010.

Manja Gärtner, Johanna Mollerstrom, and David Seim. Individual Risk Preferences and the Demand for Redistribution. *Journal of Public Economics*, 153:49–55, September 2017. ISSN 00472727. doi: 10.1016/j.jpubeco.2017.06.009.

Manja Gärtner, Johanna Möllerström, and David Seim. Income Mobility, Luck/Effort Beliefs, and the Demand for Redistribution: Perceptions and Reality. *Working Paper*, 2019.

Stephane Gauthier and Guy Laroque. Separability and Public Finance. *Journal of Public Economics*, 93(11-12):1168–1174, 2009.

Jonah B Gelbach. When do Covariates Matter? And Which Ones, and How Much? *Journal of Labor Economics*, 34(2):509–543, 2016.

Martin Gilens. *Affluence and Influence: Economic Inequality and Political Power in America*. Princeton University Press, 2012.

Robert E Goodin. Laundering Preferences. *Foundations of Social Choice Theory*, 75:81–86, 1986.

Alan Greenspan. Comments: National Association of Business Economists Conference, 2014.

António Guterres. Secretary-General's Nelson Mandela Lecture: "Tackling the Inequality Pandemic: A New Social Contract for a New Era", July 2020.

Fatih Guvenen, Gueorgui Kambourov, Burhanettin Kuruscu, Sergio Ocampo-Diaz, and Daphne Chen. Use It or Lose It: Efficiency Gains from Wealth Taxation. Working Paper 26284, National Bureau of Economic Research, 2019.

Ingar Haaland and Christopher Roth. Labor Market Concerns and Support for Immigration. *Journal of Public Economics*, 191:104256, 2020.

Ingar Haaland and Christopher Roth. Beliefs about Racial Discrimination and Support for Pro-Black Policies. *The Review of Economics and Statistics*, 105(1):1–38, March 2021. ISSN 0034-6535, 1530-9142. doi: 10.1162/rest_a_01036.

Ingar Haaland, Christopher Roth, and Johannes Wohlfart. Designing Information Provision Experiments. *Journal of Economic Literature*, Forthcoming.

Stefan Harrendorf. Prospects, Problems, and Pitfalls in Comparative Analyses of Criminal Justice Data. *Crime and Justice*, 47(1):159–207, 2018.

John C Harsanyi. Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility. *Journal of Political Economy*, 63(4):309–321, 1955.

John C. Harsanyi. Morality and the Theory of Rational Behavior. *Social Research*, 44(4): 623–656, 1977.

Orestes P Hastings. Less Equal, Less Trusting? Longitudinal and Cross-sectional Effects of Income Inequality on Trust in US States, 1973–2012. *Social Science Research*, 74:77–95, 2018.

Jonathan Heathcote, Kjetil Storesletten, and Giovanni L Violante. Presidential Address 2019: How Should Tax Progressivity Respond to Rising Income Inequality? *Journal of the European Economic Association*, 18(6):2715–2754, 2020.

James J Heckman. The Economics of Inequality: The Value of Early Childhood Education. *American Educator*, 35(1):31, 2011.

Nathaniel Hendren. Measuring Economic Efficiency using Inverse-Optimum Weights. *Journal of Public Economics*, 187:104198, 2020.

Albert O Hirschman and Michael Rothschild. The Changing Tolerance for Income Inequality in the Course of Economic Development* With A Mathematical Appendix. *The Quarterly Journal of Economics*, 87(4):544–566, 1973.

Hilary Hoynes, Jesse Rothstein, and Krista Ruffini. Making Work Pay Better Through an Expanded Earned Income Tax Credit, 2017.

Kristoffer B Hvidberg, Claus T Kreiner, and Stefanie Stantcheva. Social Position and Fairness Views. *Review of Economic Studies*, 2022.

Bas Jacobs. The Marginal Cost of Public Funds is One at the Optimal Tax System. *International Tax and Public Finance*, 25:883–912, 2018.

Bas Jacobs and Ruud A De Mooij. Pigou meets Mirrlees: On the Irrelevance of Tax Distortions for the Second-best Pigouvian Tax. *Journal of Environmental Economics and Management*, 71:90–108, 2015.

