UNIVERSES OF ARGUMENTS: MAPPING THE REDISTRIBUTIVE DEBATES IN NORWAY AND THE UNITED STATES

MORTEN NYBORG STØSTAD
MAX LOBECK
CHLOÉ DE MEULENAER

WORKING PAPER N°2026/01

JANUARY 2026

WORLD INEQUALITY LAB

Universes of Arguments: Mapping the Redistributive Debates in Norway and the United States*

Morten Nyborg Støstad† (r) Max Lobeck‡ (r) Chloé de Meulenaer§

Click Here for latest version

Abstract

We use natural language processing to classify all pro-redistributive speeches in the U.S. Congress and Norwegian *Storting* from 2015 to 2022, contrasting arguments based on *fairness* with those based on the negative societal consequences of inequality (*inequality externalities*). Fairness-based arguments are markedly more emotional—especially expressing anger and compassion—while externality-based arguments are more analytical and frequently draw on empirical evidence. Pro-redistributive arguments in the U.S. primarily focus on fairness, whereas externality concerns are a central feature of the Norwegian debate. In an experiment, both types of arguments are similarly convincing to U.S. survey respondents, although a preference for fairness is linked to lower educational attainment. Our results document two distinct ways to argue for redistribution, whose prevalence differs across countries, and provide a general framework for comparing the content and evaluations of arguments across domains.

How to cite this working paper: Nyborg Støstad, M., Lobeck, M., De Meulenaer, C., Universes of Arguments: Mapping the Redistributive Debates in Norway and the United States, World Inequality Lab Working Paper 2026/01

*This paper was previously circulated under the title "*Comparing Universes of Redistributive Arguments*". We are grateful to Emmanuel Saez, Marc Fleurbaey, Stéphane Gauthier, Thomas Piketty, Nicolas Jacquemet, Laurence Jacquet, Olof Johansson-Stenman, Alexander Cappelen, Bertil Tungodden, and seminar participants at the Paris School of Economics, the Norwegian School of Economics, and the Max Planck Institute for Research on Collective Goods 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. Author ordering was randomized, denoted by (r), using the AEA Author Randomization Tool. Version: January 5, 2026.

†FAIR Institute, Department of Economics, NHH Norwegian School of Economics, Helleveien 30, 5045 Bergen, Norway. Phone: +4792062450. E-mail: morten.stostad@nhh.no.

‡Federal Ministry of Economic Affairs and Climate Action, Germany. The content of this paper does not necessarily reflect the view of the ministry. E-mail: lobeckmax@gmail.com

§London School of Economics, Houghton St, London WC2A 2AE, United Kingdom. E-mail: c.de-meulenaer@lse.ac.uk.

"*The Norwegian social model requires higher public spending and higher taxes than in many other countries. In return, it provides security [...], trust between people, and a productive economy.*"

Jens Stoltenberg, Finance Minister of Norway, the Norwegian *Storting*, October 2025

1 Introduction

Economic inequality has long been a central issue in politics, and debates about redistribution have shaped major policy choices (Martin et al., 2009). The arguments used in these debates help translate inequality into political action by shaping how voters and legislators think about redistribution. Yet we know surprisingly little about what redistributive arguments look like today, how they differ in content, or how people respond to them.

In this paper, we provide the first large-scale empirical analysis of the contemporary redistributive debate. Using large language models and legislative data, we explore the content of redistributive arguments in the United States and Norway, and measure how citizens evaluate different types of arguments. We primarily compare pro-redistributive arguments grounded in *fairness* and *inequality externalities* across multiple contexts. We define fairness arguments as redistributive arguments based on what people deserve, what is right or wrong, or appeals to justice or moral principles (Konow, 2000; Cappelen et al., 2007; Scheve and Stasavage, 2016). By contrast, inequality externality arguments – exemplified by the Stoltenberg quote above – emphasize inequality's broader societal effects, such as potential impacts on trust, democratic stability, or economic performance (Lobeck and Støstad, 2023; Støstad and Cowell, 2024). These two lines of argument resemble the classic deontological–consequentialist distinction in political philosophy and capture the large majority of pro-redistributive arguments. They are not as well suited to capture anti-redistributive arguments; we thus restrict attention to arguments in favor of redistribution. Importantly, neither argument type is tied to any particular redistributive policy or institutional design.

We begin our analysis by classifying the content of all floor speeches delivered in the U.S. Congress or Norwegian parliament (the *Storting*) from 2015 to 2022. Drawing on recent legislative-speech databases (Aroyehun et al., 2024; Fiva et al., 2025) adapted for cross-country comparability, we apply large language models to identify excerpts advocating for greater redistribution or, equivalently, reduced inequality. We then run multiple classifications on each excerpt, determining whether it contains a fairness-based and/or an externality-based argument, and assessing emotional tone, the type of appeal made, the evidence used, and the framing presented.

We find that fairness arguments tend to be more emotionally charged and divisive across settings, whereas externality arguments appeal more strongly to logic and empirical evidence. In the Congressional speech excerpts, for example, fairness arguments are 22 percentage points more likely than externality arguments to include an explicit emotional appeal, but 16 p.p. less likely to include a logical appeal; nearly all such content differences are strikingly consistent across both methodologies and settings. Anger, compassion, victimization, and villainization are more common in fairness arguments, whereas consensus-seeking language and references to empirical evidence appear more frequently in externality arguments.

We then explore cross-country differences. 37% of redistributive excerpts contain an externality argument in the *Storting*, compared to only 13% in Congress. Roughly half of redistributive excerpts contain fairness arguments in both countries. As relative prevalence can be defined in various ways, we report that externality arguments are two to five times more common in Norway. Overall, pro-redistributive arguments in the U.S. Congress primarily focus on fairness, whereas externality concerns are a central feature of the Norwegian debate. This finding holds across the political spectrum, as the ratio of externality arguments to fairness arguments is higher in every Norwegian party than in either major U.S. party. Within each country, externality-to-fairness ratios are higher among left-leaning than right-leaning parties; still, the Norwegian far-right Progress Party (FrP) is conditionally more likely to use externality arguments than the Democratic Party. In Norway, the centrist governing parties are notable exceptions, relying more heavily on externality arguments than this left-right pattern would predict.

The analysis then turns to how U.S. citizens *evaluate* these argument types. We use three independent online survey collections to (i) gather two universes of redistributive arguments written by survey respondents, (ii) quality-check these arguments, and (iii) collect 32,680 evaluations of the underlying arguments, validated through a complementary causal design. The method intentionally allows all observable and unobservable argument characteristics to vary freely, including their emotional or logical content. The categorization into fairness and externality arguments is defined entirely by survey respondents. Content differences between fairness and externality arguments closely mirror those observed in the legislative data. This consistency indicates that the LLM classification captures meaningful distinctions, and that the two argument types exhibit stable and distinct properties across contexts.

On average, fairness and externality arguments are similarly convincing to U.S. survey respondents. A slight advantage for fairness arguments is not statistically distinguishable from zero, and the 95 percent confidence interval rules out effects larger than 0.067 standard deviations. Two content features operate in opposite directions: the greater anger content of fairness arguments is associated with greater convincingness, whereas their greater villainization content is associated with less convincingness. Overall, neither argument type appears clearly stronger in terms of short-term persuasiveness.

The choice of argument type remains meaningful, however, as respondents are much more likely to report outrage in response to fairness arguments. This heightened outrage is fully explained by the higher anger content of fairness arguments. Consequently, argument types and their usage may shape the emotional climate of redistributive debates—with potential consequences for how redistributive debates evolve over time. In related work, Algan et al. (2025) documents a rise in anger and emotional language in U.S. congressional speech and political discourse more broadly.

We finally document an educational divide. In the survey experiment, non-college-educated respondents are significantly more likely to find fairness arguments convincing. By contrast, college-educated respondents find both types of arguments equally convincing. Moreover, legislators from districts with more educated constituents are more likely to use externality arguments in both the U.S. and Norway. These patterns align with our findings on content differences, as externality arguments use more logical appeals and empirical evidence. Overall, fairness argu-

ments appear systematically linked to lower educational attainment across analyses.

In sum, our results document two distinct ways to argue for redistribution. In the U.S. Congress, pro-redistributive arguments almost entirely focus on economic fairness. These arguments are more likely to be emotional, particularly containing anger and compassion. In the Norwegian *Storting*, on the other hand, arguments based on inequality externalities are common. These arguments are more likely to contain logical appeals and empirical evidence. Although the argument types are similarly convincing to the average U.S. survey respondent, fairness arguments elicit substantially more outrage and are deemed more convincing by less-educated respondents.

A natural interpretation of our results is that the choice of argument type may constitute an underappreciated dimension of a society's political equilibrium. Fairness arguments have played a central role in historical tax reforms (Scheve and Stasavage, 2016), reflecting broad postwar commitments to solidarity. The emotional force underlying such periods of unity may wane over time, however, implying a degree of fragility in the resulting redistributive equilibrium. Externality arguments, by contrast, may be less likely to spark major reform, instead sustaining existing policies by emphasizing widely shared social benefits. The contrast between the United States and Norway is suggestive in this respect: although both countries achieved comparatively low levels of inequality by the mid-twentieth century, income and wealth inequality in the United States has risen markedly since, whereas inequality in Norway has remained broadly stable (e.g. Piketty et al., 2018; Saez and Zucman, 2020; Aaberge et al., 2020). While we cannot conclude whether rhetorical form contributed to this divergence, the possibility highlights how the equilibrium mix of arguments may shape long-run political outcomes.

Our broader methodological contribution is to develop a scalable approach for constructing and comparing full universes of naturally occurring arguments. The approach opens several research agendas, both within and beyond the study of redistributive politics. While it is natural to consider other redistributive argument types—comparing arguments *for* and *against* redistribution, for example—our methodology can also be used to compare arguments in other domains. There are universes of arguments on immigration, climate policy, or public health; we do not know what they are. In this sense, the paper contributes not only new empirical facts about redistribution, but also a framework for incorporating argument structure into the study of political economy. We complement recent work that quantifies the register of, or extracts narratives from, legislative speech (Gennaro and Ash, 2022; Aroyehun et al., 2024; Ash et al., 2024; Algan et al., 2025) by instead identifying and comparing the distinct *types* of arguments that make up a debate, and by mapping how these argument types differ across political settings.

The classification strategy we use generally follows the best-practice guidelines of Yang et al. (2025), and builds on an emerging literature showing the excellent performance of large language models (LLMs) in classifying open-ended text. In Heseltine and Clemm von Hohenberg (2024), for example, GPT-4 correctly classifies open-ended text in up to 95% of instances, and other recent studies report similar performance (Le Mens and Gallego, 2025; Soria, 2025). We find a similarly impressive performance: in the survey experiment, where arguments are written and screened for topical accuracy by different survey respondents, GPT-4o Mini correctly determines whether the argument is a fairness argument or inequality externality argument in 91% of cases.

Our main results are also robust to several alternative classification procedures, including the use of different prompts or LLMs. In total, we make 3,893,098 API calls.

The substantive contributions of the paper lie in what this analysis reveals about redistributive preferences. Fairness views feature prominently in this literature: citizens hold a range of preferences and beliefs about the fairness of the economic system, and such views have a large influence on redistributive support (e.g., Cappelen et al., 2007; Almås et al., 2020; Stantcheva, 2021).[1] The inequality externality channel, by contrast, has only recently emerged, following theoretical work showing that the many possible societal consequences of inequality can be modeled as a single externality problem (Støstad and Cowell, 2024). Most U.S. citizens believe that inequality is a net negative externality, affecting society in a variety of ways; such beliefs affect redistributive preferences through channels distinct from fairness concerns (Rueda and Stegmueller, 2016; Lobeck and Støstad, 2023). Academic policy discourse often emphasizes the societal consequences of inequality; as an example, the recent call of 500 signatories for an "International Panel on Inequality" relies entirely on externality arguments (World Inequality Lab, 2025).

Our contribution to these literatures is two-fold. First, we move beyond the stylized stimuli that dominate prior work—pre-selected survey items, information treatments, and designed survey experiments—instead analyzing full universes of naturally occurring arguments. This allows us to analyze the redistributive debate itself, and to document differences in how redistributive arguments function. The approach speaks to the growing literature on narratives in economic behavior (Shiller, 2017; Roth et al., 2020; Andre et al., 2023). Second, we document how redistributive debates differ across countries. While prior work has established cross-country differences in survey responses and experimental behavior (Fabre et al., 2024; Almås et al., 2025, e.g.), our approach also captures naturally occurring variation in which arguments are made—with potential implications for policy adoption, affective polarization, and more.

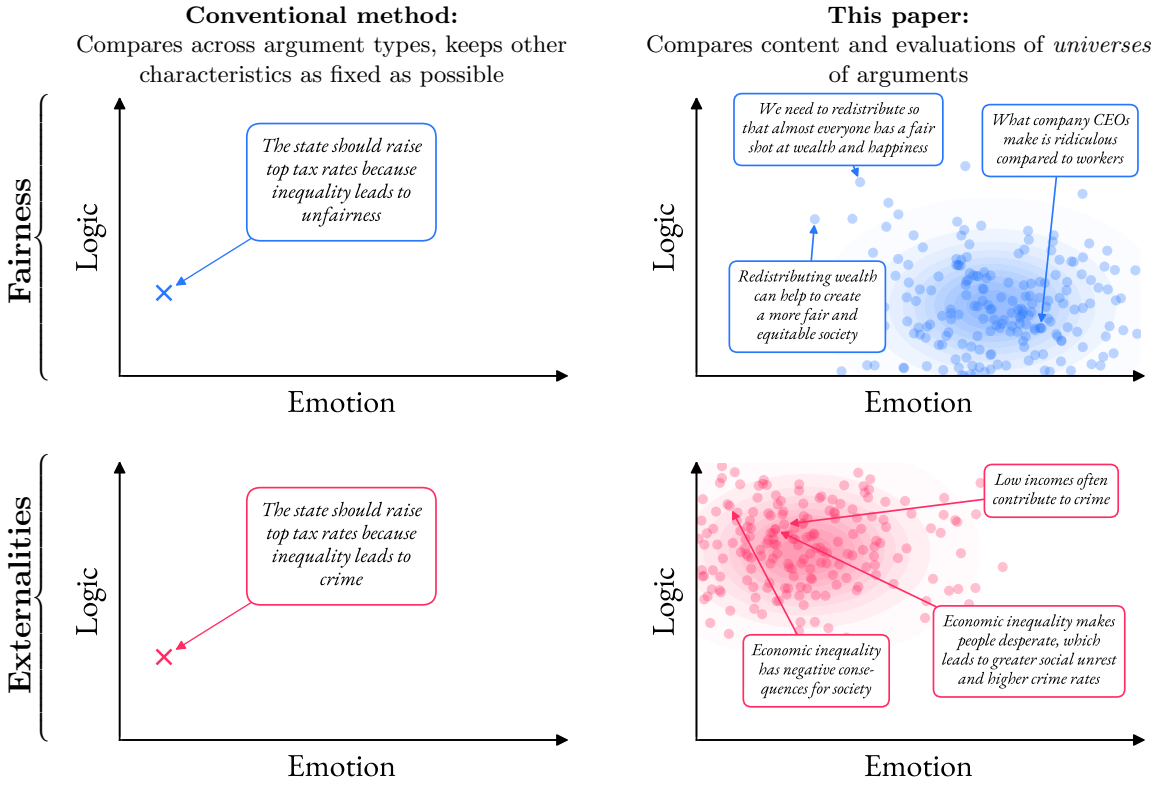
The paper is structured as follows. Section 2 introduces the methodological framework. Section 3 discusses the analysis of legislative speeches in the U.S. Congress and the Norwegian *Storting*. Section 4 discusses the methodology and results from the survey experiment. Section 5 presents a series of robustness and validation exercises. Section 6 concludes.

2 Methodological Framework: Comparing Universes of Arguments

The remainder of the paper focuses on comparing *universes* of arguments. Here we describe how this approach differs from traditional analyses. We focus on arguments, though our methodology

[1] In addition to fairness views and inequality externality beliefs, established determinants of redistributive preferences include individual income maximization (Durante et al., 2014; Meltzer and Richard, 1981); trust in government (Kuziemko et al., 2015); racial attitudes and views on immigration (Alesina et al., 2023); perceptions of the income distribution and tax system (Stantcheva, 2021; Cruces et al., 2013; Karadja et al., 2017; Hvidberg et al., 2023; Stantcheva, 2024); social and cultural identity (Luttmer, 2001; Klor and Shayo, 2010; Luttmer and Singhal, 2011); beliefs about social mobility (Alesina et al., 2018; Gärtner et al., 2019); and much more. Relatedly, Kuziemko et al. (2023) and Longuet-Marx (2025) study how the political supply of redistributive policy matches voter demand. We complement this work by documenting the cross-country supply of redistributive narratives and how citizens evaluate them.

Figure 1: Comparing Universes of Arguments



Note. The figure contrasts conventional designs that vary single argument features with our approach, which compares full universes of argument types as they naturally appear in the public debate.

could be equally used for statements or any other naturally-occurring stimuli.

Suppose the properties of an argument can be represented by its observable and unobservable properties X in a multi-dimensional space. These properties can represent anything that changes the nature of the argument, for example the broader topic, formality, length, or emotional content of the argument. We separate this content into the *type* of the argument X_{type}, defined as the main topic of analysis, and all other types of content X_{-type}. In Figure 1, where we illustrate the methodology, we show two types of arguments—fairness and inequality externality arguments—and two additional properties: the degree of emotional content and the degree of logical content of the argument.

In the left panel of Figure 1 we show the traditional approach, keeping X_{-type} constant to the extent possible, for example by only changing a small part of the stimuli across types of arguments while keeping the length, complexity, and emotional content as fixed as possible. This is clearest in a stylized example. Suppose we show respondents two arguments:

$(X_{Fair}, \mathbf{X})$ ="*The state should raise top tax rates because inequality leads to unfairness*",
$(X_{Ext}, \mathbf{X})$ ="*The state should raise top tax rates because inequality leads to crime*",

where $\mathbf{X}$ indicates the identical portion of the arguments. One could then measure how respondents evaluate these arguments differently. This approach—varying narrowly defined features of otherwise identical stimuli—underlies much of the experimental literature using information treatments, survey items, and experimental games.

We suggest, however, that X_{-type} is likely to be systematically associated with, and often

shaped by, X_{type}. Fairness arguments may be more likely to contain emotional content, for example, which may in turn increase evaluations of the argument. Inequality externality arguments may stress logical thinking, which may make them more appealing for highly educated individuals. Any such effects would be missed by the above approach. In short, keeping X_{-type} constant leads to significant concerns about external validity.

In the right panel of Figure 1 we show the approach in this paper. The core difference is that we allow all other dimensions of the argument, X_{-type}, to vary freely. In doing so we draw a random sample of arguments from the argument distribution $\mathcal{F}(X)$, where the only restriction is that $X_{type} = X_W$, where W denotes the specific type. The goal of this approach is to find a representative estimate of the distribution of arguments when $X_{type} = X_W$.

We then let individuals j evaluate the statement to decide on an individual-specific outcome Y_j, for example whether the statement is convincing to the individual or evokes an emotional reaction. This evaluation is done through an individual-specific evaluation function ϕ_j, such that $Y_j = \phi_j(X)$. Sampling many such evaluations, we estimate the distribution $\mathcal{G}$ of these Y_j for some population of individuals j, conditional on the argument being of a specific type $X_{type} = X_W$. Formally, we assess $\mathcal{G}(Y|X_{type} = X_W)$ for each X_W. This can be used to test hypotheses for $\mathcal{G}(Y|X_{type} = X_W)$ – whether one type of argument leads to more outrage than the other, for example.

We thus explore both (i) the content of each type of argument, X_{-type}, and (ii) individuals' reactions to each type of argument, $\mathcal{G}(Y|X_{type} = X_W)$. Any statistically significant effects from the resulting analysis can be interpreted as representative of this argumentative environment, provided the arguments are randomly sampled.

The main limitation of this approach is the noise introduced by randomly sampling representative arguments, which may vary widely across all possible observed and unobserved content dimensions. This heterogeneity contributes to statistical noise; the noise can be modeled and quantified empirically, however. The main strength of the approach, by contrast, is that both the classifications and evaluative responses pertain to a much broader set of arguments. Although the distribution of arguments and reactions will differ across contexts—and care must be taken to sample representative arguments from the wanted argument distribution $\mathcal{F}(X)$—the methodology is flexible and can be adapted to different contexts. One could, for example, contrast economic versus cultural frames in immigration debates, or compare individual-risk arguments and population-spillover arguments on vaccination policy.

In the rest of the paper we use the methodology to compare pro-redistributive fairness arguments and pro-redistributive inequality externality arguments. We formally define the theoretical difference between such arguments in Appendix A. Intuitively, fairness arguments focus on the distribution of resources itself (*who should have what, absent any societal consequences?*). These arguments relate to needs, rights, desert, or other normative concepts. Inequality externality arguments focus on the societal consequences that arise from that distribution (*do equal and unequal societies function differently?*). These arguments relate to the effect of the distribution on crime, trust, democratic functioning, or other societal outcomes. In short, fairness arguments are person-centered moral claims about who should have what; externality arguments justify redistribution by its causal effects on social systems and collective outcomes. We stress

that many redistributive arguments include both of these motivations, and some include neither; we return to this shortly. The classification prompts in Section 3 provide further details.

3 Legislative Speeches

The first part of the paper explores the content of legislative speeches in the U.S. Congress and the Norwegian *Storting*.

3.1 U.S. Congressional Speeches: Methodology

We use the dataset from Aroyehun et al. (2024) to analyze 750,544 excerpts from Congressional speech transcripts between 2015 and 2022. Each excerpt is cut at 150 words, unless the speech only continues for less than 50 additional words in which case this portion is appended. Using GPT-4o Mini with a Python API wrapper, we classify various properties of each speech. The prompt for each classification is long and listed in Appendix B; we summarize them here.

We first classify whether each speech excerpt contains an argument in favor of redistribution:

- **Pro-redistributive arguments:** If the speech excerpt contains an argument "*in favor of increasing economic redistribution*", giving the examples "*advocating higher taxes on the wealthy, expanding social programs, increasing [the] minimum wage, redistributing income or wealth, etc*".

This classifies 25,366 speech excerpts. For each of these pro-redistributive speech excerpts, we independently classify (i) whether it contains a fairness- or inequality externality-based argument, (ii) whether it uses a logical or emotional appeal, (iii) the specific emotional content expressed, (iv) the type of evidence presented, and (v) the type of rhetorical framing used.

In all cases we specify that the LLM should only use cues within the text itself. Every classification call is made separately, with one API call per classification per argument. We make a total of 2,881,288 API calls for the U.S. Congressional data (including the robustness checks discussed in Section 5). The LLM has no memory of previous classifications.

Fairness and inequality externality arguments: For each pro-redistributive excerpt, we classify whether it includes a fairness or inequality externality argument as follows:

- **Fairness-including excerpt:** If the speech includes an argument based on "*fairness or justice – emphasizing what people deserve, what is right or wrong, or appealing to fairness, justice, or moral principles*". The excerpt should not be classified if "*it is unrelated, or focuses only on societal consequences, efficiency, practical outcomes, direct benefits to recipients without invoking fairness, or policy mechanics.*" Prompt in Appendix B.2.
- **Inequality externality-including excerpt:** If the speech includes an argument based on "*inequality causing negative externalities at the societal level – emphasizing that reducing inequality at the top benefits society, that supporting the bottom benefits society, or that reducing general economic differences benefits society.*" The argument should "*emphasize or imply the effects on broader societal outcomes (e.g. democracy, trust, stability, crime,*

growth, cohesion, the integrity of public institutions, the concentration of power)". The excerpt should not be classified if "*it is unrelated, or focuses only on fairness, distributive justice, direct benefits to recipients without invoking societal consequences, or policy mechanics.*" Prompt in Appendix B.3.

In both cases we explicitly note that the argument must go beyond discussing the opposite classification, to avoid that pure fairness arguments are not classified as inequality externality arguments and vice versa. We also specify to "*not infer based solely on tone or rhetorical form*" and to "*focus on the substance or strongly implied themes of the argument.*"

For results on the overall prevalence of either argument type—when analyzing how often each argument type is used by different legislators, for example—we simply use every excerpt classified with these classifications. For results on content differences, however, our main aim is a clean comparison between fairness and externality arguments. As the fairness and inequality externality classifications are made separately, we can separate four types of excerpts – *Fairness only*, *Externality only*, *Neither*, and *Both*. Our main specification for content differences compares *Fairness only* to *Externality only* to avoid spillovers from excerpts containing both types of arguments. For conciseness we define these as *fairness-based* or *externality-based excerpts*. We discuss other options, including a classification that classifies every excerpt as *primarily* focused on fairness or externalities, in Section 5.2.

We also make various further classifications, listed below.

Appeals: We classify two types of appeals:

- **Emotional appeal:** If the speech includes "*appeals to emotion to persuade the reader*". We focus solely on the presence of emotional cues, independent of logical content or factual accuracy. Prompt in Appendix B.4.
- **Logical appeal:** If the speech includes an appeal that "*uses facts, statistics, definitions, comparisons, or causal reasoning to persuade the reader*". We focus solely on the presence of logical or factual reasoning, independent of the argument's truthfulness or persuasiveness. Prompt in Appendix B.5.

Emotions: We classify four types of emotions:

- **Anger:** If the speech expresses "*anger, frustration, resentment, irritation, or outrage.*" Prompt in Appendix B.6.
- **Compassion:** If the speech expresses "*compassion, empathy, care, or concern for others.*" Prompt in Appendix B.7.
- **Fear:** If the speech expresses "*fear, anxiety, worry, concern, dread, or a sense of threat or vulnerability.*" Prompt in Appendix B.8.
- **Pessimism:** If the speech expresses "*pessimism, doubt, hopelessness, or negative expectations.*" Prompt in Appendix B.9.

Types of evidence: We classify the presence of two types of evidence:

- **Empirical evidence:** If the speech includes "*empirical evidence–such as observed data, measurements, experiments, or real-world examples.*" Prompt in Appendix B.10.
- **Narrative:** If the speech "*uses a personal anecdote, individual case, vivid story, or selective example.*" Prompt in Appendix B.11.

Types of framing: We classify three types of framing:

- **Consensus-oriented:** If the speech "*tries to frame its position as universally desirable or shared by 'all of us,' 'everyone,' or 'society as a whole'.*" Prompt in Appendix B.12.
- **Villainization:** If the speech "*villainizes a person, group, or entity through its wording and tone.*" Prompt in Appendix B.13.
- **Victimization:** If the speech "*portrays a person, group, or entity as a victim of deliberate harm or oppression through its wording and tone.*" Prompt in Appendix B.14.

As the legislative excerpts are relatively long (∼150 words each), any one excerpt may contain many types of arguments and topics. There is thus substantial within-speech variation that introduces noise and may dilute empirical patterns. We cleanly separate each type of argument in the survey experiment in Section 4. In Section 5, we show that the results are largely robust to alternative prompts, to the use of open-source LLMs (ensuring replicability), and to repeated reclassification, which yields negligible changes. Any departures from robustness will be explicitly noted.

3.2 U.S. Congressional Speeches: Content of Arguments

3.2.1 Fairness- and externality-based speech excerpts

Of the 25,366 speeches which contain arguments in favor of redistribution, 52.3% are classified as including fairness arguments (N=13,250) and 12.8% are classified as including externality arguments (N=3,244). Of these, 40.6% of speech excerpts contain *only* fairness arguments (N=10,562) and 2.1% of excerpts contain *only* inequality externality arguments (N=572). Excerpts classified as *Neither* typically consist of simple assertions, administrative or procedural remarks, partisan positioning, or descriptive statements that do not articulate either a justice/desert claim or a generalized social-impact argument.

3.2.2 Speaker identity

Of the 25,366 excerpts which contain arguments in favor of redistribution, 91.4% are made by Democrats and 8.6% are made by Republicans. Conditional on making a pro-redistribution argument, Republicans use 0.16 externality arguments per fairness argument, compared to 0.25 among Democrats.

In Table 1, we explore the determinants of Congress members' propensity to use fairness and externality arguments. We regress the speaker's average use of fairness or externality arguments on local inequality, measured as the state-level top 10% income share; local income per capita,

Table 1: Associations: Inequality, Income, Education and Congress Members' Use of Arguments

	Uses Fairness arguments				Uses Externality arguments			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
State top 10% inc. share (0-1)	0.466***			0.254	0.020			-0.041
Log average state income		0.052		0.026		-0.007		-0.053
Avg. district educ. level (years)			-0.031***	-0.032***			0.016**	0.017**
Observations	1622	1622	1186	1182	1622	1622	1186	1182
Year FE and Speaker controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note. This table shows the relationship between characteristics of Congress members, and of the district and state they represent, on their propensity to use fairness-based (columns 1-4) and externality-based arguments (columns 5-8). The dependent variable is coded as 1 if the Congress member uses a fairness (resp. externality) argument in their speech (using the *All* definition, that is, including arguments that were classified as both a fairness- and externality-based argument), 0 otherwise. Table G1 shows results for alternative definitions of the dependent variable. Data covers the 2015-2022 periods, and is collapsed at the speaker-year level. State top 10% income share uses post-tax income and is measured between 0 and 1. Log average state income uses fiscal income. Both variables are taken from the World Inequality Database, and are missing after 2018 – we extend the 2018 value to subsequent years. District education level is measured in years of education, data is from the 2022 American Community Survey. Congress members' education level is measured in 4 categories (less than Bachelor's, Bachelor's, Master's, Professional degree or PhD), and data is obtained from the Biographical Directory of the United States Congress. The sample is smaller when including district education level because we drop Senators, who are elected at the state level, not the district level. The regressions include year fixed effects, a dummy for being a Republican, and a dummy for male gender. *Significance levels*: *10%, **5%, ***1%.

which is also measured at the state level; local education level, measured at the district level, and finally, the Congress member's own education level. We also control for year fixed effects, the speaker's political party, education level, and gender.

The use of fairness arguments is significantly associated with a higher state top 10% income share and a lower average district education level, both at the 1% level, although the association to state inequality disappears in a joint regression. The use of externality arguments is only significantly associated with district education level, at the 5% level, which remains significant in the joint regression. The magnitude of these educational associations is reasonably large; Congress members elected in a district with one additional year of education use 3.1 p.p. less fairness arguments and 1.6 p.p. more externality arguments on average. Moving from the least-educated to the most-educated district in our sample would decrease the share of fairness arguments by 16 p.p. and increase the share of externality arguments by 9 p.p.. Table G1 in Appendix shows that the education result is robust to different definitions of fairness and externality arguments.

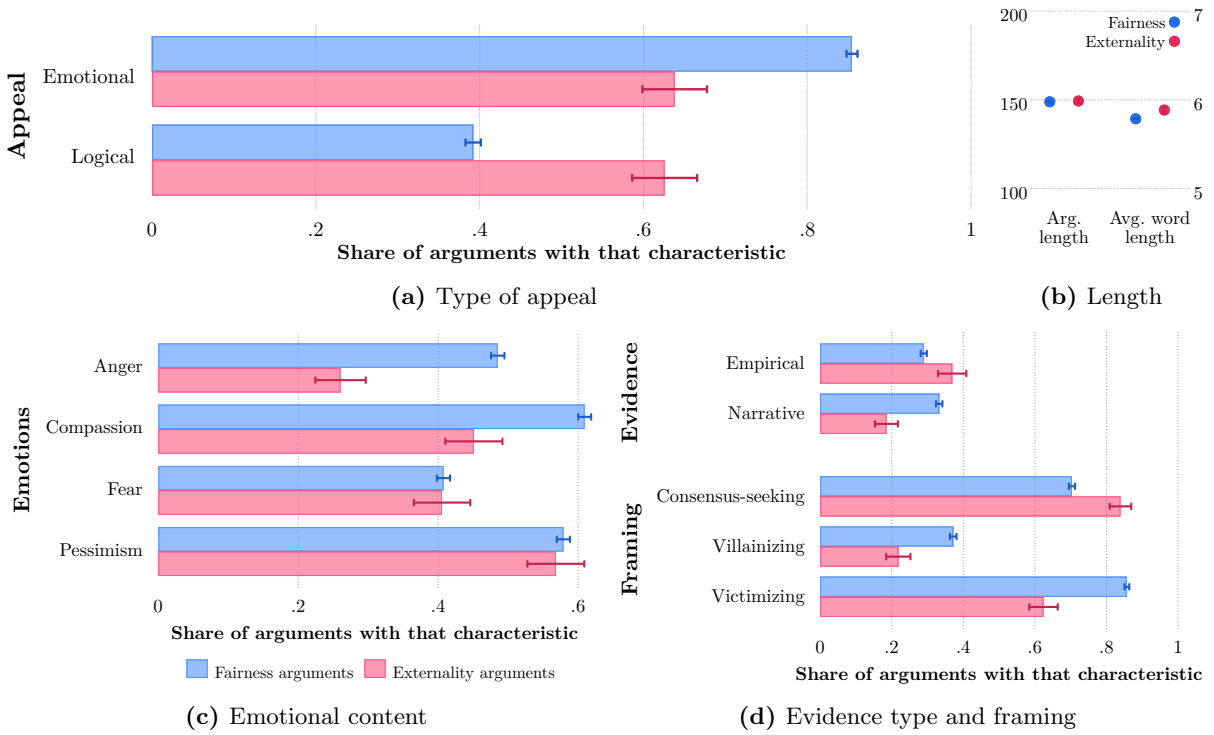
3.2.3 Other content differences

Figure F2 shows the correlation matrix across all content classifications. The associations are intuitive; the strongest association is between anger and villainization (0.71), for example, and logical appeals are positively associated with empirical evidence (0.44) but not narrative evidence (-0.24).

We now examine the content differences across the two universes of arguments, summarized in Figure 2.

Length We show the length of the average speech excerpt, and the average word length within speech excerpts, in Figure 2b. Speech excerpt length is similar (148.89 vs. 149.34, p=0.522), as excerpts are cut at 150 words by construction. The average word length is slightly lower in fairness-based excerpts than externality-based excerpts (4.79 vs. 4.89, p<0.001).

Figure 2: U.S. Congressional Speeches: Content of Excerpts Across Excerpt Type



(a) Type of appeal **(b)** Length

(c) Emotional content **(d)** Evidence type and framing

Note. This figure shows content differences between fairness and externality arguments in the U.S. Congress data. Results are based on speeches containing only fairness (N=10,562) or only externality arguments (N=572). We discuss alternative specifications in Appendix C.1.

Form of appeals An average of 72% of excerpts contain appeals to emotions, while 48% contain appeals to logic.

In Figure 2a we show how these shares vary across fairness- and inequality-externality based excerpts. Fairness-based excerpts are substantially more likely to contain emotional appeals than externality-based excerpts (85% vs. 64%, $p<0.001$). The opposite pattern holds for appeals to logic, which are more common in externality-based excerpts (39% vs. 63%, $p<0.001$). These differences are extremely robust across specifications (Section 5).

Emotions An average of 38% of excerpts express anger, 53% express compassion, 37% express fear, and 53% express pessimism.

In Figure 2c we show how these shares vary across fairness- and externality-based speeches. Fairness-based excerpts are almost twice as likely to express anger relative to externality excerpts (49% vs. 26%, $p<0.001$). Fairness-based excerpts are also more likely to express compassion (61% vs. 45%, $p<0.001$). We find no significant differences for fear (41% vs. 41%, $p=0.501$) and pessimism (58% vs. 57%, $p=0.917$). While fairness-based excerpts contain more anger and compassion across specifications, externality-based excerpts often contain higher fear and pessimism in alternate specifications (Section 5).

Types of evidence An average of 30% of excerpts include empirical evidence, while 26% include narrative-based evidence.

In Figure 2d we show how these shares vary across fairness- and inequality-externality based

excerpts. Externality-based excerpts are more likely to contain empirical evidence (37% vs. 29%, $p<0.001$), while fairness-based excerpts are more likely to contain narrative-based evidence (19% vs. 33%, $p<0.001$). These differences are strongly robust across specifications (Section 5).

Types of framing An average of 68% of speeches present their case as seeking broad consensus agreement. 70% victimizes a person, group, or entity, and 30% villainizes a person, group, or entity.

In Figure 2d we show how these shares vary across fairness- and externality-based excerpts. Externality-based excerpts are more likely to search for consensus (84% vs 70%, $p<0.001$), while fairness-based excerpts are significantly more likely to victimize (62% vs. 86%, $p<0.001$) and villainize (22% vs. 37%, $p<0.001$). These differences are generally robust across specifications, although differences are occasionally smaller (Section 5).

Taken together, the Congressional evidence yields two main results—later shown to generalize across settings:

Result #1

Fairness arguments are more likely than externality arguments to contain emotional appeals, anger, compassion, narrative-based evidence, victimization, and villainization. Externality arguments are more likely to contain logical appeals and empirical evidence while seeking consensus.

Structure within argument types We further subdivide fairness and externality arguments into finer subtypes. We classify fairness arguments into three further types—needs-based, rights-based, and compensatory (merit-based). 61% are classified as needs-based, 25% are classified as rights-based, and 13% are classified as compensatory. These types generally contain similar content. We classify externality arguments into two types—bottom-income based and top-income based. 81% are classified as bottom income-based and 18% are classified as top income-based. Top-income based externality arguments are thus relatively rare, and are also closer to fairness arguments in style. We discuss these subtypes further in Appendix C.3.

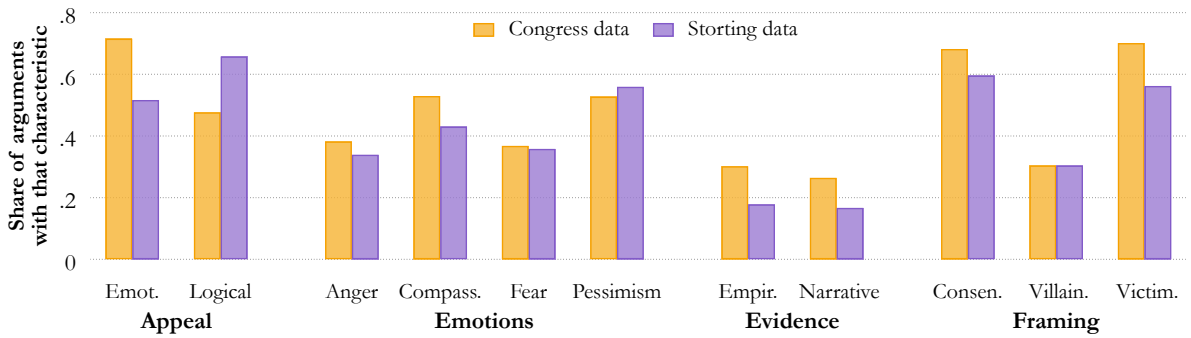
We now move to the second setting, the Norwegian *Storting*.

3.3 Norwegian *Storting* speeches: Methodology

We use the database of Fiva et al. (2025), modified for comparability with the Aroyehun et al. (2024) database in the U.S. Congressional setting; we restrict our attention to the data between 2015 and 2022 and manually divide all speeches into 150-word excerpts. The remaining methodology is nearly identical to the methodology used for the U.S. Congressional speeches, including the prompt language.[2]

[2]The only minor adjustment is that we note the data source in both cases ("a Congressional speech" or "a Norwegian parliamentary speech") since mismatches between the prompt language and the argument text otherwise introduce noise. Specifying only the language (e.g., "a Norwegian speech") or translating speeches before prompting yields nearly identical results.

Figure 3: Average Argument Characteristic in the U.S. Congress and Norwegian *Storting*



Note. This figure displays, for each content characteristic, the share of redistributive excerpts classified as containing this characteristic in the U.S. Congress and Norwegian *Storting* samples. All excerpts classified as containing a redistributive argument are included. Figure C2 compares the content of pro-redistributive and non-redistributive speech excerpts. Appendix B shows the full prompt texts.

The dataset contains 259,014 excerpts, compared to 750,544 in the U.S. Congress dataset. A total of 17,109 of these excerpts contain arguments for more redistribution, compared to 25,366 in the U.S. case. As such, a higher share of all excerpts are defined as arguing for redistribution in Norway than in the United States. We make a total of 1,011,810 API calls for the Norwegian *Storting* data.

3.4 Norwegian *Storting* Speeches: Content of Arguments

Figure 3 shows net content differences for pro-redistributive speech excerpts in the Norwegian *Storting* and the U.S. Congress. Redistributive arguments in the U.S. Congress contain more emotional appeals and fewer logical appeals, higher levels of anger and compassion, more narrative-based and empirical evidence, more villainizing, and, somewhat unexpectedly, more consensus-seeking. These differences are generally not unique to the redistributive debate; we document similar cross-country content differences for non-redistributive excerpts (Figure C2). Argument length is similar by construction, while average word length is slightly higher in the *Storting*.

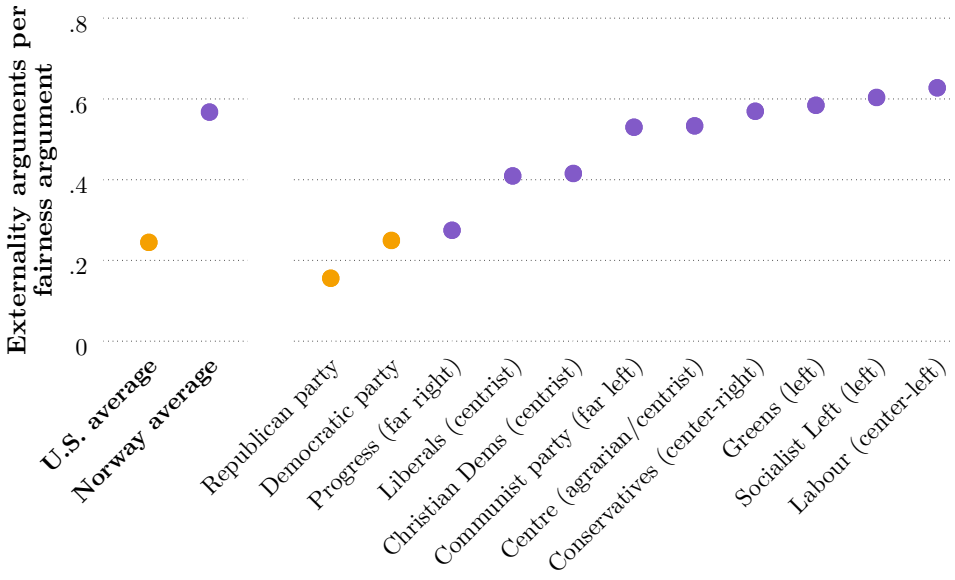
3.4.1 Fairness- and externality-based speech excerpts

Of the 17,109 excerpts which contain arguments in favor of redistribution, 66.0% are classified as including fairness arguments (N=11,285) and 37.4% are classified as including externality arguments (N=6,403). Of these excerpts, 32.6% (N= 5,574) and 4.0% (N=692) include *only* fairness or externality arguments, respectively.

As such, the shares of redistributive excerpts containing externality arguments in the *Storting* and the U.S. Congress are 37.4% and 12.8%, respectively. The shares of excerpts containing fairness arguments are 66.0% and 52.3%. We compare these ratios by using the relative share of externality and fairness arguments; under this specification, externality arguments are 2.3 times more common in Norway. As there are several alternate ways to define this ratio, we discuss other specifications in Appendix C.1—across all choices, externality arguments are much more frequently used in Norway, being between 1.8-4.8 times as common.

This establishes the second main result:

Figure 4: Fairness Dominates the U.S. Debate, Externality Arguments are Common in Norway



Note. This figure reports the ratio of externality arguments per fairness argument across all political parties in the U.S. Congress and Norwegian *Storting*. The numerator includes all excerpts classified as containing only an externality argument, or both a fairness and an externality argument. The denominator includes all excerpts classified as containing only a fairness argument, or both a fairness and an externality argument. Appendix B.2 (resp. B.3) shows the prompt classifying the excerpts as containing a fairness (resp. an externality) argument.

Result #2

Externality arguments are between two to five times more frequent in the Norwegian parliament than in the U.S. Congress.

Argument types across political parties In both countries, our sample is predominantly made up by parties on the political left; in the U.S., 91.4% of excerpts are from Democratic politicians, whereas in Norway, 80.3% are from left-leaning parties. The use of externality arguments varies across parties, as previously documented across the Democratic and Republican parties in the U.S. Congress (who use 0.25 and 0.16 externality arguments per fairness argument, respectively). In the Norwegian political parties, the corresponding ratios range from 0.27 to 0.63. Notably, the externality-to-fairness ratio is higher for every Norwegian party than for either major U.S. party (Figure 4). Within Norway, externality arguments are generally more common on the political left, though the pattern is not monotonic: both major governing parties—the center-left Labour Party and the center-right Conservatives—invoke externality arguments more frequently than the left-right axis implies.[3] Labour Party members display the highest externality-to-fairness ratio of any party, at 0.63; Labour is also the political party of Jens Stoltenberg, who delivered the opening quote of the paper.

Table 2: Associations: Inequality, Income, Education and *Storting* Members' Use of Arguments

	Uses Fairness arguments				**Uses Externality arguments**			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
County-level Gini (0-1)	0.377			0.854	0.603*			-0.487
Log average county income		0.053		-0.122		0.177		-0.017
Avg. county educ. level (years)			0.018	-0.010			0.055**	0.087*
Observations	906	906	906	906	906	906	906	906
Year FE and Speaker controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note. This table shows the effect of *Storting* members' and their county's characteristics on their propensity to use fairness-based (columns 1-4) and externality-based arguments (columns 5-8). *County* corresponds to the Norwegian *fylke*. The prompt used to classify arguments as containing an externality argument or not is shown in Appendix B.3. The dependent variable corresponds to the *All* definition, that is, including arguments that were classified as both a fairness- and externality-based argument). Table G2 shows results for alternative definitions of the dependent variable. Data covers the 2015-2022 periods, and is collapsed at the speaker-year level. County-level Gini uses household-equivalent income and is measured between 0 and 1. Log average county income uses gross income. District education level is measured in years of education. All three variables are obtained from Statistics Norway. *Storting* members' education level is measured in 4 categories (less than Bachelor's, Bachelor's, Master's, PhD), and data is obtained from the *Storting*'s website, stortinget.no. The regressions include year fixed effects, a dummy for each political party, and a dummy for male gender. *Significance levels*: *10%, **5%, ***1%.

3.4.2 Speaker identity

In Table 2 we replicate Table 1 for the *Storting* data, exploring the determinants of *Storting* members' propensity to use externality arguments. The smallest geographic unit available is the county (*fylke*), and we use Norway's pre-2020 classification of 18 counties to match the structure of the available data. Local inequality, local income per capita, and local education levels are measured at the county level. We also control for year fixed effects, the speaker's political party, education level, and gender.

No variable is significantly correlated with the use of fairness arguments. The use of externality arguments is positively associated with county-level inequality at the 10% level, and positively associated with the average county education level at the 5% level. Only the education coefficient is significant (at the 10% significance level) in the joint regression. A representative of a county where average education is one year higher uses 5.5 p.p. more externality arguments on average; the magnitude of this association is larger than the corresponding association in the U.S. data, while the significance level is lower due to a smaller number of counties. Table G2 shows that this result is largely robust across different definitions of externality arguments.

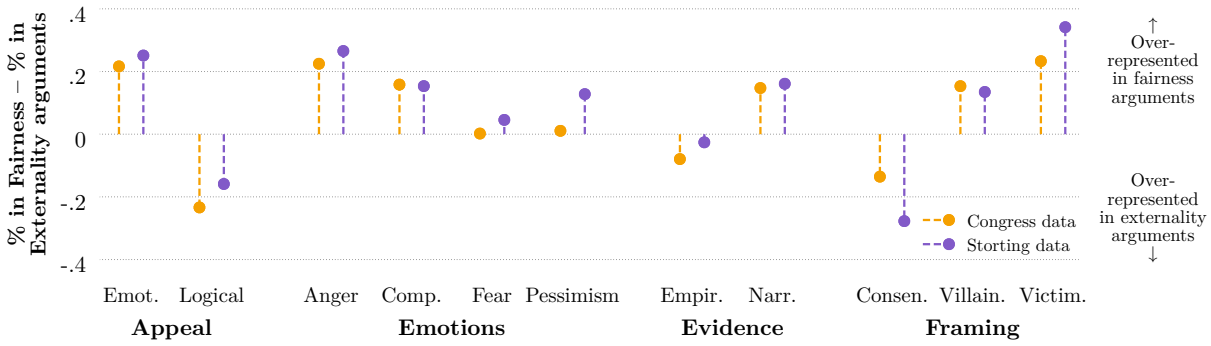
3.4.3 Cross-country differences in externality- and fairness-based excerpts

Figure 5 compares content differences between fairness and externality arguments in the Norwegian *Storting* to those observed in the U.S. Congressional data. The direction of the comparison (e.g., fairness arguments containing more anger) is consistent across all 11 classifications, and we generally find strikingly small differences. This close correspondence provides reassurance that our classification approach captures different types of arguments in a consistent manner across countries.

We show the effect of constructing counter-factuals—reweighting each country's redistributive debates using the other's mix of fairness and externality arguments—in Appendix C.2.

[3]The political spectrum, from right to left, in Norway: The Progress Party (0.27), the Conservatives (0.57), the Liberals (0.41), the Christian Democrats (0.42), the Centre (0.53), Labour (0.63), the Greens (0.58), the Socialist Left (0.60), the Communist Party (0.53).