Bas Jacobs, Egbert LW Jongen, and Floris T Zoutman. Revealed Social Preferences of Dutch Political Parties. *Journal of Public Economics*, 156:81–100, 2017.

Laurence Jacquet and Etienne Lehmann. How to Tax Different Incomes? *CEPR Discussion Paper Series no. 16571, IZA Institute of Labor Economics*, 2021.

Laurence Jacquet, Etienne Lehmann, and Bruno Van der Linden. Optimal Redistributive Taxation with Both Extensive and Intensive Responses. *Journal of Economic Theory*, 148(5): 1770–1805, 2013.

Charles I Jones. Taxing Top Incomes in a World of Ideas. *Journal of Political Economy*, 130 (9):2227–2274, 2022.

Ravi Kanbur and Matti Tuomala. Relativity, Inequality, and Optimal Nonlinear Income Taxation. *International Economic Review*, 54(4):1199–1217, 2013.

Ravi Kanbur, Michael Keen, and Matti Tuomala. Optimal Non-Linear Income Taxation for the Alleviation of Income-Poverty. *European Economic Review*, 38(8):1613–1632, 1994.

Louis Kaplow. *The Theory of Taxation and Public Economics*. Princeton University Press, 2010.

Mounir Karadja, Johanna Mollerstrom, and David Seim. Richer (and Holier) Than Thou? The Effect of Relative Income Improvements on Demand for Redistribution. *The Review of Economics and Statistics*, 99(2):201–212, May 2017. ISSN 0034-6535, 1530-9142. doi: 10.1162/REST_a_00623.

Morgan Kelly. Inequality and Crime. *Review of Economics and Statistics*, 82(4):530–539, 2000.

Henrik Jacobsen Kleven, Claus Thustrup Kreiner, and Emmanuel Saez. The Optimal Income Taxation of Couples. *Econometrica*, 77(2):537–560, 2009.

Wojciech Kopczuk. A Note on Optimal Taxation in the Presence of Externalities. *Economics Letters*, 80(1):81–86, 2003.

Ilyana Kuziemko, Michael I Norton, Emmanuel Saez, and Stefanie Stantcheva. How Elastic Are Preferences for Redistribution? Evidence from Randomized Survey Experiments. *The American Economic Review*, 105(4):1478–1508, 2015.

Arthur B Laffer. The Laffer Curve: Past, Present, and Future. *Backgrounder (Washington, D.C.)*, 1765:1–16, 2004.

Etienne Lehmann, Laurent Simula, and Alain Trannoy. Tax Me If You Can! Optimal Nonlinear Income Tax Between Competing Governments. *The Quarterly Journal of Economics*, 129(4): 1995–2030, 2014.

Max Lobeck and Morten Nyborg Støstad. The Consequences of Inequality: Beliefs and Redistributive Preferences. *Working Paper*, 2023.

Benjamin B Lockwood and Matthew Weinzierl. Positive and Normative Judgments Implicit in US Tax Policy, and the Costs of Unequal Growth and Recessions. *Journal of Monetary Economics*, 77:30–47, 2016.

Stefan Lollivier and Jean-Charles Rochet. Bunching and Second-Order Conditions: A Note on Optimal Tax Theory. *Journal of Economic Theory*, 31(2):392–400, 1983.

Robert E. Jr. Lucas. The Industrial Revolution: Past and Future, September 2004.

N Gregory Mankiw, Matthew Weinzierl, and Danny Yagan. Optimal Taxation in Theory and Practice. *Journal of Economic Perspectives*, 23(4):147–74, 2009.

Alan Manning. Top Rate of Income Tax. *Centre for Economic Performance's Election Analysis*, 2015.

Ruben Berge Mathisen, Wouter Schakel, Svenja Hense, Lea Elsässer, Mikael Persson, and Jonas Pontusson. Unequal Responsiveness and Government Partisanship in Northwest Europe. *Unequal Democracies: Working Papers*, 31, 2021.

Allan H. Meltzer and Scott F. Richard. A Rational Theory of the Size of Government. *Journal of Political Economy*, 89(5):914–927, 1981. ISSN 00223808, 1537534X.

Atif R Mian, Ludwig Straub, and Amir Sufi. The Saving Glut of the Rich. Working Paper 26941, National Bureau of Economic Research, 2020.