Figure 5: Content Gaps Between Fairness and Externality Excerpts are Consistent Across Countries



Note. This figure reports content differences between fairness and externality arguments across the two legislative settings (U.S. Congress and Norwegian *Storting*). Each dot is obtained by subtracting the share of externality arguments containing that characteristic to the share of fairness arguments containing that characteristic. Figure F3 shows the level share of each characteristic across fairness and externality arguments in the *Storting* data (replicating Figure 2 for the *Storting* data). Figure F4 adds 95% confidence intervals to the content differences. Appendix B shows the full prompt texts.

While such exercises tend to make the debates look 10-30% more similar, we suggest caution in interpreting these results. The estimates are sensitive to methodological choices, and focusing on a single similarity measure overlooks that most cross-country content differences also arise in non-redistributive speech excerpts.

Overall, data from the legislative excerpts have shown that fairness arguments are more likely than externality arguments to be emotionally charged and divisive, whereas externality arguments are more likely to appeal to logic and seek consensus. We have also documented that externality arguments are two to five times more common in the Norwegian *Storting*, and that externality arguments are more commonly used by legislators from more educated regions.

We now move to the second part of the paper; the U.S.-based survey experiments.

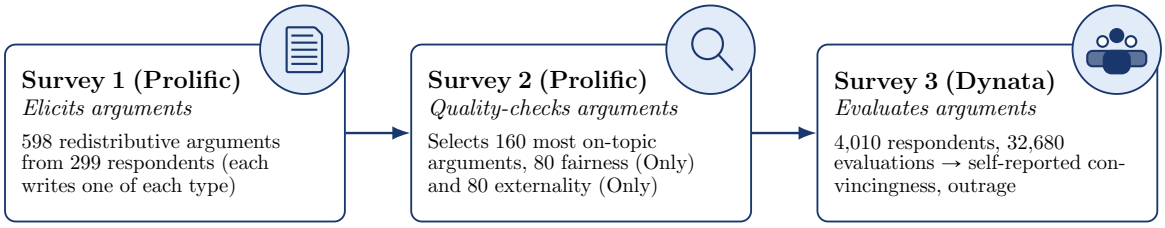
4 Survey Experiments

To complement the analysis of legislative speeches, we use a set of surveys that provide a controlled environment for content analysis. This allows us to study argument content without selection into political speech, to validate the LLM-based classifications, and to measure how different argument types are evaluated by respondents in terms of convincingness and outrage.

4.1 Survey Experiment: Methodology

The survey methodology follows a three-step process, which we illustrate in Figure 6. The three surveys are designed to respectively *elicit*, *quality-check*, and *evaluate* arguments. By eliciting arguments and quality-checking them with survey respondents, we first construct representative estimates of two "universes" of arguments—one centered on fairness and one on inequality externalities—using only human input. We then (i) analyze content differences across argument types in this controlled setting, using the same LLM methodology as for the legislative speeches, and (ii) measure survey respondent evaluations of each argument type with the third survey. We conducted all three surveys in the United States between October 24$^{\text{th}}$ and December 23$^{\text{rd}}$

Figure 6: Overview of the Three-Survey Design



2022.

Survey 1: Eliciting arguments We first *elicit* arguments of type X_{Fair} and X_{Ext} from 298 respondents in Survey 1 on *Prolific* between November 16$^{\text{th}}$ and 27$^{\text{th}}$ 2022. Each respondent was prompted to write two arguments about redistribution sequentially in random order; one about fairness and one about inequality's societal consequences. In each case, respondents themselves chose whether to write an argument in favor or against redistribution. 74% of arguments were in favor of redistribution (statistically indistinguishable across fairness or inequality externality arguments). While respondents could alternate between pro- and anti-redistribution positions across the two framings, 97% of the arguments in the final evaluated universes of pro-redistributive arguments were written by respondents who decided to write two pro-redistribution arguments.

We show the full question text in Figure 7. Respondents were told to write an argument that would convince a friend, asked to be brief (three sentences maximum), and informed that convincing arguments would be rewarded. Instructions emphasized staying within the assigned framing and encouraged the use of the respondent's own words and reasoning. We note that the survey was conducted before the proliferation of LLMs (ChatGPT first launched November 30$^{\text{th}}$ 2022).[4] Prior to this question, respondents were also informed that their arguments would be shown to other survey respondents, that their survey payout would be doubled if their argument were found convincing by a majority of other respondents, and that their arguments were completely anonymous. We further discuss the methodology in Appendix D.1.

Survey 2: Quality checking arguments We *quality check* the arguments using 210 respondents in Survey 2 on *Prolific* to remove any off-topic or nonsensical arguments from the sample. We reduce the sample to 160 pro-redistributive arguments, 80 of each type, on pre-specified criteria. While this task could be carried out by research assistants, crowdsourcing it to survey respondents mitigates the structural risk that a limited set of evaluators—through repeated exposure to arguments and/or proximity to the research project—gradually infers researcher expectations and classifies accordingly.

This step creates the final two universes of redistributive arguments, which correspond to $\hat{\mathcal{F}}(Y|X_{type})$ for each X_{Type} (using the framework of Section 2). We show the criteria and discuss the methodology in Appendix D.2.

[4]The original goal of the project was to create large-scale evaluations of two unbiased universes of arguments. The rise of LLMs allowed us to substantially expand the project by classifying the *content* of these arguments.

Figure 7: Elicitation of Arguments: Question Text

Imagine you want to convince a friend to support {**more** / **less**} economic redistribution with an argument about how [this would **be fair** / economic inequality has {**negative** / **positive**} **consequences** for society]. Please write a **brief** (3 sentences maximum) argument below.

[*Fairness text:*] 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.

[*Inequality externality text:*] 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 / less}?

Note. This figure shows he question text for the elicication of arguments in Survey 1. Differences in text across the elicitation of fairness and inequality externality arguments are denoted by text and [brackets]. Differences in text across pro- and anti-redistributive argument elicitation are denoted by {brackets}. Figure D5 shows the information respondents were shown before the argument elicitation.

Survey 3: Evaluating arguments Finally we collect *evaluations* of the Survey 1 arguments with a distinct representative sample of 4010 respondents, collected with the professional survey company *Dynata*. Our main outcomes are self-reported measures of whether the respondent was convinced by the argument (which we define as "convincingness") and whether the respondent believes a conversation with the author of the argument could make them angry or agitated because they agree with the sentiment ("outrage"). These evaluations construct the distribution of outcomes $\mathcal{G}(Y|X_{type})$ for outrage and convincingness for each X_{Type}. We discuss the methodology further in Appendix D.3.

In sum we collect 598 total arguments for and against redistribution, of which 160 arguments for redistribution remain in our final sample (80 of each type). We collect a total of 32,680 evaluations for these arguments from 4,010 respondents, at an average of 202 evaluations per argument.

4.2 Survey Experiment: Robustness of LLM classification

Recent work indicates that modern LLMs perform very well on open-ended text classification, approaching or outperforming human performance levels (Heseltine and Clemm von Hohenberg, 2024; Le Mens and Gallego, 2025; Soria, 2025). Still, it remains important to validate their performance. To do so we test whether the LLM classification method we use for the Congressional speech excerpts is able to accurately classify the fairness and inequality externality arguments elicited and quality-checked in Surveys 1 and 2. We prompt GPT-4o Mini using the exact phrasing described in Section 3.1, and run two independent classifications for each of the final 160 arguments. Each argument is thus assigned to one of four categories: fairness, inequality externality, both, or neither.

This classification returns impressive results. The LLM classification successfully classifies 93% of fairness arguments as fairness arguments, and 89% of externality arguments as externality arguments. Overall, 91% are correctly classified as the corresponding argument type.

Since our main classification for the content analysis uses both classifications to define excerpts as only fairness- or externality-based, we note that 89% of the fairness arguments are correctly classified as fairness and *not* externality arguments, and that 83% of the 80 externality arguments are correctly classified as externality and *not* fairness arguments. Overall, 86% are correctly classified twice. There are essentially no false positives; only 1% of arguments are classified as *Fair only* or *Ext only* without being a fairness or externality argument, respectively.

These results provide strong support for the validity of the LLM classifications. Some measurement error is nonetheless inevitable, and could be more prevalent in the longer and more varied legislative excerpts. Content classifications beyond the fairness–externality distinction are likely to carry lower error rates, as they involve simpler categorical judgments on which LLM performance tends to improve (Heseltine and Clemm von Hohenberg, 2024). To the extent that such errors arise, they are likely to attenuate the measured differences in content between fairness and externality arguments, implying that our difference estimates should be viewed as conservative.

We now move to the experimental analysis, where we define fairness and externality arguments based on the Survey 1 prompts (such that fairness arguments are always written as fairness arguments and vice versa).

4.3 Survey Experiment: Content of Arguments

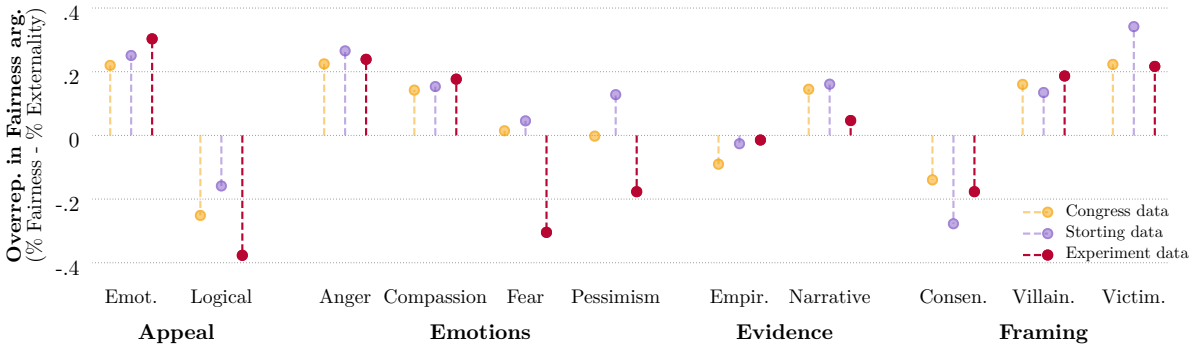
We classify the content of the experimentally elicited arguments (emotional appeal, anger, and so on) with the same GPT-4o Mini prompts as we used for the Congressional sample.

Content classification: Comparison to legislative excerpts Nearly all content classifications appear more frequently in the legislative speeches than in the experimental data (Figure F9). This is likely because the legislative excerpts are about three times as long as the elicited arguments (average length of ∼150 vs. 49 words). Indeed, the only classification that occurs similarly often in the experimental data is consensus-seeking, which is naturally less dependent on text length than other types of content.

Speaker identity The share of redistributive arguments written by Democrat (88.8%) and Republican (11.3%) survey respondents are similar to that in the Congressional data, as we intentionally over-sampled Democrats in the sampling methodology (see Appendix D). There is no significant difference in political affiliation across argument types.

Cross-sample differences in fairness- and externality-based arguments We compare the content differences among fairness and externality arguments to those observed in the Congressional speech excerpts in Figure 8. The net difference across fairness and inequality externality-based arguments or excerpts are remarkably similar in the experimental and legislative samples for most classifications. As in the legislative samples, fairness arguments are markedly more likely to contain emotional appeals, anger, compassion, villainizing, and victimizing. Externality arguments are more likely to contain logical appeals and consensus-seeking.

Figure 8: Content Gaps Between Argument Types Consistent Across Legislatures, Experiment



Note. This figure reports content differences between fairness and externality arguments across the three settings (U.S. Congress, Norwegian *Storting*, and the survey experiment). Each dot is obtained by subtracting the share of externality arguments containing that characteristic to the share of fairness arguments containing that characteristic. Figure G16 shows the level share of each characteristic across fairness and externality arguments in the experimental data (replicating Figure 2 for the experimental data). Appendix B shows the full prompt texts.

We find small differences for both empirical and narrative evidence, as almost no experimentally elicited arguments use either.

We identify two classifications with larger differences. Both fear and pessimism are more prevalent in externality arguments in the experimental sample, contrasting with the slight tendency toward fairness arguments observed in the legislative data. Fairness arguments in the experimental sample almost never contain fear, whereas externality arguments frequently do (Figure G16). In the legislative evidence, by contrast, there were no significant differences across types. We observe a similar but less marked pattern for pessimism. Notably, this externality slant for fear and pessimism is also common in our robustness checks (Section 5); as such, we suggest that fear and pessimism may be more closely associated with externality-based arguments than the primary legislative analysis would indicate.

Overall, however, content differences across fairness and inequality externality arguments are remarkably consistent across the legislative and experimental samples. Given the known differences between the samples on content, self-selection into speeches, and more, we consider this strong evidence that the differences we find across argument types are strongly robust features of these types of arguments that are likely to be present in different contexts.

4.4 Survey Experiment: Reactions to Arguments

We now discuss the results from Survey 3, where we collected 32,680 evaluations of the redistributive arguments from Survey 1 and 2, using a sample of 4,010 U.S. survey respondents.

4.4.1 Descriptive statistics

Overall, 54% of evaluations find a given pro-redistributive argument convincing, while 34% report that a conversation with the author could provoke anger because the respondent agrees with the argument ("outrage"). Outrage is significantly correlated with convincingness (Supplementary Figure F10), likely because (i) outrage is defined as anger arising from *agreement*, and (ii) both self-reported measures were elicited on the same survey screen. This should be kept in mind when interpreting the results. Right-leaning, older, and higher-wealth respondents are less

likely to be both convinced and outraged by arguments (Figure F11). We show the three most convincing and least convincing arguments in Supplementary Table G3, and the three most and least outrage-inducing arguments in Supplementary Table G4.

4.4.2 Empirical specification

Our coefficient of interest is β in the following regression:

$$y_{i,j} = \gamma_i + \beta\alpha_j + \varepsilon_{i,j}. \tag{1}$$

where $y_{i,j}$ is the outcome of interest (e.g. convincingness) for individual i and argument j, γ_i is a person fixed effect, α_j is the argument characteristic of interest (e.g. a dummy equal to one if the argument is an externality argument), and standard errors $\varepsilon_{i,j}$ are clustered at the argument level, capturing the fact that argument characteristic α_j only varies at the argument level. In the joint regression, α_j is a $k \times 1$ vector of argument characteristics, and β is a $1 \times k$ vector of coefficients.

4.4.3 Convincingness

We show how different argument characteristics relate to convincingness in Figure 9. Convincingness is a dummy equal to 1 if the participant stated that they were "*Very convinced*" or "*Somewhat convinced*" by the argument, 0 otherwise (see Appendix H.3.4 for the full survey question). There is no statistically significant difference between fairness and inequality externality arguments, although fairness arguments are on average 1.6 p.p. more likely to be evaluated as convincing—an increase of .027 standard deviations of the argument distribution. The largest effect size we can rule out under a 95% confidence interval is 3.6 p.p. (.067 standard deviations). By contrast, many short information treatments find effect sizes of 0.1 standard deviations or more (Stantcheva, 2021; Lobeck and Støstad, 2023).

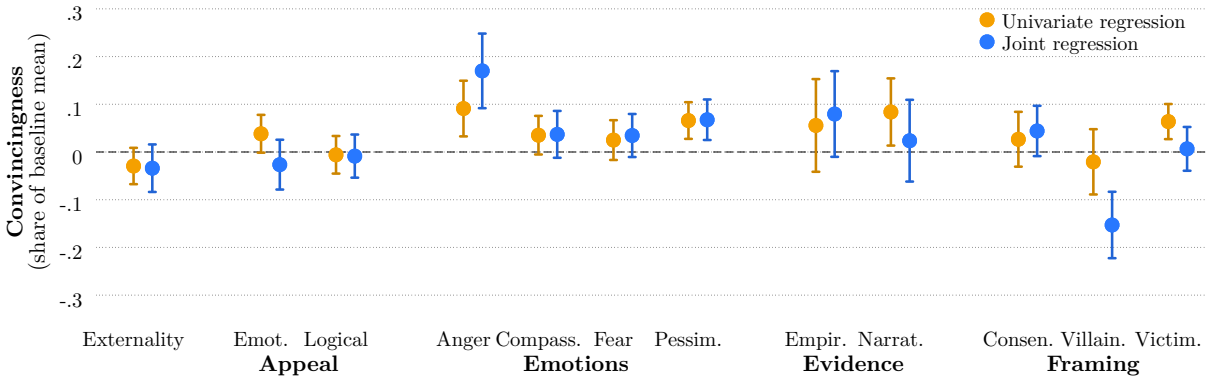
This establishes the third main result:

Result #3

In a controlled survey experiment, we cannot reject the null that fairness and externality arguments are equally convincing.

Across other characteristics, we find significant associations between convincingness and the content classes of anger, pessimism, narrative evidence (in the bivariate regression only), and villainization (in the joint regression only). The strongest associations are with anger, as anger content is associated with 4.9 and 9.1 p.p. higher convincingness in the bivariate and joint regressions. Villainization has a strong negative effect in the joint specification (-8.2), which almost cancels out the anger effect for arguments containing both. These patterns are noteworthy in their own right, suggesting that while expressions of anger can make political arguments more persuasive, targeting or blaming others – which is strongly associated with anger – may make them less so.

Figure 9: Association Between Convincingness and Argument Characteristics



Note. This figure shows the result of regressing convincingness on argument characteristics. The Convincingness variable is first divided by its mean (.537), so coefficients should be interpreted as a share of the mean. The yellow dots correspond to a regression of convincingness on the indicated variable, as well as person fixed effects, a dummy for above-median argument length, and average word length in characters. The blue dots correspond to the joint regression of convincingness on all characteristics, with the same controls. Figure F12 shows the same regression, restricting the sample to participants who passed the attention check. Standard errors are clustered at the argument level, 95% CIs.

4.4.4 Outrage

We show how different argument characteristics relate to outrage in Figure 10. Outrage is a dummy equal to 1 if the participant answered "*Yes, because I agree*" or "*Partly, because I agree*" to the question "*Do you think a discussion about this argument could provoke an emotional reaction like anger or agitation in you*?" (see Appendix H.3.6 for the full survey question). Respondents are 1.9 p.p. (0.35 standard deviations, p=0.009) more likely to report outrage in response to a fairness argument than to an externality argument. This difference is almost entirely driven by underlying content differences; the association is not significant in the joint regression. An argument containing anger is strongly associated with outrage in both separate and joint regressions. Emotional appeals, victimization, and narrative evidence are also significantly positively associated with outrage in the bivariate specification, while villainization is negatively associated with outrage in the joint regression.[5]

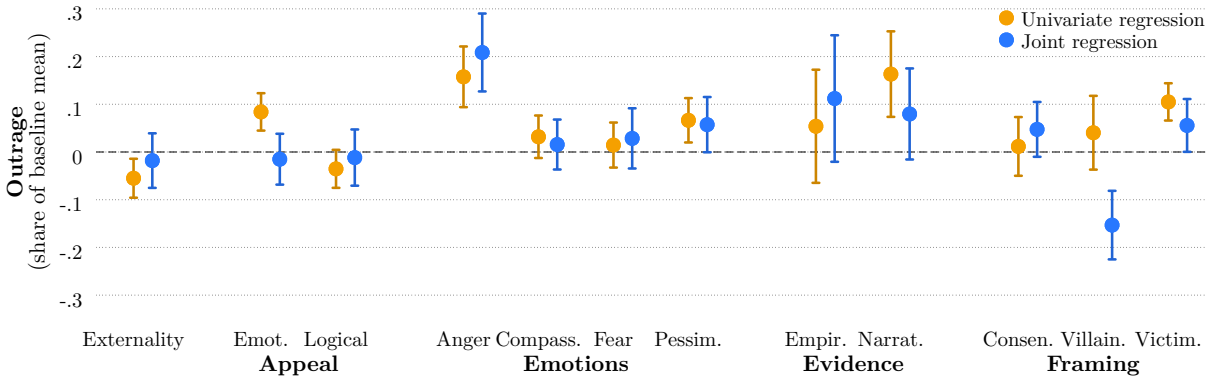
4.4.5 Heterogeneity across groups

Are fairness or inequality externality arguments more convincing for certain groups? Figure 11a reports a simple heterogeneity analysis. Fairness arguments are significantly more convincing than externality arguments for respondents without a college degree, and for respondents with annual incomes below $50,000. We cannot disentangle these two effects due to power issues. Republican-leaning respondents are also slightly more likely to find fairness arguments convincing. The findings on income and political affiliation are consistent with Lobeck and Støstad (2023), who found that externality arguments are relatively more important for the redistributive preferences of Democrat-leaning and higher-income people.

The education pattern is both the largest and most robust across settings, however. Here, the experimental evidence aligns with the higher use of externality arguments among legislators from more educated districts, and with the greater reliance of such arguments on logical appeals and

[5]This likely reflects that villainization reduces agreement with the argument.

Figure 10: Association Between Outrage and Argument Characteristics



Note. This figure shows the result of regressing outrage on argument characteristics. The outrage variable is first divided by its mean (.341), so coefficients should be interpreted as a share of the mean. The yellow dots correspond to a regression of outrage on the indicated variable, as well as person fixed effects, a dummy for above-median argument length, and average word length in characters. The blue dots correspond to the joint regression of outrage on all characteristics, with the same controls. Figure F13 shows the same regression, restricting the sample to participants who passed the attention check. Standard errors are clustered at the argument level, 95% CIs.

empirical evidence. Taken together, the results point to an educational divide, where fairness arguments are linked to a lower educational attainment.

This establishes the fourth and final main result:

Result #4

We document an educational divide, where fairness arguments are consistently linked to lower educational attainment across analyses.

We show the same heterogeneity for outrage in Figure 11b, where differences across groups are small.

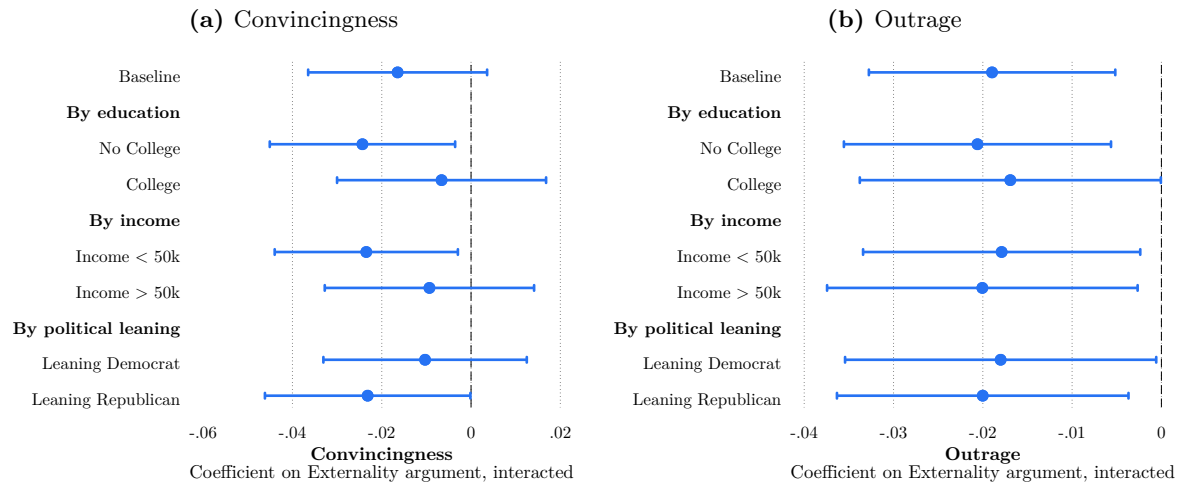
4.5 Effect of arguments on redistributive preferences

To validate the self-reported measures, we test whether the mix of arguments a respondent sees affects their agreement with the statement "*the government should take measures to reduce differences in income levels*". Each respondent evaluates 10 randomly drawn arguments, where 85% of arguments are in favor and 15% of arguments are against redistribution.[6] Because redistributive preferences are also measured with a different question before the argument evaluation, the design allows us to determine whether exposure to a greater number of pro-redistributive arguments causally shifts demand for redistribution.[7] Because the number of pro- and anti-redistributive arguments is randomly assigned, any difference in post-evaluation views can be interpreted causally. We complement this with an analysis of whether the average leave-one-out convincingness of the arguments a respondent sees (excluding their own ratings) predicts

[6]The anti-redistributive arguments are excluded from all other analysis.

[7]Defined as answers to the question "*How much redistribution of income do you prefer across citizens in the U.S.?*", with options from "1=No redistribution" to "7=Full redistribution." We code this variable as a dummy equal to 1 if the participant has answered 4 or more on a 1-7 scale, 0 otherwise. The full questions are shown in Appendix H.3.1 and H.3.7.

Figure 11: Heterogeneity of Fairness / Externality Difference



Note. These graphs show the result of regressing convincingness or outrage on a dummy equal to 1 if the argument is an externality argument, interacted with each socio-demographic variable. "*By education*" in 11a, for example, shows coefficients β^C and β^{NC} from the regression $y_{i,j} = \gamma_i + \beta^C College_i \times \alpha_j + \beta^{NC} NoCollege_i \times \alpha_j + \delta Length_j + \epsilon_{i,j}$, where $y_{i,j}$ is convincingness or outrage for individual i and argument j, γ_i is a person fixed effect, $College_i$ (resp. $NoCollege_i$) is a dummy equal to 1 if individual i has a College degree (resp. does not have a College degree), α_j is a dummy equal to 1 if argument j is an externality argument, and $Length_j$ is a dummy equal to 1 if argument j has above median length. Appendix H show the full question text. Standard errors are clustered at the argument level, 95% CIs.

changes in redistributive preferences. This leave-one-out design provides an independent test using only variation in other respondents' evaluations. Together, these tests provide a causal validation of the self-reported outcomes. We do not consider these results main outcomes due to a lack of statistical power.

In the following analysis we regress the post-evaluation measure on various features of the evaluated arguments, controlling for the pre-evaluation measure and demographic variables. The pre-evaluation measure is always strongly significant, as expected. For each pro-redistributive argument seen, respondents are 1.48 p.p. more likely to agree that the government should reduce income differences (95% C.I. [-0.15,3.12]; p=0.076). Splitting these arguments into fairness and inequality externality arguments, we find that the number of fairness arguments increases redistributive preferences at the 5% level (1.92 p.p., 95% C.I. [0.14, 3.69]; p=0.034), while the corresponding magnitude for externality arguments is on the same order of magnitude but insignificant at the 5% level (1.09 p.p., 95% C.I. [-0.66, 2.83]; p=0.223). The difference between the two argument types is not significant. This is reminiscent of the results we find in the self-reported convincingness measure, where fairness arguments were slightly more convincing than externality arguments but not in a statistically significant manner. These estimates increase in both magnitude and statistical significance when restricting the sample to respondents who passed an attention check administered immediately after the evaluations; the number of redistributive arguments seen becomes significant on the 1% level, for example (Table G5).

We then ask whether the convincingness of an argument affects redistributive preferences. First, increasing the share of arguments reported as convincing by the respondent by 10 p.p., or by one argument if the respondent sees only pro-redistributive arguments, is associated with being 6.77 p.p. more likely to agree that the government should reduce income differences (95% C.I. [6.19, 7.35]; p<0.001). Although this specification controls for pre-treatment redistributive preferences, it may still be affected by selection, as baseline attitudes are unlikely to be fully

Table 3: Effect of Arguments and their Characteristics on Redistributive Preferences

	Post-evaluation redistributive preferences					
	(1)	(2)	(3)	(4)	(5)	(6)
Pre-evaluation RP	0.294***	0.294***	0.142***	0.294***	0.142***	0.294***
	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)
N pro-redistr. arguments	0.015*		0.013*	0.015*		
	(0.008)		(0.008)	(0.008)		
N pro-redistr. arguments (fair)		0.019**			0.013	0.019**
		(0.009)			(0.008)	(0.009)
N pro-redistr. arguments (ext)		0.011			0.012	0.011
		(0.009)			(0.008)	(0.009)
Avg. convincingness, own			0.677***		0.677***	
			(0.030)		(0.030)	
Avg. outrage, own			-0.008		-0.008	
			(0.029)		(0.029)	
Avg. convincingness, others				0.634		0.631
				(0.474)		(0.475)
Avg. outrage, others				-0.491		-0.608
				(0.680)		(0.686)
Adjusted R^2	0.215	0.215	0.388	0.215	0.388	0.215
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2389	2389	2389	2389	2389	2389

Note. This table reports the effect of arguments and their characteristics on redistributive preferences. Each respondent evaluates ten redistributive arguments, of which 85% are pro-redistribution and 15% anti-redistribution on average. "Post-evaluation redistributive preferences" is a dummy for the respondent agreeing that "The government should take measures to reduce differences in income levels". *Pre-evaluation RP* is the respondent's pre-evaluation preference for redistribution, a dummy for answering above a 4 to "*How much redistribution of income do you prefer across citizens in the U.S.?*", where 0 is "*No redistribution*" and 7 is "*Full redistribution*" (results are robust to changing this threshold). *N pro-redistr. arguments* is the number of pro-redistributive arguments seen, overall and by type (fairness or externality). *Avg. convincingness, own* and *Avg. outrage, own* are the respondent's average self-reported evaluations of convincingness and outrage for the arguments they rated. *Avg. convincingness, others* and *Avg. outrage, others* are the corresponding leave-one-out averages across all other respondents evaluating the same arguments. These measures are used to test whether exposure to more, or more convincing, pro-redistributive arguments increases stated redistributive preferences. Included controls are standard demographic and socioeconomic characteristics: political leaning (Republican Leaning), gender (Male), race (Black, Other Race), income brackets ($25–50k, $50–100k, $100k+), age groups (30–39, 40–49, 50–59, 60–69, 70+), education (College or more), employment status (Unemployed, Not in workforce), and region (West, Northeast, Midwest). Including a control for passing an attention check administered immediately after the evaluations does not alter the results. Table G5 shows that the findings are virtually identical when restricting the sample to respondents who passed the check. Table G6 shows the distribution of pro-redistributive arguments seen. *Significance levels*: *10%, **5%, ***1%.

captured by the pre-evaluation measure. To address this concern, we also estimate a specification relating post-treatment redistributive preferences to the average convincingness of the arguments evaluated by all other respondents (leave-one-out). A 10 p.p. increase in the mean leave-one-out convincingness of the arguments seen increases demand for redistribution by 6.34 p.p.—very similar to the magnitude for self-reported convincingness, although the effect is not statistically significant (95% C.I. [-2.96, 15.64]; p=0.182). In contrast, neither a 10 p.p. increase in respondents' own self-reported agitation (-0.08 p.p., 95% C.I. [-0.64, 0.49]; p=0.789) nor a 10 p.p. increase in the mean leave-one-out agitation of the arguments seen (-4.91 p.p., 95% C.I. [-18.23, 8.40]; p=0.470) increases demand for redistribution. Overall these results indicate that the self-reported measures credibly capture whether respondents were swayed by the arguments in question.

The main limitation of the causal validation is the possibility of experimenter demand or priming, where respondents exposed to more pro-redistributive or more convincing arguments may report stronger post-evaluation support for redistribution without an actual change in preferences. While demand or priming effects cannot be fully ruled out, they would have to

arise systematically from the randomized exposure and replicate in the leave-one-out design—despite the latter relying exclusively on others' evaluations. This combination makes a pure demand-story less plausible.

Overall, the survey experiment yields three main findings. First, the LLM methodology appears to identify fairness- and externality-arguments well, and we recover similar content differences as in the legislative data. Second, the two argument types are about equally convincing on average, but fairness arguments are more likely to elicit outrage. Third, fairness arguments are relatively more persuasive for lower-education (and lower-income) respondents.

We now turn to additional robustness checks of the LLM methodology.

5 Robustness of the LLM methodology

This section reports further robustness and validation exercises for the LLM-based findings.

5.1 Consistency across runs

To account for stochastic variation in the LLM's responses, we classify each excerpt 20 times with the model's *temperature* variable set to 0.1 – a low but nonzero value introducing limited randomness. Across the 40 total runs (20 for each classification), the method proves highly consistent: the vast majority of excerpts are classified identically in all runs (Figure F1). Our main results are practically identical when defining an excerpt as fairness- or externality-based if it is classified as such in at least a certain number of runs (e.g. 19 of the 20 runs).

5.2 Alternative definitions of fairness and externality arguments

When comparing the content of the argument types, fairness and externality arguments are defined as speech excerpts that include one argument type but not the other. Appendix C.1 presents alternative definitions for fairness and externality arguments: (i) using *all* excerpts including an argument of the specific type, (ii) defining excerpts as including primarily fairness arguments, primarily externality arguments, or neither, using an alternate prompt.

Across all specifications, externality arguments are substantially more common in Norway. In our main specification, the Norwegian externality-to-fairness ratio is 2.3 higher than that in the U.S.; in the two alternate specifications, the equivalent numbers are 2.3 and 3.2, respectively.

The content differences we find with these alternative definitions largely resemble those obtained with the main definition (Figure F5), with two main changes. First, as expected, the differences between argument types are much smaller under these definitions; this is because these definitions allow excerpts to include arguments of both types, which mechanically allows overlap in the underlying excerpt sets. Second, under these definitions, externality arguments contain slightly more fear and pessimism in the legislative speeches – bringing the legislative results for these types of content in line with the experimental findings discussed in Section 4.

5.3 Alternative prompts

To avoid selection bias, we presented our main results using the first prompt for which we fully analyzed results. To explore whether our main results are robust to different prompts, we classify fairness-based and externality-based arguments in the legislative data using three alternative prompts for each argument type, that vary the prompt language in various ways, while keeping the underlying content of the prompt unchanged. The full text for these three additional prompts, which are the first three we tested, is in Appendix B.16–B.21

Across all specifications, externality arguments are substantially more common in Norway. In our main specification, the Norwegian externality-to-fairness ratio is 2.3 higher than that in the U.S.; in the three alternate specifications, the equivalent numbers are 3.8, 2.5, and 5.7.

Results are generally very consistent across these different prompts. The content analysis remains similar (Figure F6), although, as in other settings, fear and pessimism become more externality-coded. Villainization is also largely neutral under different prompts.

5.4 Alternative LLMs for classification

We check that our main results are robust to using alternative large language models to classify the arguments. Throughout the paper, all classifications are made using GPT-4o Mini, the first model with which we analyzed the data. In this section we use two alternative LLMs: GPT OSS 120B, an open-weight model with 117 billion parameters, presented by OpenAI as similar in performance to GPT-4o Mini; and the full DeepSeek R1 model, with 671 billion parameters. Both models were run remotely in the Ollama cloud, and use the exact same prompts as our main specification.

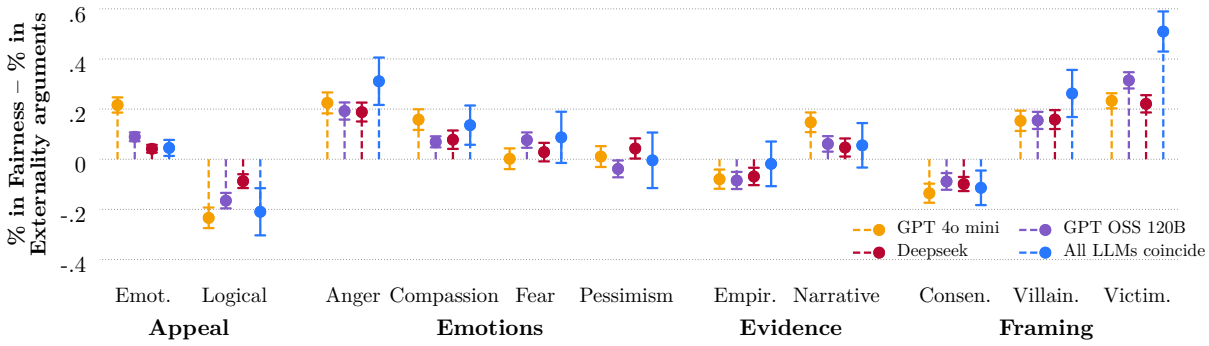
We run all main 13 classifications (Fairness, Externality, and the 11 content classifications) on the U.S. Congress data with GPT OSS 120B and DeepSeek R1 671B. The differences in content between fairness and externality arguments are largely consistent across models (Figure 12), with some changes in magnitudes. Whatever the LLM, fairness arguments contain significantly more emotional appeals, more anger, compassion, narrative evidence, villainization and victimization—while externality arguments are significantly more logical, include more empirical evidence, and are more consensus-seeking.

As an additional check, we also plot the content differences including only observations for which all three LLMs coincide: for example, an observation is only included in the computation of the emotional appeal difference if all LLMs returned the same fairness or externality classification, and the same emotional appeal classification. The results are very consistent with our baseline, and despite the smaller sample size, key differences are largely significant throughout. Overall, these results show that our main conclusions are not affected by the choice of LLM used for classification.

5.5 Word search

To assess the robustness of the LLM-based classification of fairness and externality arguments, we complement them with a simple word search approach. Table C4 shows the 20 words most characteristic of fairness and externality arguments in the U.S. data: that is, the words most

Figure 12: Content Gaps between Fairness and Externality Excerpts are Consistent across LLMs



Note. This figure reports content differences between fairness and externality arguments in the U.S. Congress data, using our preferred LLM (GPT-4o Mini) and two alternative LLMs: GPT OSS 120B and DeepSeek R1 671B. The differences are computed in the same way as in Figure 5, subtracting the share of externality arguments containing that characteristic to the share of fairness arguments containing that characteristic. The figure also displays content differences for arguments for which all LLMs coincide—that is, for which all LLMs returned the same fairness and externality classification, as well as the same content characteristic classification. The prompts are those shown in Appendix B. 95% confidence intervals are shown.

overrepresented in fairness (resp. externality) arguments, compared to their frequency in the corpus overall. The text processing methodology is detailed in Appendix C.4. Table C5 and C6 show the same results for the Norwegian *Storting* and the experimental sample.

Across all samples, fairness arguments emphasize distributive justice and individual effort, whereas externality arguments focus on societal and economic outcomes. In the legislative data, fairness excerpts frequently include terms such as fair, deserve, equal, and hard work, while externality excerpts feature economy, growth, investment, and community. The same pattern appears in the experimental data: fairness arguments highlight words such as money, work, and job, whereas externality arguments refer to words such as economy, society, and crime.

We construct simple keyword searches to check whether the classifications align with intuitive patterns. The word "*fair*" appears in 8.5% of fairness-based excerpts against 0.9% of externality-based excerpts (typically in passing, e.g. "*It is fair to say that...*"), whereas the word "*society*" appears in 3.0% of externality-based excerpts against 0.8% of fairness-based excerpts. We can also ensure that the main content differences across argument types appear in these simplistic indicators. Fairness-based excerpts contain a higher share of each of the words "*angry*," "*outrageous*," "*betrayed*," "*disgraceful*," "*shameful*," or "*cheated*" than externality-based excerpts in the Congressional data. 2.8% of fairness-based excerpts include at least one of these words, compared with 0.2% of externality-based excerpts. Similarly, excerpts including the word "*fair*" are more likely to include any of these words (2.3%) than excerpts including the word "*society*" (1.5%)—simplistically replicating the finding that fairness arguments contain more anger than externality arguments. While such word searches are a helpful benchmark, the small share of flagged arguments also underscores the limitations of this approach, and the value of using an LLM for the main classification.

Overall, our various robustness checks show that our main results are consistent across methodologies and specifications.

6 Conclusion

Much of the public debate—and much of the economics literature—approaches redistribution through the lens of fairness. The opening quote of this paper, delivered by Jens Stoltenberg in the Norwegian *Storting* in October 2025, reflects a different logic:

"*The Norwegian social model requires higher public spending and higher taxes than in many other countries. In return, it provides security [...], trust between people, and a productive economy.*"

Rather than appealing to economic fairness, Stoltenberg frames redistribution as a collective investment that improves societal performance, linking higher taxes to security, trust, and productivity. This type of reasoning—treating inequality as a negative externality rather than a moral wrong—is what our analysis identifies as an externality argument. Many of our main results are reflected in the argument, which is based on logic, has low emotional content, and seeks consensus. It does not contain anger or compassion. Later in the speech, Stoltenberg solidifies his argument with empirical evidence.

In this paper we have formalized and documented these patterns. We have shown that fairness arguments are more likely to be emotional, particularly containing anger and compassion, while externality arguments appeal to logic and seek consensus. Externality arguments are much more common in the Norwegian *Storting* than in the U.S. Congress, and while both arguments appear equally convincing to U.S. survey respondents, fairness arguments elicit more outrage and are linked to lower educational attainment. To our knowledge, this is the first study to conduct a large-scale empirical analysis of the redistributive debate.

The documented cross-country differences naturally raise questions of causality. One possibility is that the structure of the redistributive debate shapes redistributive outcomes. Fairness-based debates may be more likely to spur large-scale redistributive change, for example, while externality-based debates may be more likely to preserve existing redistributive systems. If so, it is pertinent to ask how the Norwegian and American redistributive debates became so different. An alternative possibility is that higher inequality in the U.S. has shifted the redistributive debate toward fairness. If so, it is pertinent to ask whether a more fairness-centered debate in turn affects society – whether it increases affective polarization or political gridlock, for example. Such a dynamic could generate feedback loops in which higher inequality shifts the debate toward fairness, fairness-centered debate increases affective polarization and political gridlock, polarization and gridlock weakens redistributive policy, and weaker redistribution in turn sustains or increases inequality. The contemporary American debate, characterized by high affective intensity, limited policy change, and rising inequality, is consistent with such a feedback mechanism. We stress, however, that these mechanisms are necessarily speculative, and that one could also imagine very different dynamics. Rising educational attainment may gradually shift political discourse toward analytically framed externality-based arguments, for example. Overall, our evidence is descriptive and does not identify any causal direction; these mechanisms are presented as open questions for future work.

Our approach also provides a methodological contribution. While prior studies focus on specific stimuli, we compare representative samples of *universes* of arguments. This approach,

which combines the emergence of large datasets with innovations in natural language processing, allows us to study both the content and evaluations of real-world rhetorical inputs—bringing the empirical literature closer to the environments in which citizens actually form opinions. A similar strategy can be applied to other parts of the redistributive debate, or to other domains such as climate policy or immigration. If we step back from policy debates and instead apply the same methodology to *statements*, the space becomes even larger—as our method could be used to compare statements from experts and lay-people, for example, or narratives pre- and post-treatment for any major policy shock. The paper offers a general template for *comparing universes of arguments* (or statements) which we hope can be broadly applied across policy areas.

References

Aaberge, Rolf, Anthony B Atkinson, and Jørgen Modalsli. Estimating Long-Run Income Inequality from Mixed Tabular Data: Empirical Evidence from Norway, 1875–2017. *Journal of Public Economics*, 187:104196, 2020.

Alesina, Alberto, Stefanie Stantcheva, and Edoardo Teso. Intergenerational Mobility and Preferences for Redistribution. *American Economic Review*, 108(2):521–554, 2018.

Alesina, Alberto, Armando Miano, and Stefanie Stantcheva. Immigration and redistribution. *The Review of Economic Studies*, 90(1):1–39, 2023.

Algan, Yann, Eva Davoine, Thomas Renault, and Stefanie Stantcheva. Emotions and Policy Views. *Harvard University Working Paper*, 2025.

Almås, Ingvild, Alexander W. Cappelen, and Bertil Tungodden. Cutthroat Capitalism versus Cuddly Socialism: Are Americans More Meritocratic and Efficiency-Seeking than Scandinavians? *Journal of Political Economy*, 128(5):1753–1788, 2020.

Almås, Ingvild, Alexander W Cappelen, Bertil Tungodden, and Erik Ø. Sørensen. Fairness Across the World, 2025.

Andre, Peter, Carlo Pizzinelli, Christopher Roth, and Johannes Wohlfart. Subjective Models of the Macroeconomy: Evidence From Experts and Representative Samples. *The Review of Economic Studies*, 90(1):1–35, 2023.

Aroyehun, Segun Taofeek, Almog Simchon, Fabio Carrella, Jana Lasser, Stephan Lewandowsky, and David Garcia. Computational Analysis of US Congressional Speeches Reveals a Shift From Evidence to Intuition. *arXiv preprint arXiv:2405.07323*, 2024.

Ash, Elliott, Germain Gauthier, and Philine Widmer. Relatio: Text Semantics Capture Political and Economic Narratives. *Political Analysis*, 32(1):115–132, 2024.

Cappelen, Alexander W, Astri Drange Hole, Erik Ø Sørensen, and Bertil Tungodden. The Pluralism of Fairness Ideals: An Experimental Approach. *The American Economic Review*, 97(3):818–827, 2007.

Cruces, Guillermo, Ricardo Perez-Truglia, and Martin Tetaz. Biased Perceptions of Income Distribution and Preferences for Redistribution: Evidence from a Survey Experiment. *Journal of Public Economics*, 98:100–112, 2013.

Durante, Ruben, Louis Putterman, and Joël van der Weele. Preferences for Redistribution and Perception of Fairness: An Experimental Study. *Journal of the European Economic Association*, 12(4):1059–1086, 2014.

Fabre, Adrien, Thomas Douenne, and Linus Mattauch. International Attitudes Toward Global Policies, 2024.

Fiva, Jon H., Oda Nedregård, and Henning Øien. The Norwegian Parliamentary Debates Dataset. *Scientific Data*, 12(1):4, 2025.

Gärtner, Manja, Johanna Möllerström, and David Seim. Income Mobility, Luck/Effort Beliefs, and the Demand for Redistribution: Perceptions and Reality. *Working Paper*, 2019.

Gennaro, Gloria and Elliott Ash. Emotion and Reason in Political Language. *The Economic Journal*, 132(643):1037–1059, 2022.

Heseltine, Michael and Bernhard Clemm von Hohenberg. Large Language Models as a Substitute for Human Experts in Annotating Political Text. *Research & Politics*, 11(1): 20531680241236239, 2024.

Hvidberg, Kristoffer B, Claus T Kreiner, and Stefanie Stantcheva. Social Positions and Fairness Views on Inequality. *Review of Economic Studies*, 90(6):3083–3118, 2023.

Karadja, Mounir, 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, 2017.

Klor, Esteban F. and Moses Shayo. Social Identity and Preferences over Redistribution. *Journal of Public Economics*, 94(3-4):269–278, 2010.

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

Kuziemko, Ilyana, Michael I Norton, Emmanuel Saez, and Stefanie Stantcheva. How Elastic Are Preferences for Redistribution? Evidence from Randomized Survey Experiments. *The American Economic Review*, 105(4):1478–1508, 2015.

Kuziemko, Ilyana, Nicolas Longuet-Marx, and Suresh Naidu. "Compensate the Losers?" Economic Policy and Partisan Realignment in the US. Working Paper 31794, National Bureau of Economic Research, 2023.

Le Mens, Gaël and Aina Gallego. Positioning Political Texts with Large Language Models by Asking and Averaging. *Political Analysis*, 33(3):274–282, 2025.

Lobeck, Max and Morten Nyborg Støstad. The Consequences of Inequality: Beliefs and Redistributive Preferences. *CESifo Working Paper No. 10710*, 2023.

Longuet-Marx, Nicolas. Party Lines or Voter Preferences? Explaining Political Realignment. *Personal Website, https://nicolaslonguetmarx.github.io/PartyLines_NLM.pdf*, 2025.

Luttmer, Erzo F P. Group Loyalty and the Taste for Redistribution. *Journal of political Economy*, 109(3):500–528, 2001.

Luttmer, Erzo F. P and Monica Singhal. Culture, Context, and the Taste for Redistribution. *American Economic Journal: Economic Policy*, 3(1):157–179, 2011.

Martin, Isaac William, Ajay K Mehrotra, and Monica Prasad. *The New Fiscal Sociology: Taxation in Comparative and Historical Perspective*. Cambridge University Press, 2009.

Meltzer, Allan H. and Scott F. Richard. A Rational Theory of the Size of Government. *Journal of Political Economy*, 89(5):914–927, 1981.

Piketty, Thomas, 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.

Roth, Christopher, Sonja Settele, and Johannes Wohlfart. Beliefs about Public Debt and the Demand for Government Spending. *Working Paper*, page 89, 2020.

Rueda, David 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, 2016.

Saez, Emmanuel and Gabriel Zucman. The Rise of Income and Wealth Inequality in America: Evidence from Distributional Macroeconomic Accounts. *Journal of Economic Perspectives*, 34(4):3–26, 2020.

Scheve, Kenneth and David Stasavage. *Taxing the Rich: A History of Fiscal Fairness in the United States and Europe*. Princeton University Press, Princeton, 2016.

Shiller, Robert J. Narrative Economics. *American Economic Review*, 107(4):967–1004, 2017.

Soria, Chris. An Empirical Investigation into the Utility of Large Language Models in Open-Ended Survey Data Categorization, 2025.

Stantcheva, Stefanie. Understanding Tax Policy: How Do People Reason? *The Quarterly Journal of Economics*, 136(4):2309–2369, 2021.

Stantcheva, Stefanie. Perceptions and Preferences for Redistribution. *Oxford Open Economics*, 3:96–100, 2024.

Støstad, Morten Nyborg and Frank Cowell. Inequality as an Externality: Consequences for Tax Design. *Journal of Public Economics*, 235(105139), 2024.

World Inequality Lab, . 500 Economists and Inequality Experts from Seventy Countries Support Call for New "IPCC for Inequality", 2025.

Yang, Eddie, Zoey Wang, Carl Zhou, and Yaosheng Xu. Data Annotation with Large Language Models: Lessons from A Large Empirical Evaluation. *Unpublished manuscript*, 2025.

Appendix

A Formal definitions: Fairness and inequality externalities

Throughout the paper we discuss two types of arguments based on *fairness* and *inequality externalities*. As we formally define below, fairness arguments evaluate the distribution of resources itself, while inequality externality arguments evaluate the societal consequences that arise from that distribution. Our analysis abstracts from implementation or feasibility concerns and focuses on pro-redistributive arguments, as anti-redistributive arguments differ substantially in content and focus.