James A Mirrlees. An Exploration in the Theory of Optimum Income Taxation. *The Review of Economic Studies*, 38(2):175–208, 1971.

Michael I Norton and Dan Ariely. Building a Better America—One Wealth Quintile at a Time. *Perspectives on Psychological Science*, 6(1):9–12, 2011.

B. Obama. Remarks by the President on the Economy in Osawatomie, Kansas, 2011.

OECD. International tax reform: OECD Releases Technical Guidance for Implementation of the Global Minimum Tax, 2023.

Arthur M. Okun. *Equality and Efficiency: The Big Tradeoff*. Brookings Institution Press, 1975. ISBN 978-0-8157-2654-8.

Andrew J Oswald. Altruism, Jealousy and the Theory of Optimal Non-Linear Taxation. *Journal of Public Economics*, 20(1):77–87, 1983.

Mats Persson. Why are Taxes so High in Egalitarian Societies? *The Scandinavian Journal of Economics*, pages 569–580, 1995.

The Pew Research Center. Middle Easterners See Religious and Ethnic Hatred as Top Global Threat, October 2014.

Thomas Piketty. *Capital in the Twenty-First Century*. Harvard University Press, Cambridge, 2014.

Thomas Piketty, Emmanuel Saez, and Stefanie Stantcheva. Optimal Taxation of Top Labor Incomes: A Tale of Three Elasticities. *American Economic Journal: Economic Policy*, 6(1): 230–71, 2014.

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

Jukka Pirttilä and Matti Tuomala. Income Tax, Commodity Tax and Environmental Policy. *International Tax and Public Finance*, 4:379–393, 1997.

Plato. *The Laws of Plato*. CreateSpace Independent Publishing Platform, 2016.

Plutarch. *Parallel Lives, Solon*. Loeb Classical Library, 1923.

Pope Francis. @Pontifex Twitter, April 2014.

John Rawls. *A Theory of Justice*. Harvard University Press, Cambridge, 1971.

Paul R. Rosenbaum. Multiple Control Groups. In *Observational Studies*, pages 253–275. Springer New York, New York, NY, 2002.

Christopher Roth and Johannes Wohlfart. Experienced Inequality and Preferences for Redistribution. *Journal of Public Economics*, 167:251–262, November 2018. ISSN 0047-2727. doi: 10.1016/j.jpubeco.2018.09.012.

Christopher Roth, Sonja Settele, and Johannes Wohlfart. Beliefs about Public Debt and the Demand for Government Spending. *Working Paper*, page 89, 2020.

David Rueda and Daniel Stegmueller. The Externalities of Inequality: Fear of Crime and Preferences for Redistribution in Western Europe. *American Journal of Political Science*, 60 (2):472–489, April 2016. ISSN 00925853. doi: 10.1111/ajps.12212.

Hector Rufrancos, Madeleine Power, Kate E Pickett, and Richard Wilkinson. Income Inequality and Crime: A Review and Explanation of the Time Series Evidence. *Sociology and Criminology-Open Access*, 2013.

Efraim Sadka. On Income Distribution, Incentive Effects and Optimal Income Taxation. *The Review of Economic Studies*, 43(2):261–267, 1976.

Emmanuel Saez. Using Elasticities to Derive Optimal Income Tax Rates. *The Review of Economic Studies*, 68(1):205–229, 2001.

Emmanuel Saez. Optimal Income Transfer Programs: Intensive Versus Extensive Labor Supply Responses. *The Quarterly Journal of Economics*, 117(3):1039–1073, 2002.

Emmanuel Saez. Do Taxpayers Bunch at Kink Points? *American Economic Journal: Economic Policy*, 2(3):180–212, 2010.

Emmanuel Saez and Stefanie Stantcheva. Generalized Social Marginal Welfare Weights for Optimal Tax Theory. *American Economic Review*, 106(1):24–45, 2016.

Emmanuel Saez and Stefanie Stantcheva. A Simpler Theory of Optimal Capital Taxation. *Journal of Public Economics*, 162:120–142, 2018.

Emmanuel Saez and Gabriel Zucman. *The Triumph of Injustice: How the Rich Dodge Taxes and How to Make Them Pay*. WW Norton & Company, 2019.