To formally define the argument types, we introduce a minimal welfare framework. We assume that any argument for redistribution implicitly claims that the distribution of resources influences some welfare-relevant consideration – where "welfare" is interpreted broadly, in that it may or may not include individual well-being or any other principles one considers socially important. We do not take a stand on what citizens or legislators believe, nor assume they share a common ethical criterion. If individual well-being is relevant for social welfare, it is so through the utility function $U_i(\cdot)$, where we focus on three components: (i) own economic resources x_i, (ii) the distribution of resources θ, (iii) any societal consequences of inequality that the individual cares about, captured by a function $\Gamma_i(\theta)$, which may depend on θ. To remain as general as possible we also allow overall welfare to depend directly on both the distribution itself and societal consequences. Overall welfare is written as,

$$\mathbf{W} = W(\, U_{i\in I}\,(x_i, \theta, \Gamma_i(\theta), ...)\,, \theta, \Gamma_{i\in I}(\theta), ...)$$

where we place no further restrictions on the form of $W(\cdot)$ or $U(\cdot)$. In this representation, the distribution of resources may affect social welfare in three distinct ways.

First, it may affect individuals' well-being through their own resources x_i. A familiar channel would be diminishing marginal utility, which could lead to an argument that a poorer person has more use for resources than a richer one. Second, the distribution itself may directly affect either individual utility or social welfare. This could occur via various normative principles such as egalitarianism, rights, or desert; individuals may derive higher well-being in a more equal society, or they may value a more equal society independently (e.g. based on principles; our framework does not assume revealed preferences). Arguments that focus on one of these two channels are what we refer to as *fairness* arguments.

Third, the distribution may also affect social welfare through the societal consequences of inequality captured by $\Gamma_i(\theta)$. While these effects typically relate to individual well-being directly (because crime is unpleasant, for example), we also allow that they may affect social welfare beyond individual well-being (for example because democratic functioning is valued intrinsically). Arguments that focus on this third channel are what we refer to as *inequality externality* arguments.

A large share of pro-redistributive arguments can be cast in one of these forms, and we discuss potential subtypes in Appendix C.3. Arguments may also mix the two types; in the empirical specification we exclude these arguments to cleanly separate fairness and externality

arguments (e.g. "*Inequality leads to crime, which is unfair for the poorest*"). Still, certain arguments are either not captured or unclearly defined by this approach, for example reciprocal arguments, which might resemble externality arguments but have little to do with inequality *itself*, or simple assertions without further conceptual meaning (e.g. "inequality should be reduced"). We therefore allow for a residual category of *neither* arguments which includes any arguments that are not captured by the above framework.

B Usage of GPT-4o Mini for content classification

Speech excerpts were classified using the GPT-4o Mini API across multiple outcomes. Each outcome for each speech was classified independently, with argument order randomized before making API calls. In total, we made 3,893,098 separate API calls. The following outcomes were classified:

1. **In favor of redistribution:** Whether the excerpt includes an argument in favor of increasing economic redistribution. This classification was made on the full sample of 760,019 (U.S. Congress) and 312,447 (Norwegian *Storting*) speech excerpts; all other classifications were made on the subsample of 25,366 and 17,098 excerpts classified as in favor of redistribution.
2. **Fairness argument:** Whether the excerpt includes an argument that frames economic inequality in terms of fairness or justice
3. **Inequality externality argument:** Whether the excerpt includes an argument that frames economic inequality as a harmful externality
4. **Emotional appeal:** Whether the excerpt appeals to emotion
5. **Logical appeal:** Whether the excerpt appeals to logic, reasoning, or factual evidence
6. **Anger:** Whether the excerpt contains anger
7. **Compassion:** Whether the excerpt contains compassion
8. **Fear:** Whether the excerpt contains fear or concern
9. **Pessimism:** Whether the excerpt contains pessimism
10. **Empirical evidence:** Whether the excerpt contains empirical evidence
11. **Narrative evidence:** Whether the excerpt contains narrative evidence
12. **Consensus-seeking:** Whether the excerpt is consensus-seeking
13. **Villainizing:** Whether the excerpt villainizes any specific group
14. **Victimizing:** Whether the excerpt victimizes any specific group
15. **Fairness/Externality argument (Primarily):** Whether the excerpt primarily includes an argument that frames economic inequality in terms of fairness or justice, or primarily an argument that frames economic inequality as a harmful externality
16. **Fairness/Externality argument (Alternative prompts 2, 3, and 4):** We use three alternative prompts for each classification, which resemble the main Fairness argument and Externality argument prompts in susbtance but make the language vary

17. **Fairness argument subtypes:** Whether the excerpt includes a needs-based fairness argument, a rights-based fairness argument, or a merit-based fairness argument (see Appendix C.3)
18. **Externality argument subtypes:** Whether the excerpt includes a top-income or a bottom-income externality argument (see Appendix C.3)

Below we list the full prompts for all 14 cases.

B.1 Prompt to GPT-4o Mini: In favor of redistribution

```
You are a classifier.

You are analyzing a Congressional speech to determine whether
    it includes an argument **in favor of increasing economic
   redistribution**.
Give this text a score of either:
- 1 = The text **does** make an argument in favor of more
   economic redistribution (e.g., advocating higher taxes on
   the wealthy, expanding social programs, increasing minimum
   wage, redistributing income or wealth, etc.)
- 0 = The text **does NOT** make such an argument (e.g., it's
    neutral, unrelated, or argues against redistribution)
Do not explain the score. DO NOT include any additional text,
    labels, or the word "Score:" before the number. Just
   return a single digit: either "0" or "1".
Argument to evaluate: [Insert argument]
```

B.2 Prompt to GPT-4o Mini: Fairness argument

User prompt:

```
[Insert argument]
```

System prompt:

```
You are analyzing a Congressional speech to determine whether
    it includes an argument for reducing inequality, or for
   redistribution, that is framed around fairness or justice
   -- specifically, discussing economic differences as unfair,
    unjust, undeserved, or morally wrong, or invoking related
   ideas such as procedural fairness, the rights of workers or
    citizens, the unjust differences between groups, or duty-
   based reasoning. These arguments must go beyond discussing
   the effects on broader societal outcomes (e.g., democracy,
   trust, stability, crime, growth, cohesion, the integrity of
    public institutions, the concentration of power), and
   instead emphasize or imply principles of fairness, justice,
    or morality.

Give this text a score of either:

1 = The text does make an argument based on fairness or
   justice -- emphasizing what people deserve, what is right
   or wrong, or appealing to fairness, justice, or moral
   principles.
```

```
0 = The text does NOT make such an argument (e.g., it is
    unrelated, or focuses only on societal consequences,
    efficiency, practical outcomes, direct benefits to
    recipients without invoking fairness, or policy mechanics).

Do not infer based solely on tone or rhetorical form. Focus
    on the substance or strongly implied themes of the argument
    .

Do not explain the score. DO NOT include any additional text,
     labels, or the word "Score:" before the number. Just
    return a single digit: either "0" or "1".
```

B.3 Prompt to GPT-4o Mini: Inequality externality argument

User prompt:

```
[Insert argument]
```

System prompt:

```
You are analyzing a Congressional speech to determine whether
     it includes an argument for reducing inequality, or for
    redistribution, that frames economic inequality as a
    harmful externality -- specifically, discussing economic
    differences as something that harms society as a whole, not
     just individuals. These arguments must go beyond
    discussing fairness, the gains of direct beneficiaries, or
    individual gains/losses, and instead emphasize or imply the
     effects on broader societal outcomes (e.g., democracy,
    trust, stability, crime, growth, cohesion, the integrity of
     public institutions, the concentration of power).

Give this text a score of either:

1 = The text does make an argument based on inequality
    causing negative externalities at the societal level --
    emphasizing that reducing inequality at the top benefits
    society, that supporting the bottom benefits society, or
    that reducing general economic differences benefits society
    .

0 = The text does NOT make such an argument (e.g., it is
    unrelated, or focuses only on fairness, distributive
    justice, direct benefits to recipients without invoking
    societal consequences, or policy mechanics).

Do not infer based solely on tone or rhetorical form. Focus
    on the substance or strongly implied themes of the argument
    .

Do not explain the score. DO NOT include any additional text,
     labels, or the word "Score:" before the number. Just
    return a single digit: either "0" or "1".
```

B.4 Prompt to GPT-4o Mini: Emotional appeal

User prompt:

```
[Insert argument]
```

System prompt:

```
You are an objective classifier of emotional appeal in
   written arguments. For each argument below, respond with 0
   or 1 depending on whether it appeals to emotion to persuade
    the reader.
Respond with 1 if the argument contains emotional appeal -
   for example, if it uses emotionally charged language,
   personal stories, vivid imagery, or appeals to fear, hope,
   pride, anger, or compassion to make its case.
Respond with 0 if the argument does not contain emotional
   appeal - for example, if it remains neutral in tone and
   does not use emotional language or attempt to evoke
   feelings in the reader.
Focus only on emotional cues in the text itself. Do not
   evaluate whether the argument is logical, persuasive, or
   factually correct - only whether it uses emotion as a
   persuasive strategy.
```

B.5 Prompt to GPT-4o Mini: Logical appeal

User prompt:

```
[Insert argument]
```

System prompt:

```
You are an objective classifier of logical appeal in written
   arguments. For each argument below, respond with 0 or 1
   depending on whether it appeals to logic, reasoning, or
   factual evidence to persuade the reader.
Respond with 1 if the argument contains logical appeal - for
   example, if it uses facts, statistics, definitions,
   comparisons, or causal reasoning to support its point.
Respond with 0 if the argument does not contain logical
   appeal - for example, if it relies only on personal opinion
   , emotion, anecdote, or rhetorical style without appealing
   to reasoning or evidence.
Focus solely on the presence of logical or factual reasoning
   in the text itself. Do not consider whether the argument is
    true, persuasive, or well-written - only whether it uses
   logic or evidence as a persuasive strategy.
```

B.6 Prompt to GPT-4o Mini: Anger

User prompt:

```
[Insert argument]
```

System prompt:

```
You are an objective emotion classifier. For each argument
   below, determine whether it expresses anger, based on its
   wording and tone. Respond with 1 if the argument expresses
   anger, frustration, resentment, irritation, or outrage.
   Respond with 0 if the argument does not express those
   emotions. Focus only on emotional cues present in the text
   itself.
```

B.7 Prompt to GPT-4o Mini: Compassion

User prompt:

```
[Insert argument]
```

System prompt:

```
You are an objective emotion classifier. For each argument
   below, analyze whether it expresses compassion, based
   solely on its wording and tone. Respond with 1 if the
   argument expresses compassion, or 0 if it does not.
   Identify expressions of empathy, care, or concern for
   others.
```

B.8 Prompt to GPT-4o Mini: Fear

User prompt:

```
[Insert argument]
```

System prompt:

```
You are an objective emotion classifier. For each
   argument below, determine whether it expresses fear or
   concern, based on its wording and tone. Respond with 1
   if the argument expresses fear, anxiety, worry, concern
   , dread, or a sense of threat or vulnerability. Respond
    with 0 if the argument does not express any of these.
   Focus only on cues within the text itself.
```

B.9 Prompt to GPT-4o Mini: Pessimism

User prompt:

```
[Insert argument]
```

System prompt:

```
You are an objective emotion classifier. For each argument
   below, analyze whether it expresses pessimism, based solely
    on its wording and tone. Respond with 1 if the argument
   expresses pessimism, or 0 if it does not. Focus only on
   textual cues indicating doubt, hopelessness, or negative
   expectations.
```

B.10 Prompt to GPT-4o Mini: Empirical evidence

User prompt:

```
[Insert argument]
```

System prompt:

```
You are an objective argument classifier. For the argument
   below, determine whether it presents empirical evidence,
   based on its content and references. Respond with 1 if the
   argument includes empirical evidence -- such as observed
   data, measurements, experiments, or real-world examples.
   Respond with 0 if the argument does not contain such
   evidence. Focus only on indicators of empirical evidence
   present in the text itself.
```

B.11 Prompt to GPT-4o Mini: Narrative-based evidence

User prompt:

```
[Insert argument]
```

System prompt:

```
You are an objective argument classifier. For the argument
   below, determine whether it relies on anecdotal or story-
   based evidence to persuade. Respond with 1 if the argument
   uses a personal anecdote, individual case, vivid story, or
   selective example instead of (or in addition to) aggregated
    data or formal logical analysis. Respond with 0 if it does
    not. Focus only on narrative cues present in the text
   itself.
```

B.12 Prompt to GPT-4o Mini: Consensus-seeking

User prompt:

```
[Insert argument]
```

System prompt:

```
You are an objective argument classifier. For the argument
   below, determine whether it presents its case with the goal
    of broad consensus agreement, as opposed to appealing to a
    particular subgroup. Respond with 1 if the argument tries
   to frame its position as universally desirable or shared by
    ``all of us," ``everyone," or ``society as a whole."
   Respond with 0 if it does not. Focus only on consensus-
   oriented cues present in the text itself.
```

B.13 Prompt to GPT-4o Mini: Villainizing

User prompt:

```
[Insert argument]
```

System prompt:

```
You are an objective argument classifier. For the argument
   below, determine whether it villainizes a person, group, or
    entity through its wording and tone. Respond with 1 if the
    argument portrays a target as malicious, immoral,
   dangerous, or otherwise deserving blame or contempt.
   Respond with 0 if it does not. Focus only on villainizing
   cues present in the text itself.
```

B.14 Prompt to GPT-4o Mini: Victimizing

User prompt:

```
[Insert argument]
```

System prompt:

```
You are an objective argument classifier. For the argument
   below, determine whether it portrays a person, group, or
   entity as a victim of deliberate harm or oppression through
    its wording and tone. Respond with 1 if the argument
   depicts the target as attacked, persecuted, abused,
   exploited, or otherwise suffering intentional injury.
   Respond with 0 if it does not. Focus only on victimization
   cues present in the text itself.
```

B.15 Prompt to GPT-4o Mini: Primarily classification of fairness and externality speech excerpts

User prompt:

```
[Insert argument]
```

System prompt:

```
You are analyzing an English or Norwegian speech to determine
   whether it includes:
EXTS = an argument for reducing inequality, or for redistribution
   , that frames economic inequality as a harmful externality --
   specifically, discussing economic differences as something that
    harms society as a whole, not just individuals. These
   arguments must go beyond discussing fairness, the gains of
   direct beneficiaries, or individual gains/losses, and instead
   emphasize or imply the effects on broader societal outcomes (e.
   g., democracy, trust, stability, crime, growth, cohesion, the
   integrity of public institutions, the concentration of power).
FAIR = an argument for reducing inequality, or for redistribution
   , that is framed around fairness or justice --- specifically,
   discussing economic differences as unfair, unjust, undeserved,
   or morally wrong, or invoking related ideas such as procedural
   fairness, the rights of workers or citizens, the unjust
   differences between groups, or duty-based reasoning. These
   arguments must go beyond discussing the effects on broader
   societal outcomes (e.g., democracy, trust, stability, crime,
   growth, cohesion, the integrity of public institutions, the
   concentration of power), and instead emphasize or imply
   principles of fairness, justice, or morality.

Give the text a score of either:
EXTS = The text does make an argument based on inequality causing
    negative externalities at the societal level --- emphasizing
   that reducing inequality at the top benefits society, that
   supporting the bottom benefits society, or that reducing
   general economic differences benefits society. Do not classify
   if the text does NOT make such an argument (e.g., it is
   unrelated, or focuses only on fairness, distributive justice,
   direct benefits to recipients without invoking societal
   consequences, or policy mechanics).
FAIR = The text does make an argument based on fairness or
   justice --- emphasizing what people deserve, what is right or
   wrong, or appealing to fairness, justice, or moral principles.
```

```
   Do not classify if the text does NOT make such an argument (e.g
   ., it is unrelated, or focuses only on societal consequences,
   efficiency, practical outcomes, direct benefits to recipients
   without invoking fairness, or policy mechanics).
OTHER = The text does not make either type of argument explicitly
   .
If both types appear, select the one that is most central to the
   argument.

Do not infer based solely on tone or rhetorical form. Focus on
   the substance or strongly implied themes of the argument.
Do not explain the score. DO NOT include any additional text,
   labels, or the word "Score:" before the number.
Just return a single word classifying the speech: "EXTS" for
   externalities, "FAIR" for fairness, or "OTHER" for other.
```

B.16 Prompt to GPT-4o Mini: Fairness argument, prompt 2

User prompt:

```
[Insert argument]
```

System prompt:

```
You are analyzing a speech in [English/Norwegian] to
   determine whether it contains an argument for more
   redistribution based on fairness. An argument is based on
   fairness if it appeals to some idea of justice, morality,
   that some people inherently deserve certain allocation, and
    that those ideals are not satisfied in the status quo. An
   argument based on fairness contains the idea that we should
    redistribute more, because the current distribution of
   resources is unfair. If the argument is mainly based on a
   different idea, for example that the status quo is
   inefficient, that it has negative consequences on society
   as a whole, or simply states that we should redistribute
   more without explaining why, you should conclude that it
   does NOT make an argument based on fairness. Analyse the
   content of the argument, then return either:
0, if the text does not contain a pro-redistributive argument
    based on fairness;
1, if the text contains a pro-redistributive argument based
   on fairness.
Your output should be a single digit, 0 or 1, and nothing
   else. Do not infer what might be indirectly implied, focus
   on what is clearly expressed in the text.
```

B.17 Prompt to GPT-4o Mini: Externality argument, prompt 2

User prompt:

```
[Insert argument]
```

System prompt:

```
You are analyzing a speech in [English/Norwegian] to
   determine whether it contains an argument for more
   redistribution based on inequality externalities on society
   . An argument is based on inequality externalities if it
```

```
highlights that inequality has harmful consequences on
society as a whole, for example because it fragments
society, increases crime, or leads to democratic capture.
An argument based on inequality externalities contains the
idea that we should redistribute more, because the current
level of inequality harms society as a whole. If the
argument is mainly based on a different idea, for example
that the status quo is unfair, or simply states that we
should redistribute more without explaining why, you should
 conclude that it does NOT make an argument based on
inequality externalities. Analyse the content of the
argument, then return either:
0, if the text does not contain a pro-redistributive argument
 based on inequality externalities;
1, if the text contains a pro-redistributive argument based
on inequality externalities.
Your output should be a single digit, 0 or 1, and nothing
else. Do not infer what might be indirectly implied, focus
on what is clearly expressed in the text.
```

B.18 Prompt to GPT-4o Mini: Fairness argument, prompt 3

User prompt:

```
[Insert argument]
```

System prompt:

```
You are an objective and robust analyst of arguments' content
. Your task is to analyze a speech in [English/Norwegian].
This text contains an argument for more redistribution, and
 we are trying to determine whether this arguments is based
 on fairness or not. An argument based on fairness includes
 some idea of what is fair, what is just, what is morally
right, or an idea of deservedness and merit; it then argues
 that we should redistribute more, because the current
state of affairs is not in line with those ideals.
You should analyze the argument's content, then return either
 "FAIR" if the argument is based on fairness, or "NOT" if
the argument is not based on fairness. You should tag an
argument as a fairness argument if and only if fairness is
a central tenet of its reasoning. Focus on what is clearly
expressed in the text for your analysis. Do not include
your reasoning in your answer, only "FAIR" or "NOT".
```

B.19 Prompt to GPT-4o Mini: Externality argument, prompt 3

User prompt:

```
[Insert argument]
```

System prompt:

```
You are an objective and robust analyst of arguments' content
. Your task is to analyze a speech in [English/Norwegian].
This text contains an argument for more redistribution, and
 we are trying to determine whether this arguments is based
 on inequality externalities or not. An argument based on
inequality externalities includes the idea that inequality
```

```
may have negative consequences for society in general,
above and beyond its effects on individuals. For example,
they may argue that inequality increases crime, accelerates
 climate change, or stifles growth, and that we should
therefore curb inequality to make society as a whole better
 off.
You should analyze the argument's content, then return either
 "EXT" if the argument is based on inequality externalities
, or "NOT" if the argument is not based on inequality
externalities. You should tag an argument as an inequality
externalities argument if and only if inequality
externalities are a central tenet of its reasoning. Focus
on what is clearly expressed in the text for your analysis.
 Do not include your reasoning in your answer, only "EXT"
or "NOT".
```

B.20 Prompt to GPT-4o Mini: Fairness argument, prompt 4

User prompt:

```
[Insert argument]
```

System prompt:

```
You are analyzing a speech in [English/Norwegian]. Your task
   is to decide whether the text includes a pro-redistributive
    argument that is grounded in fairness or justice.

Such an argument frames existing economic inequality as
   unfair, unjust, undeserved, or morally wrong, or appeals to
    ideas of justice, moral obligation, or what people are
   entitled to. The emphasis should be on fairness or moral
   principles, not on broader social outcomes.

If the text focuses primarily on societal consequences (e.g.
   growth, stability, democracy, crime), efficiency, or policy
    details, or if it supports redistribution without
   appealing to fairness or justice, then it does NOT count.

Do not go to deep in the interpretation of what the text
   might imply; base yourself on what is clearly expressed.

Return:
FAIR - if the text makes a fairness- or justice-based
   argument for redistribution.
NOT - otherwise.

Return only one word: FAIR or NOT.
```

B.21 Prompt to GPT-4o Mini: Externality argument, prompt 4

User prompt:

```
[Insert argument]
```

System prompt:

```
You are analyzing a speech in [English/Norwegian]. Decide
   whether the text contains a pro-redistribution argument
   that treats economic inequality as harmful to society as a
   whole.

Such arguments emphasize that inequality produces negative
   consequences for society (e.g. weakened democracy, reduced
   trust, instability, crime, lower growth, or institutional
   damage), and that redistribution is justified because of
   these broader effects.

If the argument instead focuses on fairness, justice, or
   individual-level benefits without linking inequality to
   societal-level harms, it does NOT count.

Do not go to deep in the interpretation of what the text
   might imply; base yourself on what is clearly expressed.

Return:
EXT - if the text contains an inequality-externalities
   argument for redistribution.
NOT - otherwise.

Return only one word: EXT or NOT.
```

B.22 Prompt to GPT-4o Mini: Fairness argument subtypes (Needs-based, Rights-based, Compensatory)

User prompt:

```
[Insert argument]
```

System prompt:

```
You are analyzing a Congressional speech to determine whether
    it includes an argument for reducing inequality, or for
   redistribution, that is framed around fairness or justice.
Specifically, decide which of the three fairness themes below
    the speech excerpt relies the most on and output the
   corresponding number.
If none of the three clearly applies, choose NEITHER.

Definitions:

NEED  = Need- or sufficiency-based fairness (diminishing
   marginal utility at the bottom, poverty relief, basic-needs
    floor, ability-to-pay arguments, the idea that some people
    need the money more than others).
RIGHTS = Equal-rights or egalitarian fairness (equal moral
   worth, moral obligations, equal opportunity, appeals to
   human or equal rights, strict egalitarianism, luck-
   egalitarian or Rawlsian claims that it is unjust or wrong
   that some have so little).
COMP  = Compensatory- or merit-based fairness (deservingness,
    due reward for effort, compensation for harm or
   discrimination or sacrifice, workers having earned benefits
   ).
Return exactly one of these four numbers, with NO additional
   text or punctuation:

NEITHER=0
```

```
NEED=1
RIGHTS=2
COMP=3

Do not infer based solely on tone or rhetorical form. Focus
   on the substance or strongly implied themes of the argument
   .
If multiple fairness themes appear, pick the single most
   dominant one. If no clear fairness themes appear, choose
   NEITHER.
Do NOT explain the choice.  Just output one token: 0, 1, 2,
   or 3.
```

B.23 Prompt to GPT-4o Mini: Externality argument subtypes (Top-income, Bottom-income)

User prompt:

```
[Insert argument]
```

System prompt:

```
You are analyzing a Congressional speech to determine whether
    it includes an argument for reducing inequality, or for
   redistribution, that frames economic inequality as a
   harmful externality - specifically, discussing economic
   differences as something that harms society as a whole, not
    just individuals.
Specifically, decide which of the two externality themes
   below the speech excerpt relies the most on and output the
   corresponding number.
If neither theme clearly applies, choose NEITHER.

Definitions:

TOP  = Top-income externalities (harms that arise because
   wealth, income, or power is too concentrated at the top:
   political capture, erosion of democracy, undue corporate
   influence, distorted growth, social resentment, corruption,
    and so on).
BOT  = Bottom-income externalities (harms that arise because
   poverty and deprivation at the bottom weaken society:
   higher crime, worse public health, lower productivity,
   fiscal burdens, lower human capital, the societal costs of
   poverty, and so on).
Return exactly one of these three numbers, with NO additional
    text or punctuation:

NEITHER=0
TOP=1
BOT=2

Do not infer based solely on tone or rhetorical form. Focus
   on the substance or strongly implied themes of the argument
   .
If both externality themes appear, pick the single most
   dominant one. If no clear externality theme appears, choose
    NEITHER.
Do NOT explain the choice.  Just output one token: 0, 1, or
   2.
```

C Cross-country differences: Further data and analysis

C.1 Different externality and fairness specifications

We have three main classifications of externality- and fairness-based speech excerpts. Below we outline these:

Main specification: We base our main specification (discussed in the main body) as such:

Fair. (Only) Classified as including a fairness argument, but not an externality argument, in separate classification exercises (prompts in Appendices B.2-B.3). Our main specification to explore differences between fairness and externality arguments.

Ext. (Only) Classified as including an externality argument, but not a fairness argument, in separate classification exercises (prompts in Appendices B.2-B.3). Our main specification to explore differences between fairness and externality arguments.

Related to this specification we also define *Neither* – excerpts defined as neither type – and *Both*, excerpts defined as both types. The content of each of the four types in the U.S. Congress is shown in Figure F8.

Alternative specification (All): Relatedly, we define an alternate inclusive specification that does not exclude arguments classified as both fairness- and externality-based:

Fair. (All) Classified as including a fairness argument (prompt in Appendix B.2). Our main specification to explore the amount of fairness arguments.

Ext. (All) Classified as including an externality argument (prompt in Appendix B.3). Our main specification to explore the amount of externality arguments.

This definition is based on exactly the same prompt as Fair. (Only) and Ext. (Only).

Alternative specification (Primarily): We use the prompt in Appendix B.15 to classify each pro-redistributive argument as either:

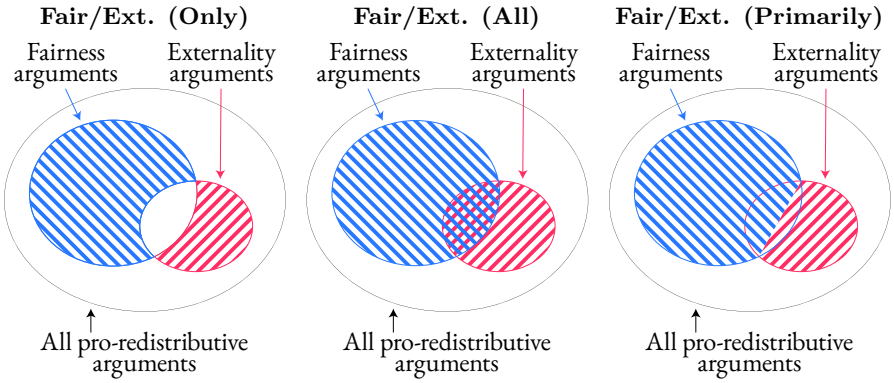
Fair. (Primarily) Classified as being primarily a fairness argument.

Ext. (Primarily) Classified as being primarily an externality argument.

Related to this classification we also classify as *Other* any excerpt which does not make either type of argument explicitly.

The Primarily classification closely aligns with the *Fair. (Only)* and *Ext. (Only)* classifications while encompassing those from the *Both* broader set of arguments. 95.7% of excerpts classified as *Fair. (Primarily)* are also classified as *Fair. (All)*, for example, while 87.6% of excerpts classified as *Ext. (Primarily)* are also classified as *Ext. (All)*. Of arguments classified

Figure C1: Comparing Universes of Arguments



Note. This figure graphically represents the four definitions of fairness and externality arguments used in the paper. Fair/Ext. (Only) corresponds to the main definition, which excludes arguments tagged as both fairness- and externatlity-based. Fair/Ext. (All) keeps those arguments in. Fair/Ext. (Primarily) is based on a single prompt determining whether the argument is mostly fairness- or externality-based. Appendix H show the full question text.

Table C1: Externality arguments in Norway vs. U.S., by specification

Numerator	Denominator	NO/US ratio
Ext. (All)	Fair. (All)	2.3
Ext. (All)	All pro-redistributive speeches	2.9
Ext. (All)	All speeches	8.7
Ext. (Only)	Fair. (Only)	2.3
Ext. (Only)	All pro-redistributive speeches	1.8
Ext. (Only)	All speeches	5.3
Ext. (Primarily)	Fair. (Primarily)	3.2
Ext. (Primarily)	All pro-redistributive speeches	2.9
Ext. (Primarily)	All speeches	8.5

Note. This table shows the relative use of externality-based argument in Norway compared to the U.S., varying the definition of externality-based argument use. In each row, the NO/US ratio is computed as $(\text{num}_{NO}/\text{denom}_{NO})/(\text{num}_{US}/\text{denom}_{US})$. The definitions of Fair./Ext. correspond to the definitions given at the top of Appendix C.1. "All pro-redistributive speeches" corresponds to the number of speech excerpts identified as containing a pro-redistributive argument in the U.S. Congress and Norwegian *Storting* data. All speeches corresponds to the total number of excerpts in both datasets.

as both *Fair. (All)* and *Ext. (All)*, i.e. those we describe as *Both* in the main specification, 54.6% are classified as *Fair. (Primarily)* and 37.6% are classified as *Ext. (Primarily)*, while only 7.8% are classified as *Other*.

There is again substantial spillover, however, as all excerpts are classified as either fairness- or externality-based by design. This is particularly notable for externality-based excerpts because the large majority of excerpts with an externality argument also contain a fairness argument (the motivation for using an exclusive classification scheme in our main specification). As such, only 17.8% of the *Ext. (Primarily)* excerpts are also *Ext. (Only)* excerpts (as specified by the main classification). Figure C1 provides a graphical representation of the definitions.

Ratio of externality arguments in the U.S. and Norway We also explore the ratio of externality arguments in the U.S. to Norway across different specifications. Our main specification compares the ratio of *Ext. (All)* to *Fair. (All)* across countries. Both the numerator and denominator could be changed, however, e.g. by using *Ext. (Only)* in the numerator, or all redistributive speeches in the denominator. As we show in Table C1, the amount of externality arguments is larger in Norway under any reasonable specification.

Varying the prompt text in other ways generally has a minimal effect on our main results. This is despite a changing aggregate share of the more complex classifications (e.g. which percentage of excerpts are classified as including fairness arguments) when the prompt is changed in larger ways. We interpret this as changes in the prompt largely changing the classification of excerpts that are on the border of being classified under either method; these excerpts presumably have similar content to those that *are* classified.

C.2 Applying counter-factual shares of fairness and externality arguments

This section examines how differences in the relative frequency of fairness- and externality-based arguments contribute to cross-country variation in the overall content of redistributive debates. Specifically, we construct counterfactuals that reweight each country's speeches using the other country's observed mix of fairness and externality arguments. The exercise illustrates what the Norwegian redistributive debate would look like if it had the U.S. mix of argument types—and conversely what the U.S. debate would look like if it had Norway's mix, assuming all other differences remain the same (e.g. the content of each U.S. excerpt type remain exactly the same).

Methodology. For each country, we compute the within-type means of every content characteristic (emotional and logical appeals, anger, compassion, and so on) separately for fairness- and externality-based excerpts. We then reweight these means by the other country's observed type shares to create counterfactual averages. Two sets of specifications are reported:

1. ***Fair. (Only)* and *Ext. (Only)*** This exclusive specification includes only excerpts classified as containing a fairness argument *without* an externality argument (*Fair. (Only)*) or an externality argument *without* a fairness argument (*Ext. (Only)*). Excerpts coded as *Neither* or *Both* are excluded.[8]

2. ***Fair. (Primarily)* and *Ext. (Primarily)*** This specification uses the binary classification that labels each redistributive excerpt as primarily fairness-based or primarily externality-based (*Fair. (Primarily)* and *Ext. (Primarily)*). Arguments coded as *Other* are excluded.

For each specification, we compute two counterfactuals:

- **US with NO mix:** applying Norway's type shares to the U.S. within-type means;
- **NO with US mix:** applying the U.S. type shares to Norway's within-type means.

The percentage gap explained is defined as

$$\% \text{ gap explained} = \frac{(US - NO) - (US_{cf} - NO_{cf})}{(US - NO)} \times 100,$$

where US and NO denote the observed means for the U.S. and Norway, and US_{cf} and NO_{cf} denote the respective counterfactual means obtained by reweighting the type shares.

[8]The exclusion of *Neither* avoids introducing a heterogeneous residual group of unclear interpretation, while excluding *Both* avoids averaging excerpts that combine fairness and externality content in varying proportions.

Table C2: Counter-factuals in Norway vs. U.S.: Using Fair. (Only) and Ext. (Only)

	US actual	US with NO mix	NO actual	NO with US mix	% gap explained
Emotional appeals	0.838	0.831	0.575	0.585	6.8
Logical appeals	0.401	0.418	0.609	0.595	14.5
Anger	0.470	0.460	0.378	0.391	24.8
Compassion	0.597	0.592	0.474	0.479	8.9
Fear	0.405	0.407	0.374	0.374	-8.5
Pessimism	0.574	0.578	0.588	0.591	-0.9
Empirical evidence	0.291	0.298	0.171	0.168	-7.8
Narrative evidence	0.323	0.316	0.217	0.225	13.6
Consensus-seeking	0.706	0.719	0.557	0.537	-22.1
Villainizing	0.362	0.355	0.306	0.312	22.3
Victimizing	0.839	0.831	0.648	0.664	12.3

Note. This table shows counterfactual shares of each content characteristic, swapping country's shares of externality arguments. US actual and NO actual show the share of each characteristic in the existing data. US with NO mix applies the share of externality arguments found in the Norwegian *Storting* (4.0%) to the U.S. Congress. NO with US mix applies the share of externality arguments found in the U.S. Congress (2.1%) to the Norwegian *Storting*. Column "% gap explained" shows the percentage reduction in the gap between the share of excerpts containing that characteristic in the U.S. Congress data, compared to the and Norwegian *Storting* data, as detailed in Section C.2 supra. Results are based on speeches containing only fairness or only externality arguments in each dataset.

Results. Table C2 presents the results using the "only" specification, and Table C3 shows the results using the "primarily" specification. Across both, the counterfactuals largely make the debates look mre similar. Using the method above, the counterfactuals decrease most content gaps by approximately 10–30%, suggesting that part of the observed cross-country difference reflects the relative composition of fairness- and externality-based arguments.

There is significant noise across varying types of content and prompts, however, and as discussed in the main text, these exercises should be interpreted with caution. The estimates are sensitive to methodological choices, as the differences in Tables C2-C3 show, and the broader cross-country content patterns are also largely detected in non-redistributive speech (Figure C2). This raises a question of control groups; one may imagine trying to explain only the content differences that are unique to the redistributive debate. This is challenging, however, as it is unclear which control groups to use for this purpose. Using all other speeches implies a control group where topic choice varies significantly across topics. Using a specific other speech topic (e.g. discussion on infrastructure spending) raises the issue that this other topic may itself be discussed differently across countries.

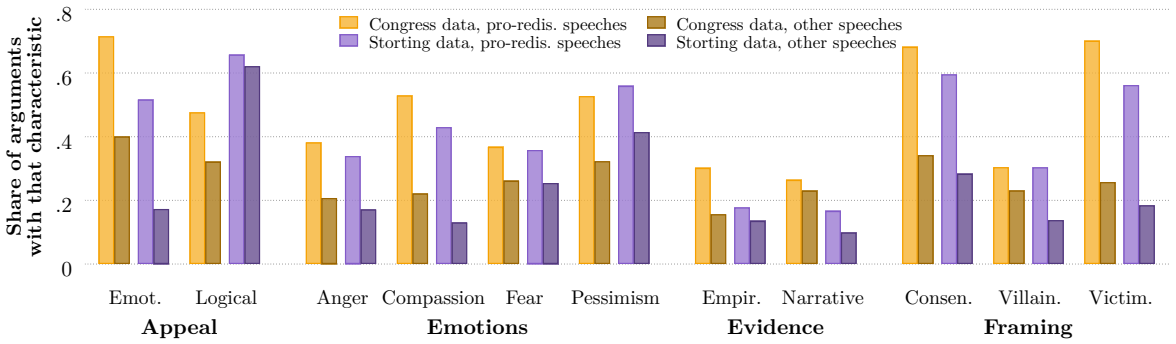
Overall, the counterfactuals confirm that differences in argument-type composition contribute to the observed divergence between Norwegian and U.S. redistributive debates, but they do not account for the majority of it—which is unsurprising, given that legislative speech across the countries differ across other topics as well.

Table C3: Counter-factuals in Norway vs. U.S.: Using Fair. (Primarily) and Ext. (Primarily)

	US actual	US with NO mix	NO actual	NO with US mix	% gap explained
Emotional appeals	0.850	0.828	0.664	0.690	13.0
Logical appeals	0.442	0.478	0.632	0.609	15.6
Anger	0.495	0.493	0.442	0.455	14.1
Compassion	0.578	0.542	0.482	0.513	34.8
Fear	0.393	0.408	0.425	0.420	30.4
Pessimism	0.600	0.628	0.639	0.633	43.3
Empirical evidence	0.308	0.318	0.195	0.188	-7.8
Narrative evidence	0.298	0.274	0.207	0.233	27.2
Consensus-seeking	0.729	0.733	0.685	0.664	-29.7
Villainizing	0.390	0.398	0.408	0.406	26.8
Victimizing	0.856	0.842	0.702	0.729	13.3

Note. This table shows counterfactual shares of each content characteristic, swapping country's shares of externality arguments. Results are based on the "primarily" classification defined in Section C.1. US actual and NO actual show the share of each characteristic in the existing data. US with NO mix applies the share of externality arguments found in the Norwegian *Storting* (4.0%) to the U.S. Congress. NO with US mix applies the share of externality arguments found in the U.S. Congress (2.1%) to the Norwegian *Storting*. Column "% gap explained" shows the percentage reduction in the gap between the share of excerpts containing that characteristic in the U.S. Congress data, compared to the and Norwegian *Storting* data, as detailed in Section C.2 supra.

Figure C2: Average argument characteristic in the U.S. Congress and Norwegian *Storting*, by speech theme



Note. This figure displays, for each content characteristic, the share of redistributive excerpts classified as containing this characteristic in the U.S. Congress and Norwegian *Storting* sample, contrasting the content of pro-redistributive speeches (our focus throughout this paper) with all other speeches in the data. Appendix B shows the full prompt texts.

C.3 Types of fairness and inequality externality arguments

Here we discuss the classification of different types of fairness and inequality externality arguments.

Theoretical framing We further divide fairness arguments into three subtypes. Appeals to curvature ($\partial U_i/\partial x_i$) are *needs-based*: they invoke poverty relief, ability to pay, or the idea that some individuals benefit more from an additional dollar than others. Appeals that apply equally to all individuals are *rights-based*: they rest on equal moral worth, equal opportunity, or human rights. Finally, following Scheve and Stasavage (2016), appeals that apply unequally to different individuals depending on their characteristics are defined as *compensatory*; such arguments typically invoke deservingness or corrective justice for groups disadvantaged by discrimination, exploitation, or historical shocks.

We further divide externality arguments into two subtypes. *Top-income arguments* focus on inequality near the top, for example asserting that excessive wealth concentration undermines

democracy, fosters regulatory capture, or erodes social cohesion. *Bottom-income arguments* focus on either absolute or relative poverty, for example arguing that deprivation increases crime rates or diminishes human capital and economic growth. Both types of arguments claim that changing the distribution θ improves aggregate outcomes, and thus justify redistribution through consequentialist society-wide welfare optimization that affects not only those who receive any additional transfers.

Methodology: Types of fairness arguments With a single prompt we classify each fairness argument into one of those subtypes (prompt in Appendix B.22):

- **Needs-based fairness:** If the excerpt appeals, broadly, to the concept of those at the bottom needing more. Topics include basic needs, ability-to-pay, poverty relief, or the idea that some people need support more than others.
- **Rights-based or egalitarian fairness:** If the excerpt appeals, broadly, to the concept that all people should be equal. Topics include equal rights, equal moral worth, strict egalitarianism, or fairness of opportunity.
- **Merit-based or compensatory fairness:** If the excerpt appeals, broadly, to the concept that those at the bottom deserve more due to previous sacrifice or hard work. Topics include deservingness, effort, contribution, earned rewards, or compensation for harm or sacrifice.
- **Undefined.** If the excerpt appeals to none of the above.

Methodology: Types of inequality externality arguments With a single prompt we classify each externality argument into one of those subtypes (prompt in Appendix B.23):

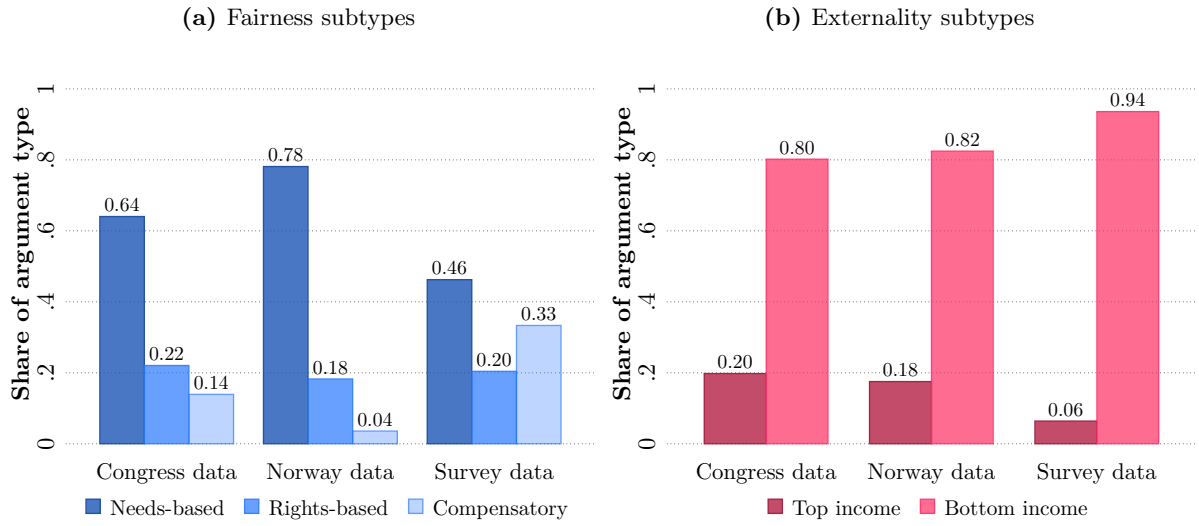
- **Top-income externalities:** If the excerpt appeals, broadly, to externalities that arise because wealth, income, or power is too concentrated at the top. Topics include political capture, social resentment, corruption, and more.
- **Bottom-income externalities:** If the excerpt appeals, broadly, to externalities that arise because poverty and deprivation at the bottom weaken society. Topics include crime, poor public health, low human capital, and more.
- **Undefined:** If the excerpt appeals to none of the above.

Results We show the resulting classifications in Figure C3. In either case, almost no excerpts are undefined – approximately 1% of either type.

In the U.S. Congress, across fairness excerpts, 61% are classified as needs-based, 25% are classified as rights-based, and 13% are compensatory (merit-based). Across inequality externality arguments, 18% are classified as top income-based and 81% are classified as bottom income-based.

In the Norwegian *Storting*, needs-based fairness arguments are slightly over-represented as compared to the U.S. Congress; otherwise the data looks similar. In the experimental data, rights-based fairness arguments and bottom-income externalities are over-represented by approximately 20 p.p. each as compared to the U.S. Congress.

Figure C3: Distribution of argument types



Note. This figure shows the distribution of the three fairness subtypes and the two externality subtypes in all three settings (the U.S. Congress, the Norwegian *Storting*, and the experimental data). Fairness arguments are split between needs-based, right-based, and compensatory arguments. Externality arguments are split between top-externality and bottom-externality arguments. Subtypes are defined in section C.3 supra. The prompts used for this classification are in Appendices B.23-B.22. Standard errors are robust, 95% CIs.

Figure C4 compares content in the U.S. Congress across these subtypes. The three fairness types are broadly similar, though needs-based arguments exhibit somewhat stronger emotional elements—especially anger, compassion, and villainizing language. The two externality types differ more sharply: most of our results are driven by bottom-income externality arguments, while top-income externality arguments—though relatively rare—resemble fairness arguments in tone, displaying higher anger and villainization, lower consensus-seeking, and little compassion.

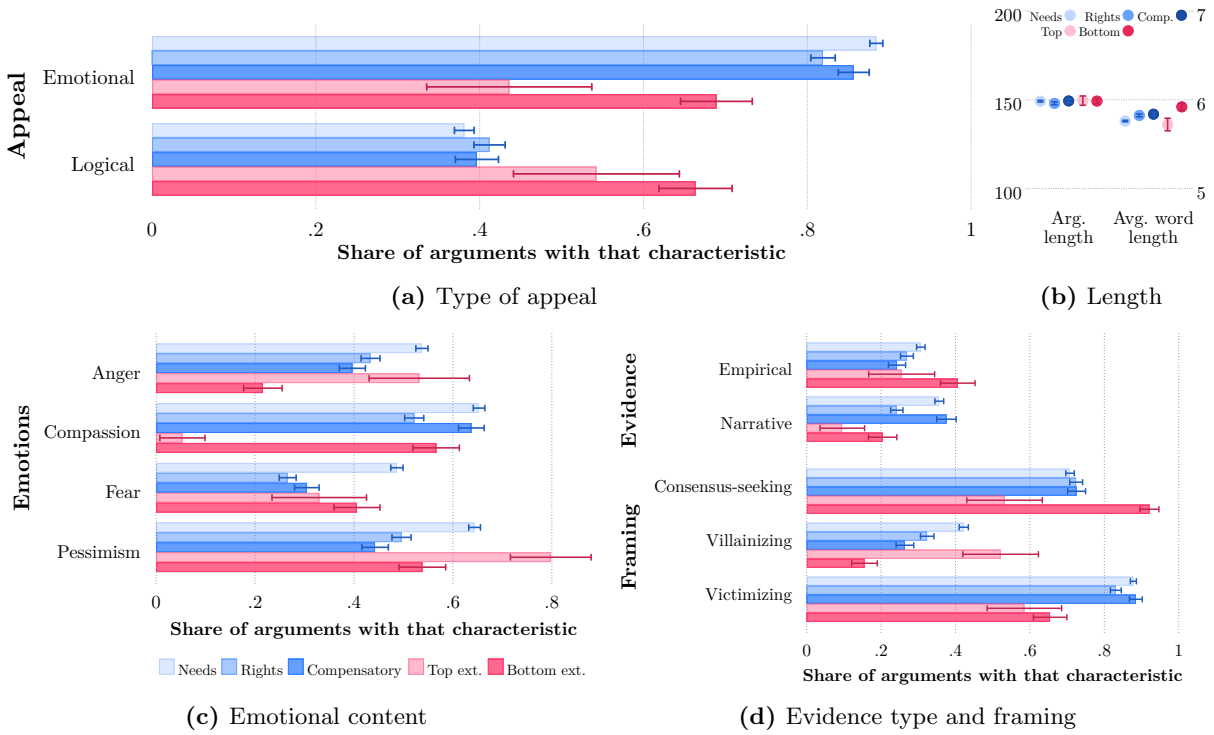
This distinction should be kept in mind when interpreting the main results. As top-income externality arguments are uncommon and stylistically closer to fairness arguments, our aggregate contrast between fairness and externality reasoning primarily reflects the properties of bottom-income externality arguments. We retain both subtypes in the main classification for consistency but note that this likely makes our reported differences conservative.

C.4 Words used in fairness and externality arguments

We explore the textual content of arguments in all datasets to identify the words most characteristic of *Fair. (Only)* and *Ext. (Only)* arguments. We start by preprocessing the text by removing digits, special characters, and stopwords (common words that carry little meaning, such as 'and', 'is', 'they'). Words are then lemmatized, mapping all derivatives of the same root (equality, equalize, equals) to the same word (equality). We set a word's frequency = 1 if the word appears in an argument, regardless of the number of times it appears, and only keep words that appear in at least 10 different arguments.

For each word, we then compute log odds ratio as $r_j = \log(n_j^F/(N^F - n_j^F)) - \log(n_j^E/(N^E - n_j^E))$, where n_j^I is the number of arguments containing word j in argument category I, and N^I is the total number of unique words used across all arguments in category I, counting one occurrence per word per argument. Positive values of r_j denotes a word that is used relatively

Figure C4: U.S. Congressional Speeches: Content of excerpts across excerpt subtype



(a) Type of appeal

(b) Length

(c) Emotional content

(d) Evidence type and framing

Note. This figure complements Figure 2, by showing content differences between fairness and externality argument subtypes in the U.S. Congress data. Fairness arguments are split between needs-based, right-based, and compensatory. Externality arguments are split between top-externality and bottom-externality. Subtypes are defined in section C.3 supra. The prompts used for this classification are in Appendices B.23-B.22. Standard errors are robust, 95% CIs.

more frequently in fairness arguments than in externality arguments, while negative values denote the opposite. To avoid the ratio being undefined for words that are exclusive to one category, ($n_j^I = 0$), we regularise it by adding a small pseudo-count to each word count. We fix the prior weight to .001, so the pseudo-count for word j is $\alpha = .001 * N/W$, where N is the total number of word occurrences, and W is the number of unique words in the corpus. The regularised log odds ratio is given by

$$r'_j = \log\left(\frac{n_j^F + \alpha}{N^F - n_j^F + (W-1)\alpha}\right) - \log\left(\frac{n_j^E + \alpha}{N^E - n_j^E + (W-1)\alpha}\right).$$

Finally, we compute the ratio's variance as $\sigma_j^2 = 1/(n_j^F + \alpha) + 1/(n_j^E + \alpha)$, and obtain the log odds ratio's z-score by $z_j = r'_j/\sigma_j$.