Emmanuel Saez and Gabriel Zucman. The Rise of Income and Wealth Inequality in America: Evidence from Distributional Macroeconomic Accounts. *Journal of Economic Perspectives*, 34(4):3–26, November 2020. ISSN 0895-3309. doi: 10.1257/jep.34.4.3.

Emmanuel Saez and Gabriel Zucman. A Wealth Tax on Corporations' Stock. *Economic Policy*, 37(110):213–227, 2022.

Emmanuel Saez, Joel Slemrod, and Seth H Giertz. The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review. *Journal of Economic Literature*, 50(1): 3–50, 2012.

Bernard Salanie. *The Economics of Taxation*. MIT press, 2011.

Agnar Sandmo. Optimal Taxation in the Presence of Externalities. *The Swedish Journal of Economics*, pages 86–98, 1975.

Ulrich Schmidt and Philipp C Wichardt. Inequity Aversion, Welfare Measurement and the Gini Index. *Social Choice and Welfare*, 52(3):585–588, 2019.

Jesus K Seade. On the Shape of Optimal Tax Schedules. *Journal of Public Economics*, 7(2): 203–235, 1977.

Amartya Sen. Real National Income. *The Review of Economic Studies*, 43(1):19–39, 1976.

Amartya Sen. Utilitarianism and Welfarism. *The Journal of Philosophy*, 76(9):463–489, 1979. ISSN 0022-362X. doi: 10.2307/2025934.

Laurent Simula and Alain Trannoy. Gini and Optimal Income Taxation by Rank. *American Economic Journal: Economic Policy*, 14(3):352–379, 2022.

Stefanie Stantcheva. Understanding Tax Policy: How Do People Reason? *The Quarterly Journal of Economics*, 136(4):2309–2369, 2021.

Joseph E Stiglitz. Self-selection and Pareto Efficient Taxation. *Journal of Public Economics*, 17 (2):213–240, 1982.

Morten Nyborg Støstad. Inequality as an Externality: Existence and Consequences. *HAL Open Science*, dumas-02407577, 2019.

Morten Nyborg Støstad and Frank Cowell. Inequality as an Externality: Consequences for Tax Design. *SSRN Electronic Journal*, 4185685, 2021. doi: 10.2139/ssrn.2580904.

Lester C Thurow. The Income Distribution as a Pure Public Good. *The Quarterly Journal of Economics*, pages 327–336, 1971.

Aleh Tsyvinski and Nicolas Werquin. Generalized Compensation Principle. Technical report, National Bureau of Economic Research, 2017.

Matti Tuomala. *Optimal Redistributive Taxation*. Oxford University Press, 2016.

Jean-Robert Tyran and Rupert Sausgruber. A Little Fairness May Induce a Lot of Redistribution in Democracy. *European Economic Review*, 50(2):469–485, February 2006. ISSN 00142921. doi: 10.1016/j.euroecorev.2004.09.014.

Vanessa Valero. Redistribution and Beliefs about the Source of Income Inequality. *Experimental Economics*, 25:876–901, 2021.

Bernd Wegener, David Mason, and International Social Justice Project (ISJP). International Social Justice Project, 1991 and 1996, 2010.

Matthew Weinzierl. The Promise of Positive Optimal Taxation: Normative Diversity and a Role for Equal Sacrifice. *Journal of Public Economics*, 118:128–142, 2014.

Martin L Weitzman. On Modeling and Interpreting the Economics of Catastrophic Climate Change. *The Review of Economics and Statistics*, 91(1):1–19, February 2009. ISSN 0034-6535. doi: 10.1162/rest.91.1.1.

Martin King Whyte. Fair Versus Unfair: How do Chinese Citizens View Current Inequalities. *Growing pains: Tensions and opportunity in China's transformation*, pages 305–332, 2010.

Richard Wilkinson and Kate Pickett. *The Spirit Level: Why More Equal Societies Almost Always Do Better*. Allen Lane, 2009.

Richard Wilkinson and Kate Pickett. *The Spirit Level: Why Greater Equality Makes Societies Stronger*. Bloomsbury Publishing USA, 2011.

Martin Wolf. Review of 'Capital in the Twenty-First Century', by Thomas Piketty, April 2014.

Yilong Xu and Ginevra Marandola. The (Negative) Effects of inequality on Social Capital. *Journal of Economic Surveys*, 2022.