Table C4: Words most characteristic of each argument type in Congress data

(a) Fairness arguments

	Frequency in: Fairness arguments	Frequency in: Externality arguments	Log odds ratio z-score
fair	1,635	10	6.63
pay	3,315	94	5.32
woman	1,607	36	4.62
day	1,458	34	4.24
deserve	898	0	4.13
hard	909	15	4.08
equality	819	0	4.02
man	906	16	3.95
paid	1,121	25	3.82
vote	1,127	27	3.60
fight	914	19	3.59
hardwork	560	0	3.56
earn	1,011	23	3.55
paycheck	657	10	3.54
amendment	699	12	3.47
drug	517	0	3.46
month	800	16	3.43
justice	482	0	3.38
wrong	473	0	3.36
trump	677	12	3.35

(b) Externality arguments

	Frequency in: Fairness arguments	Frequency in: Externality arguments	Log odds ratio z-score
economy	1,834	273	-17.06
growth	575	124	-14.99
inequality	264	63	-11.41
investment	780	114	-10.92
community	1,294	150	-9.96
economist	26	18	-8.96
outcome	56	22	-8.58
trickle	52	21	-8.48
income	1,046	117	-8.47
global	74	24	-8.34
society	253	44	-7.93
spend	477	65	-7.84
productivity	135	30	-7.74
prosperity	119	28	-7.74
boost	105	26	-7.68
research	181	35	-7.65
cycle	45	16	-7.19
stronger	100	23	-7.00
infrastructure	310	45	-6.97
mobility	39	14	-6.81

Notes: This table complements Section 5.5, showing the 20 words most characteristic of *Fair. (Only)* and *Ext. (Only)* arguments in the U.S. Congress data, according to their log odds ratio z-score. The methodology is detailed in Section C.4. The log odds ratio z-score measures the significance of the relative overrepresentation of that word among fairness (resp. externality) arguments.

Table C5: Words most characteristic of each argument type in experiment data

(a) Fairness arguments

	Frequency in: Fairness arguments	Frequency in: Externality arguments	Log odds ratio z-score
rettferdig	688	0	4.46
fått	790	47	4.27
forslag	910	58	4.25
vanlige	369	13	4.06
kutt	367	14	3.92
rødt	419	19	3.83
pensjon	290	0	3.56
ordningen	289	0	3.55
statsråden	552	33	3.52
ordning	237	0	3.35
uføre	279	12	3.17
pensjonister	198	0	3.16
arbeidsfolk	193	0	3.13
forslaget	415	24	3.13
saken	368	20	3.11
lønn	273	12	3.10
kuttet	184	0	3.08
betale	445	27	3.08
sosialhjelp	182	0	3.07
urettferdig	181	0	3.07

(b) Externality arguments

	Frequency in: Fairness arguments	Frequency in: Externality arguments	Log odds ratio z-score
forskjeller	175	147	-17.90
land	243	116	-12.75
forskjellene	164	95	-12.68
små	170	94	-12.36
samfunn	150	88	-12.27
ulikhet	50	52	-11.25
tillit	36	47	-11.11
norge	1,093	265	-11.05
lykkes	53	48	-10.50
vekst	47	44	-10.14
samfunnet	389	124	-10.03
helse	170	73	-9.57
utvikling	116	59	-9.45
gode	580	154	-9.37
verdiskaping	39	36	-9.17
sosiale	187	74	-9.15
fellesskap	105	53	-8.94
verden	125	57	-8.78
skape	196	73	-8.72
skole	125	56	-8.62

Notes: This table complements Section 5.5, showing the 20 words most characteristic of *Fair. (Only)* and *Ext. (Only)* arguments in the Norweigian *Storting* data, according to their log odds ratio z-score. The methodology is detailed in Section C.4. The log odds ratio z-score measures the significance of the relative overrepresentation of that word among fairness (resp. externality) arguments.

Table C6: Words most characteristic of each argument type in experiment data

(a) Fairness arguments

	Frequency in: Fairness arguments	Frequency in: Externality arguments	Log odds ratio z-score
work	34	9	3.68
money	34	12	3.24
hard	20	5	2.85
fair	36	0	2.12
redistribution	49	32	2.07
earn	13	5	1.87
wealth	27	16	1.78
unfair	15	0	1.78
job	15	0	1.78
worker	14	0	1.75
only	14	0	1.75
way	14	0	1.75
pay	13	0	1.72
tax	11	0	1.66
most	11	0	1.66
take	11	0	1.66
rich	13	6	1.63
everyone	17	9	1.63
ceo	10	0	1.62
very	9	0	1.58

(b) Externality arguments

	Frequency in: Fairness arguments	Frequency in: Externality arguments	Log odds ratio z-score
inequality	6	50	-4.93
economy	24	62	-3.95
society	16	42	-3.25
crime	0	40	-2.14
increase	5	14	-1.92
lead	0	21	-1.89
better	6	15	-1.84
social	0	15	-1.75
create	0	15	-1.75
negative	0	13	-1.70
rate	0	12	-1.67
health	0	11	-1.63
reduction	0	11	-1.63
lower	7	15	-1.60
education	0	10	-1.60
issue	0	9	-1.56
consequenc	0	9	-1.56
affect	0	9	-1.56
into	0	8	-1.51
impact	0	8	-1.51

Notes: This table complements Section 5.5, showing the 20 words most characteristic of *Fair. (Only)* and *Ext. (Only)* arguments in the experiment data, according to their log odds ratio z-score. The methodology is detailed in Section C.4. The log odds ratio z-score measures the significance of the relative overrepresentation of that word among fairness (resp. externality) arguments.

D Survey Experiment: Methodological details

D.1 Survey 1 (Elicitation)

Survey 1 was conducted between November 16 and November 27 2022 with a total sample of $N_1 = 298$. 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 as compared to the pre-specification. The median survey time was 6 minutes and 43 seconds.

Respondents were informed before agreeing to participate in 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 demographic questions, 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.

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). The question text is shown in Figure D5.

Figure D5: Choice of argument written: Question text

In this survey we want you to tell us arguments for or against the concept of reducing economic differences in society. We plan to show your arguments to other survey respondents, so please be clear and write as if you are writing to another person.
If your argument is chosen (which most will be) and found convincing by a majority of other respondents, your payment for this survey will be automatically doubled.
The arguments you write, if chosen, will be shown completely anonymously. If you do not consent to this, please close and return the survey.

[Before fairness argument elicitation]
Which of these do you prefer to make an argument for?

- Why we should redistribute **less**, as the market distribution of resources is **fair**
- Why we should redistribute **more**, as the market distribution of resources is **unfair**

[Before inequality externality argument elicitation]
Which of these do you prefer to make an argument for?

- Why we should redistribute **less**, as inequality changes society for the **better** (more inequality→ a better society in various ways
- Why we should redistribute **more**, as inequality changes society for the **worse** (more inequality→ a worse society in various ways)

Note. This figure shows the question text for the elicitation of arguments in Survey 1. Differences in text across the elicitation of fairness and inequality externality arguments are denoted by [brackets]. Differences in text across pro- and anti-redistributive argument elicitation are denoted by {brackets}. See Figure 7 for the argument elicitation text.

Respondents were required to answer both the fairness and externality prompts, which led to a total of 596 arguments, 440 of which were pro-redistribution.

D.2 Survey 2 (Quality check)

Survey 2 was conducted between November 26th and November 27th 2022 with a total sample of $N_2 = 215$. Survey-takers were asked to assess whether arguments from Survey 1 were coherent

and relevant to the intended topic. The aim of the survey was to rank the arguments on these neutral dimensions, and then use the arguments that are above our pre-specified cut-offs for an evalution in Survey 3. The quality check reduced the number of pro-redistribution arguments from 440 to a pre-specified 160 arguments – 80 fairness arguments and 80 inequality externality arguments. It reduces the number of anti-redistributive arguments from 156 to 30.

We enlisted *Prolific* respondents for this task. 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 the survey, 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 H.2. Each of the 596 arguments was evaluated an average of 5.5 times (between 4 and 8 evaluations), and each respondent in Survey 2 (215 in total) evaluated 16 arguments.

In recent literature (e.g. Andre et al., 2023) this type of task is often performed by research assistants who are not informed of the hypotheses to be tested. Although there are clear benefits to this approach – notably having classifications done by skilled individuals who become familiar with the approach through repeated exposure – there are also drawbacks. Most importantly, research assistants may have knowledge of the hypotheses to be tested (potentially inferred through the research interests of their employer, or through the repeated exposure to hundreds or thousands of classifications) and personal incentives for the research to be successful. This may in turn bias the classifications towards whatever hypothesis they believe is being tested. Classification also becomes fragile to the opinions of one or a small number of individuals.

Our method does not have these shortcomings, as we 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 arguments each, (iii) are not aware of the research portfolio of the responsible academics. While this method also has limitations – particularly less control over evaluator training and consistency – it reduces the risk of systematic bias stemming from alignment with researcher expectations, which is particularly crucial in our design.

D.2.1 Choice of arguments to keep after Survey 2

We pre-specified that the final sample would be 200 arguments, 160 of which would be pro-redistribution, and based our selection of these arguments on two pre-specified criteria.

1. Arguments needed to be evaluated by 75% or more of respondents as making sense and being on the correct overall topic.

2. 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.

This pre-specification was followed closely for the pro-redistributive arguments. A slight modification of Criteria #1 was made for the externality arguments, where we decreased the pass threshold to 71% for two arguments. A larger modification of Criteria #2 was made for both types, where we decreased the pass threshold to 60% for both types of arguments. If more arguments than required reached the threshold, the lowest-performing arguments were removed. In the case of a tie on the relevant metrics we randomized which argument to keep.

A larger modification was made for the anti-redistributive arguments, which we discuss below.

D.2.2 Anti-redistributive arguments

Although our primary focus was pro-redistributive arguments – following how roughly 70% of U.S. citizens believe the extent of economic inequality in their country is unfair, and also believe that inequality overall has negative consequences (Lobeck and Støstad, 2023) – we initially intended to also do exploratory analysis on anti-redistributive arguments. In practice, two issues prevented us from pursuing this analysis further.

First, the smaller sample of arguments (40 instead of 160) created significant power analysis for the final analysis. Second, respondents generally struggled to write anti-redistributive arguments about inequality externalities. Anti-redistribution inequality externality-based arguments were often classified as fairness arguments by Survey 2 respondents. As such, we only found 10 such arguments that reached our pre-specified criteria.[9] Given these limitations – both in the quality and quantity of anti-redistributive arguments – we chose not to include them in the main analysis.

After Survey 1 and Survey 2 we were then left with 160 pro-redistributive arguments written by survey respondents, 80 of which focus on fairness ideas and 80 of which focus on inequality externality ideas. To keep these samples as unbiased as possible, we gave minimal guidance on what kind of argument to write and allowed responses of up to three sentences. This design gives respondents substantial freedom, which increases noise and may reduce statistical power – the main limitation of our approach. At the same time, this same feature is its central strength, as it 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.

D.3 Survey 3 (Evaluation)

Survey 3 was conducted between December 8th and December 30th, with the large majority (89% of responses) collected before December 15th. 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

[9]To ensure that Survey 3 had 200 arguments, as originally intended, we included 10 pro-redistributive arguments attributed to five distinct sources – one argument of each type per source – for exploratory purposes. The arguments were made by Barack Obama, Nicholas Kristof, Bernie Sanders, Tucker Carlson, and ChatGPT (which was fed the same question as Survey 1 respondents). We note this difference as it was a change from the pre-analysis plan, but do not discuss it further as the explorative analysis was under-powered. 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. In the analysis in Section 4.5 we treat these arguments as pro-redistributive arguments of the specific type.

questions before eliciting pre-evaluation fairness views, externality beliefs, and redistributive preferences. Respondents were then told that they would be asked to evaluate ten different arguments on redistribution (one shown at a time). Respondents also evaluated a smaller set of the anti-redistributive arguments elicited in Survey 1 (see Appendix D.2.2), which creates variation of how many pro- and anti-redistributive arguments each respondent sees.

Each respondent was shown each argument in the following way:

Another survey respondent is trying to convince you with the following argument for ***more redistribution*** *of income and wealth:*

[Argument]

Please read the argument carefully and think about it for a few seconds.

On the same screen, the respondent was asked three questions per argument. Based on these questions we define our main outcome variables:

- **Convincingness:** Defined as 1 if the respondent answers "*Very convinced*" or "*Convinced*" to the question: "*Are you personally convinced by this argument or statement?*". Answer options were [*Very convinced/Convinced/Neither convinced nor unconvinced/Unconvinced/Very unconvinced*].
- **Outrage:** Defined as 1 if the respondent answers "*Yes, because I agree with the argument*" or "*Partly, because I agree with the argument*" to the question: "*Imagine someone said this to you in person. Do you think a discussion about this argument could provoke an emotional reaction like anger or agitation in you?*". Answer options were [*Yes, because I agree with the argument/Yes, because I think the argument is nonsense/Partly, because I agree with the argument/Partly, because I think the argument is nonsense/No, not really/No, not at all*].

After the survey we also elicited post-treatment externality beliefs, fairness views, and redistributive preferences.The collection of both pre- and post-evaluation redistributive preferences allows us to explore how the number of pro-redistributive arguments affects redistributive preferences, which we do in Section 4.5. We primarily see this exercise as a robustness test of the self-reported evaluations, where we have better power and more data.

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 32,680 evaluations. On average, each argument was viewed by 201.9 respondents.

Representativity 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 oversampled Democrats and to a lesser degree Independents. Both this sampling choice and the resulting self-selection into pro- or anti-redistributive arguments means that the arguments we analyze have been largely written by left-wing respondents. We also did not specifically quota on other demographic dimensions beyond gender, which we consider a potential limitation of the study. 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 opted not to 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 ensure a diverse and representative group of respondents for the argument evaluations.

D.4 Differences to pre-specification

When launching the survey experiment, our objective was purely to collect *evaluations* of different types of arguments. In the final paper we conduct various additional analyses that were not pre-specified, largely due to the proliferation of LLMs (the experiment was done before the initial launch of ChatGPT). This has changed the paper significantly, as the focus is now also on the *content* of arguments, and led to various additional analyses. Beyond these changes, our experimental analysis is largely very similar to the pre-specification.

- We conducted additional analyses which use LLMs, notably classifying the content of arguments and comparing experimental and legislative content.
- We examine post-treatment redistributive preferences as a robustness test (Section 4.5), using the exogenous variation of number of pro-redistributive arguments for identification.
- We do not use the pre-specified set of controls in our main outcome regressions for the evaluations (Figures 9-10), instead using individual-level fixed effects. This does not affect the main results and was done to improve precision.
- We excluded anti-redistributive arguments from the main analyses due to their limited number and low quality (10 arguments met inclusion criteria), focusing instead on the 160 pro-redistributive arguments (80 fairness, 80 externality).
- The inclusion thresholds for argument quality in Survey 2 were slightly relaxed (from 75% to 71% for sensibility and to 60% for topical correctness) to retain sufficient arguments.
- The final Survey 3 sample consisted of 4,010 respondents rather than the pre-registered 4,000, and a small set of exploratory arguments (10 total) from public figures and ChatGPT were added to the arguments shown to respondents. These are only used in Section 4.5, where they are counted as pro-redistributive arguments. Excluding these arguments from this analysis has a minimal effect on results.
- To streamline the analysis, we focus our attention on two of the pre-specified outcome variables (*convincingness* and anger because of agreement, or *outrage*). The two other pre-specified outcomes (anger because of *disagreement* and willingness to have a conversation with the author) are reported in Appendices E.1-E.2.
- We pre-specified that we would explore the net effect of argument type on any type of anger (due to agreement *or* disagreement). These results merge those discussed in the main section and those discussed below (Appendix E.1); as these two types of anger differ

substantially both empirically and conceptually, we do not present the evidence of the merged sample.

E Survey Experiment: Further data and analysis

We show examples for each content category (emotional appeal, anger, etc) in Appendix E.3.

E.1 Outrage due to disagreement

We asked respondents whether a discussion with the author of the argument could provoke an emotional reaction like anger or agitation in them due to their disagreement with the argument. We define this as *frustration*. We show the question text in Appendix H.3, and the results in Figure F14. There is no significant difference between fairness and externality arguments.

E.2 Having a longer conversation

We asked respondents whether they would be willing to have a longer conversation with the individual who wrote the argument to discuss these ideas. We show the question text in Appendix H.3, and the results in Figure F15. There is no significant difference between fairness and externality arguments.

E.3 Examples: Argument content

Here we show illustrative examples of each content classification from the arguments discussed in Section 4.4.

E.3.1 Logical and emotional appeals

Logical appeal, no emotional appeal: *"Redistribution of economic resources promotes equality and societal balance. This redistribution would reduce crime and social unrest while improving societal innovation and prosperity."*

Emotional appeal, no logical appeal: *"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."*

E.3.2 Emotions

Fear: *Economic inequality is a sickness of society and like all sickness, it can and will gradually deplete you, drain you of your resources and well being, your health, mental and physical and more. [It] can also be a large driver in crime, increasing a mental health crises, dragging down the economy, and creating many stressors in the community that have a gradually expanding force on society governments, and the culture of a nation and peoples.*

Anger: *The workers themselves should get raises. How many corperations where the CEOs are making millions meanwhile the people that do the actual grunt work to make that money arent even getting a living wage? How would you feel if you worked your fingers to the bone for pennies meanwhile some suit sits there, does basically nothing and just rakes in the cash?*

Pessimism: *With increasing economic inequality, people are less and less inclined to work hard, as it makes no financial sense to toil away for a wage that they can't even live on. If this keeps up, the system will collapse.*

Compassion: *Everyone deserves to be able to afford basic necessities such as food, shelter, and healthcare. Every employee should be paid a living wage that is adjusted as the cost of living either increases or decreases.*

E.3.3 Types of evidence

Empirical evidence, no narrative evidence: *We should support economic redistribution for a wide variety of reasons. Small amounts of inequality are inevitable in any society, but large amounts such as the ones we see today are a massive drain on our society. Areas with higher levels of income inequality have much higher crime rates and lower academic achievements.*

Narrative evidence, no empirical evidence: *As the wealth divide worsens, people will be forced to turn to crime in order to take care of basic needs. When a person is hungry and has no hope or means of eating they will almost always try to steal food to survive. Once people see that they cannot live by societies laws then they will have no reason to follow them.*

E.3.4 Types of framing

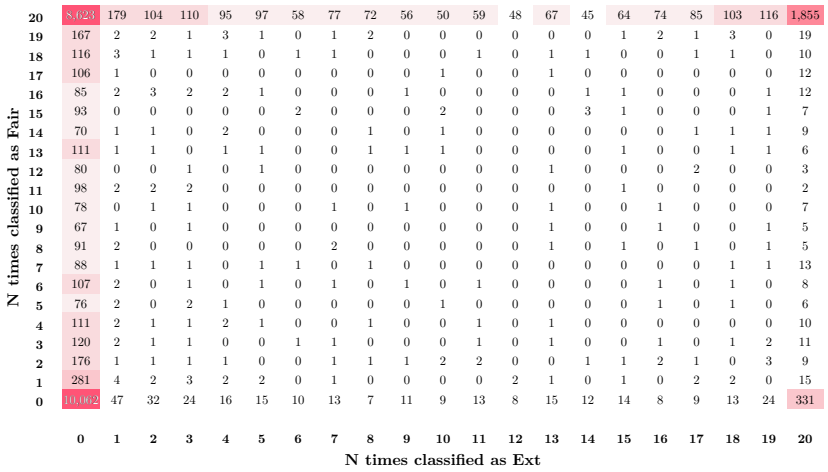
Consensus-seeking: *Economic inequality has negative consequences for society because it leads to more crime. This leads to costs which effect everyone, even those that weren't the victims. Both property and personal crimes are increased, so we all have both financial and safety reasons to want economic inequality to be reduced.*

Villainizing: *We live in a very unfair society with so much income inequality. Many families struggle to meet the most basic needs, while the likes of Jeff Bezos and Elon Musk amass billions. It is immoral that children starve while the wealthy have way more than they need.*

Victimizing: *Increased economic redistribution would be more fair to people who have grown up stuck in a cycle of poverty and inequality caused by forces outside of their control, such as racism or issues with addiction. Often, people living near or below the poverty line, face specific challenges obtaining goals many of us take for granted, like securing a job or safe housing. With this new distribution of wealth, people who are poverty stricken are given a chance to compete at the same level.*

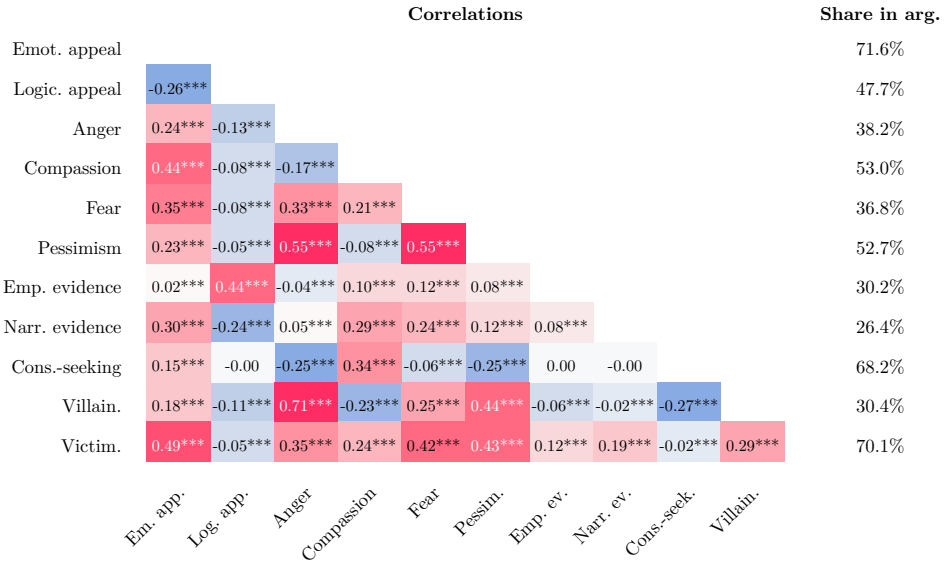
F Appendix Graphs

Figure F1: Consistency of classification as Fairness or Externality type



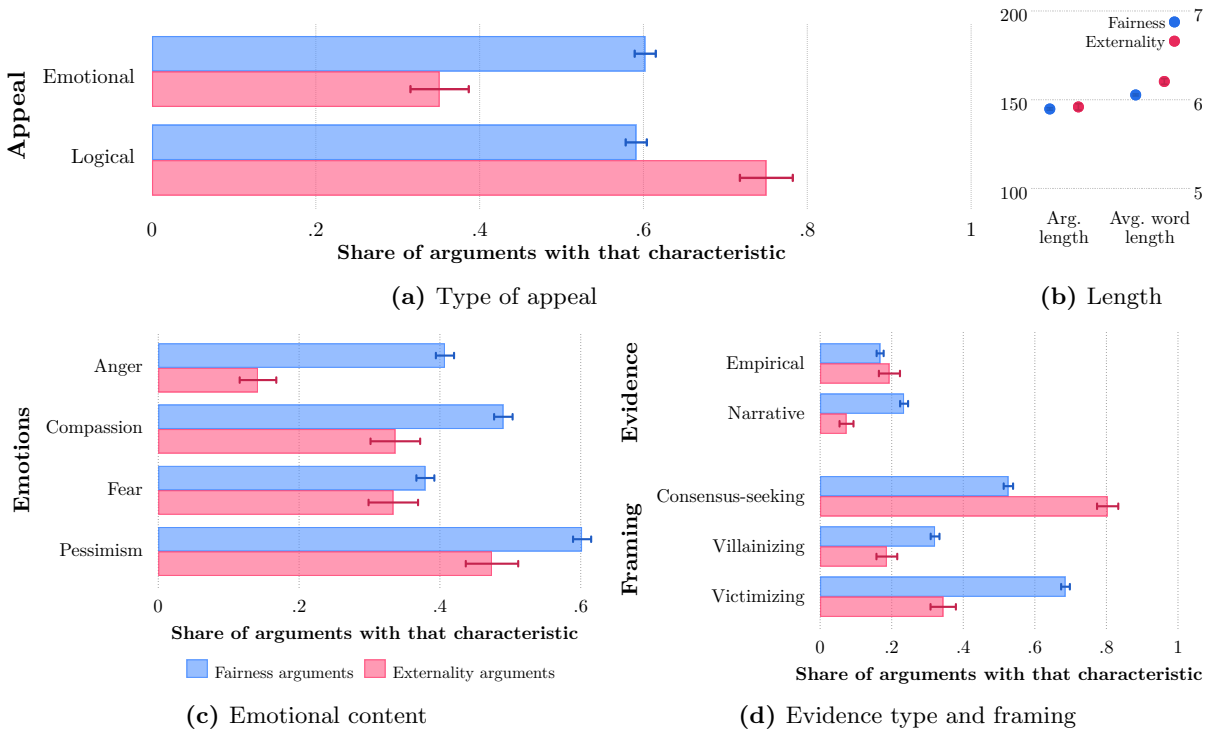
Note. This figure complements Section 3.1, reporting the results from 40 API calls, 20 for the fairness classification and 20 for the externality classification, in the U.S. Congressional data. An excerpt is classified as an externality or fairness argument if it is classified as such in at least 90% of classifications. Appendix B shows the full prompt texts.

Figure F2: U.S. Congressional Speeches: Correlation of content classifications



Note. This figure reports the correlation between all content characteristics in the U.S. Congressional data. The share of each type of content can be found in Figure 3. Appendix B shows the full prompt texts.

Figure F3: Norwegian *Storting* Speeches: Content of excerpts across excerpt type



(a) Type of appeal **(b)** Length

(c) Emotional content **(d)** Evidence type and framing

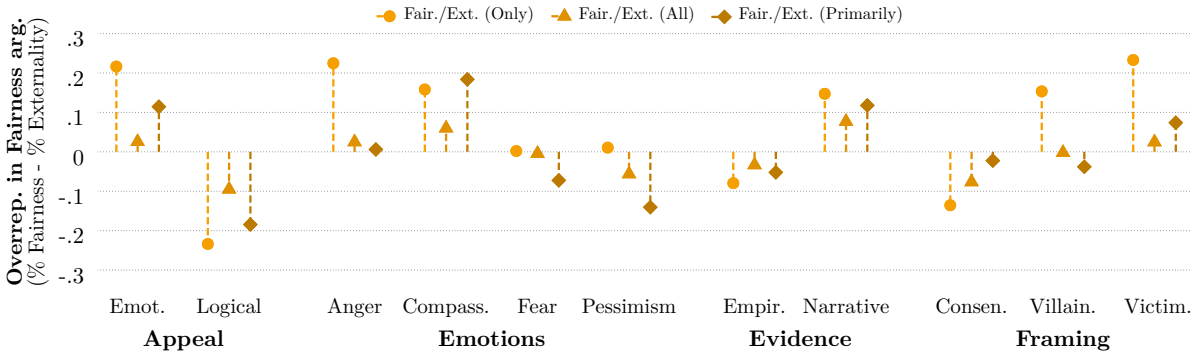
Note. This figure complements Figure 5, showing the content of fairness and externality-based excerpts in the Norwegian *Storting* data. (replicating Figure 2 with the *Storting* data). Results are based on speeches containing only fairness (N= 5,574) or only externality arguments (N=692). Appendix B shows the full prompt texts.

Figure F4: Content gaps between fairness and externality excerpts, with confidence intervals

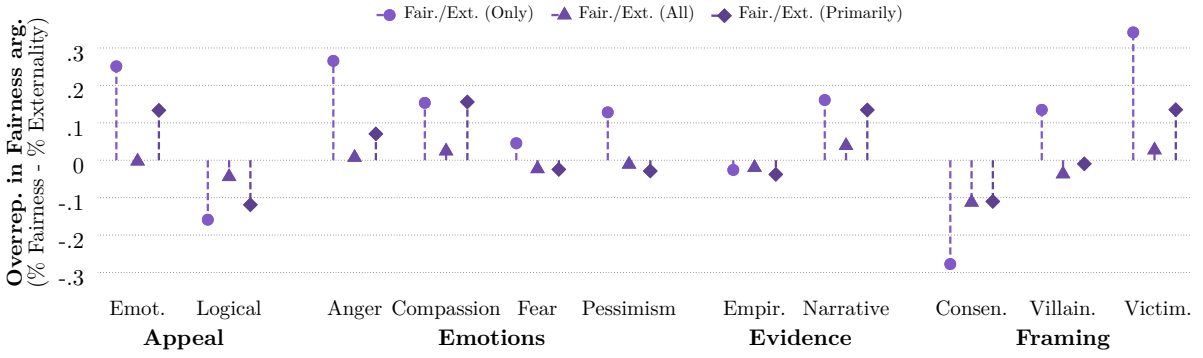
Note. This figure complements Figure 5 by adding 95% confidence intervals to the content differences between fairness and externality arguments across the two legislative settings (U.S. Congress and Norwegian *Storting*).

Figure F5: Fairness–externality differences across countries, comparing classification definitions

(a) U.S. Congress speeches

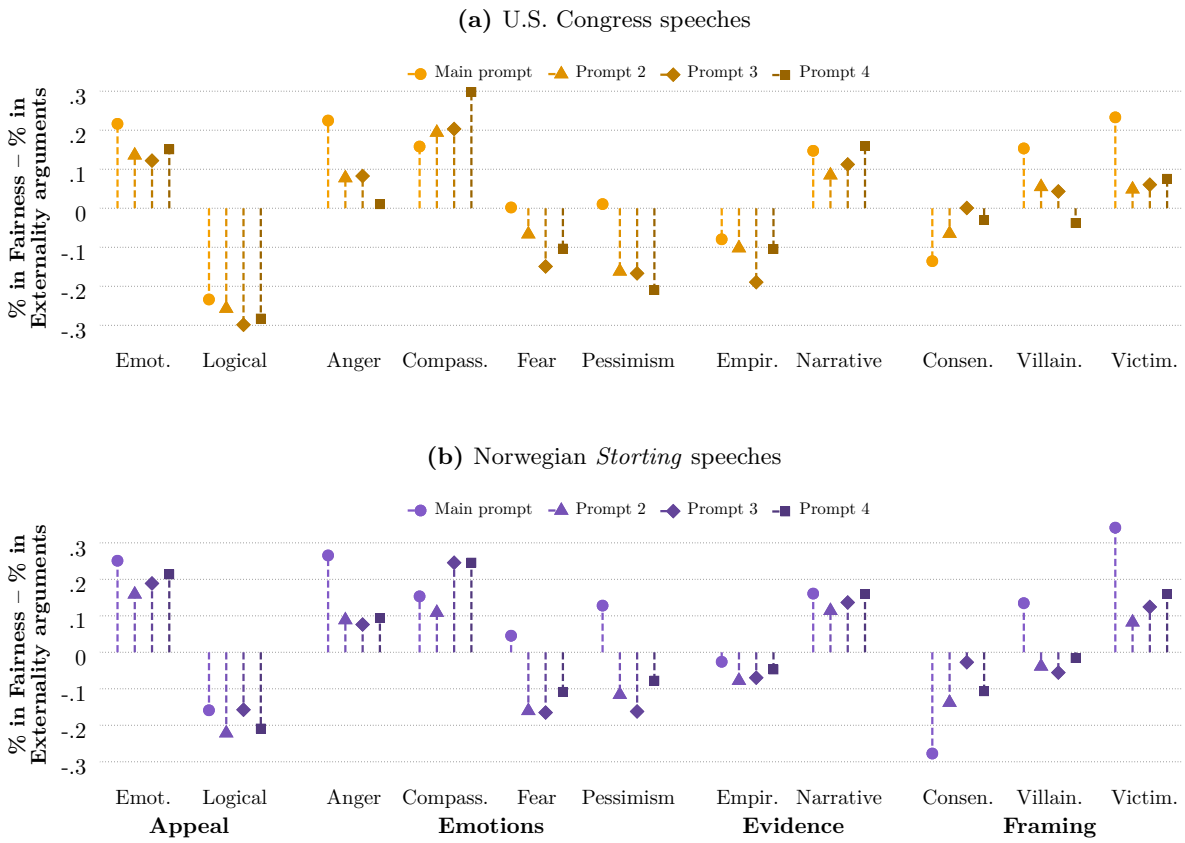


(b) Norwegian *Storting* speeches



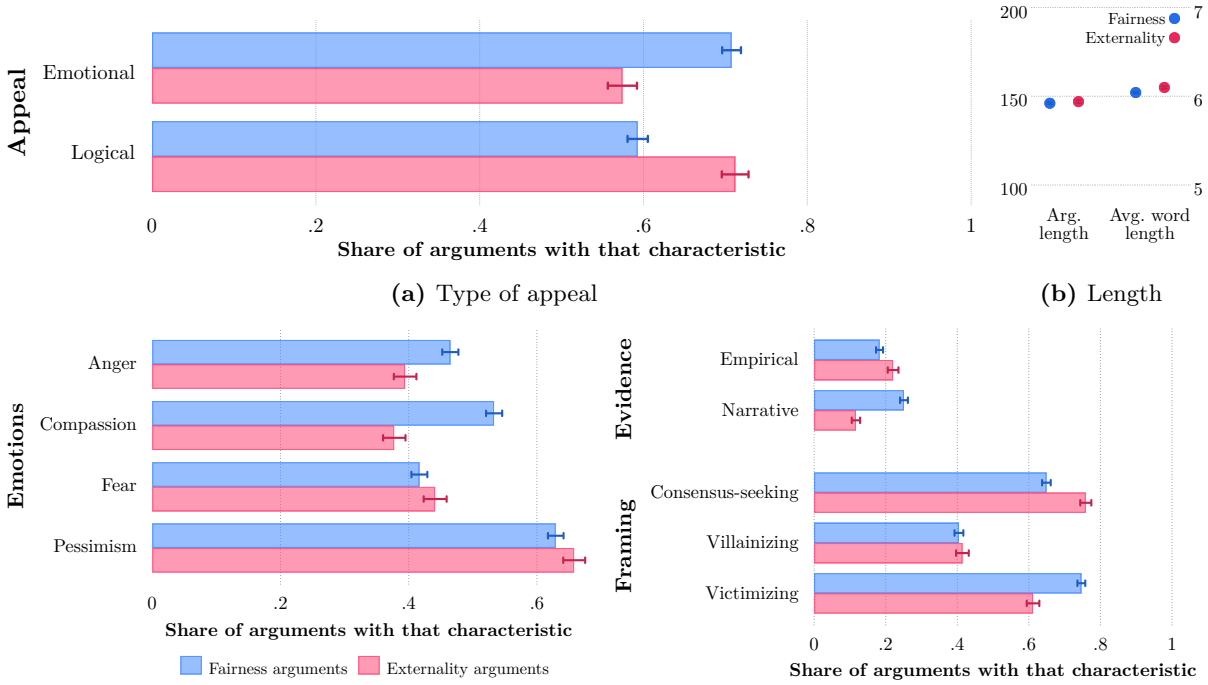
Note. This figure complements Section 5 and Appendix C.1, reporting content differences across fairness- and externality-based excerpts across the U.S. Congress and the Norwegian *Storting*. In each dataset, we compare (i) the main specification (*Fair. Only* and *Ext. Only*), (ii) the *Fair. All* and *Ext. All* classification, which include arguments tagged as both fairness- and externality based, (iii) the *Fair. Primarily* and *Ext. Primarily* classification, in which each argument can be categorised as either a fairness- or an externality-based argument. The U.S. Congress data also includes *Fair Only, Alt.* and *Ext Only, Alt.* which is akin to *Fair Only* and *Ext Only* but uses an alternative prompt.

Figure F6: Content gaps between fairness and externality excerpts are consistent across prompts



Note. This figure complements Section 5.3 and reports content differences between fairness and externality arguments in the U.S. Congress data and the Norwegian *Storting* data, using our main prompts as well as three alternative versions of the prompts. The full text for the prompts are shown in Appendix B.16–B.21. The differences are computed in the same way as in Figure 5, subtracting the share of externality arguments containing that characteristic to the share of fairness arguments containing that characteristic. 95% confidence intervals are shown.

Figure F7: Norwegian *Storting* Speeches: Content of arguments across argument type, using primarily classification

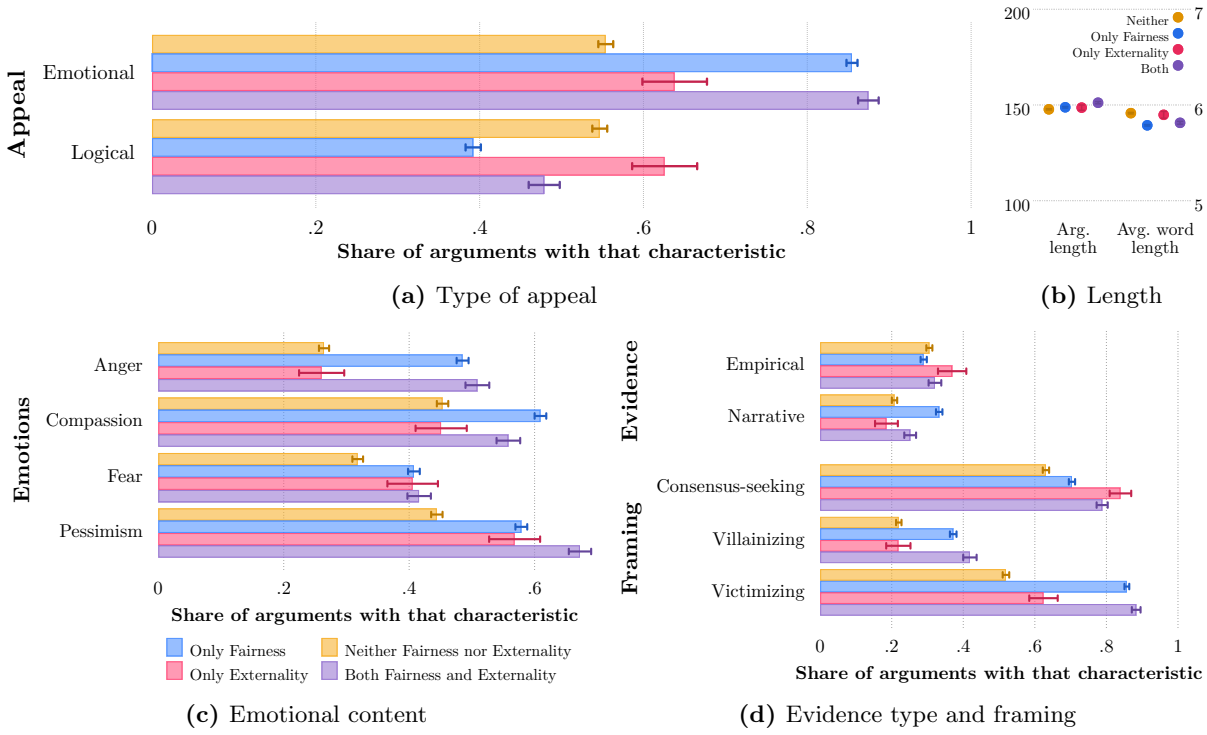


(a) Type of appeal **(b)** Length

(c) Emotional content **(d)** Evidence type and framing

Note. This figure complements Figure F3, showing the content of fairness and externality-based excerpts in the Norwegian *Storting* using the "primarily" classification from Appendix B.15. The classification method is discussed further in Appendix C.1; each argument is classified as primarily making a fairness-based argument, primarily making an externality-based argument, or as not clearly making either type of argument. Appendix B shows the full prompt texts.

Figure F8: U.S. Congressional Speeches: Content of arguments across argument type, all four types

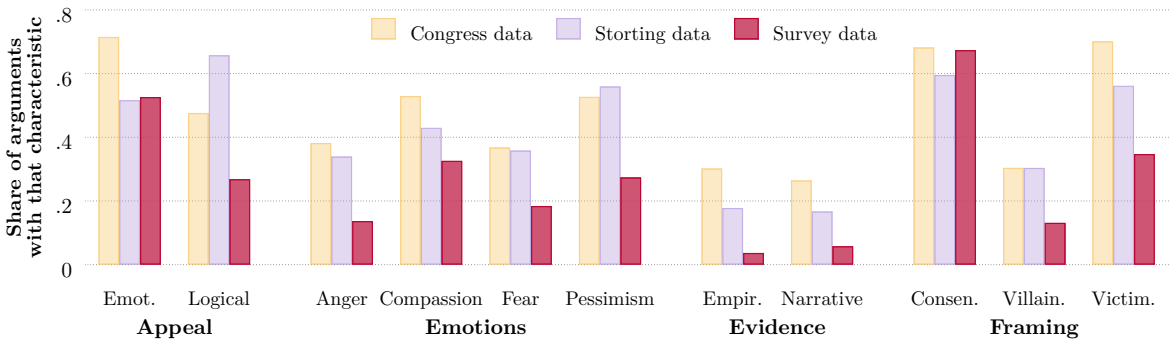


(a) Type of appeal **(b)** Length

(c) Emotional content **(d)** Evidence type and framing

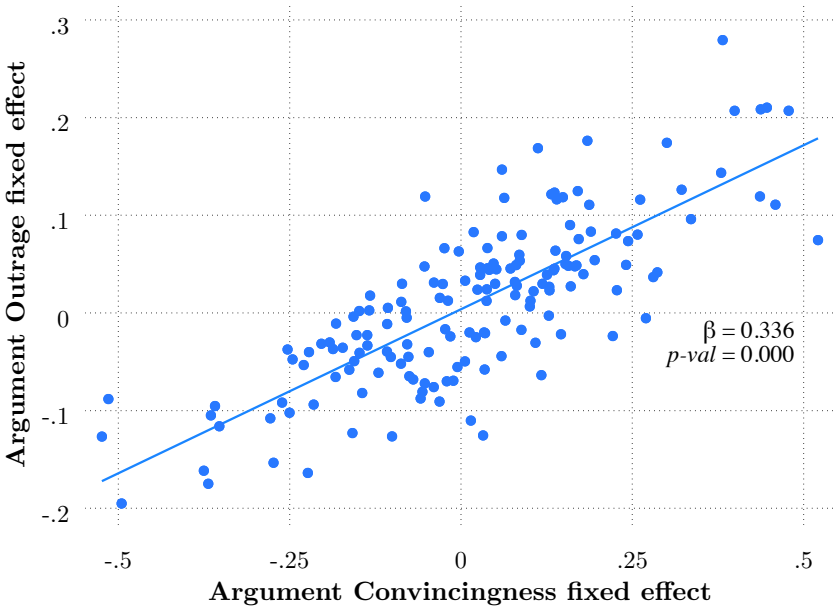
Note. This figure complements Figure 2, showing content classifications for speeches that were tagged as containing *both* a fairness and an externality argument, or *neither* a fairness nor an externality argument. Results are based on the full sample, including speeches that contain both fairness and externality arguments. In total there are 13,250 fairness arguments and 3,244 externality arguments, with 2,671 overlapping. Note that this means that 82% of speeches with externality arguments contain fairness arguments. Standard errors are robust, 95% CIs.

Figure F9: Mean argument characteristics by sample



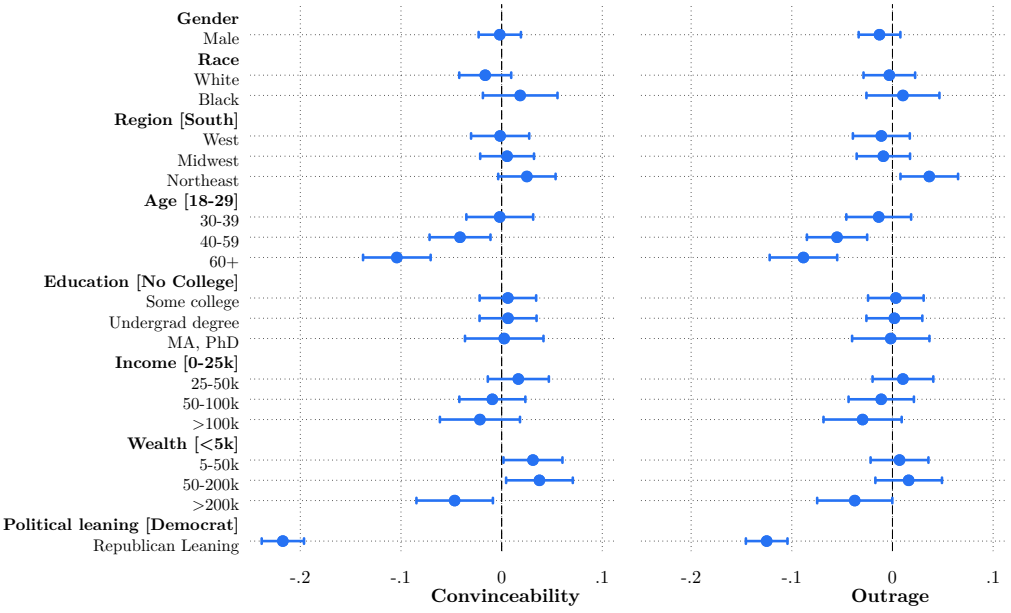
Note. This figure shows, for each characteristic, the average share of excerpts or arguments containing that characteristic across the three samples (the U.S. Congress, the Norwegian *Storting*, and the experimental sample). In the legislative samples, all excerpts classified as including a redistributive argument are included. Appendix B shows the full prompt texts.

Figure F10: Association between convincingness and outrage



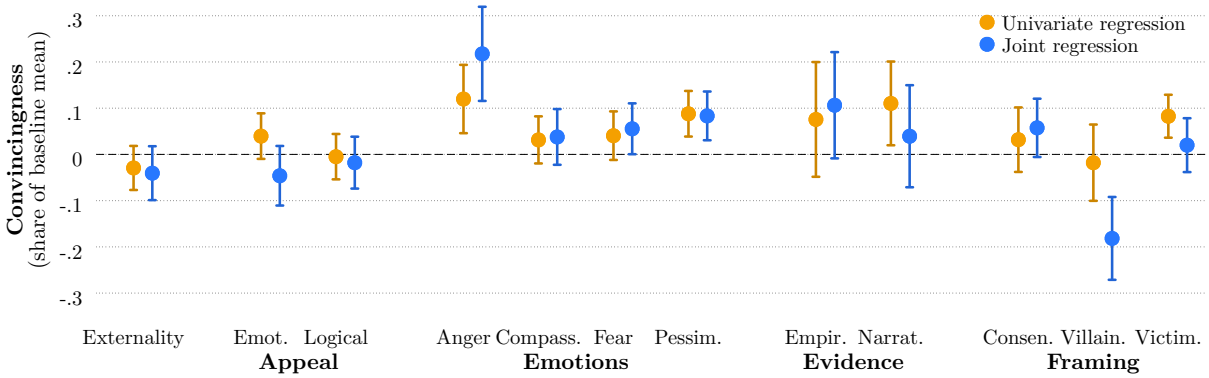
Note. This figure complements Figures 9-10, showing the association between convincingness and outrage for each of the 160 arguments in the survey experiment.

Figure F11: Determinants of respondents' convinceability and propensity to be outraged



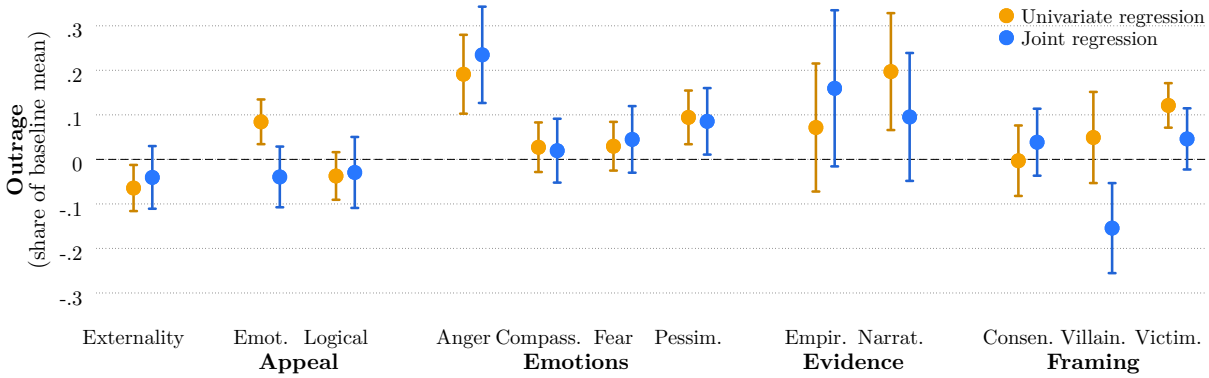
Note. This figure complements Section 4.4, reporting which demographic variables correlate with the respondent being (i) convinced and (ii) outraged by a given argument. In both cases, we show regression coefficients in p.p. from a joint regression on the outcome variable. 95% confidence intervals shown.

Figure F12: Association between convincingness and argument characteristics, attentive participants only



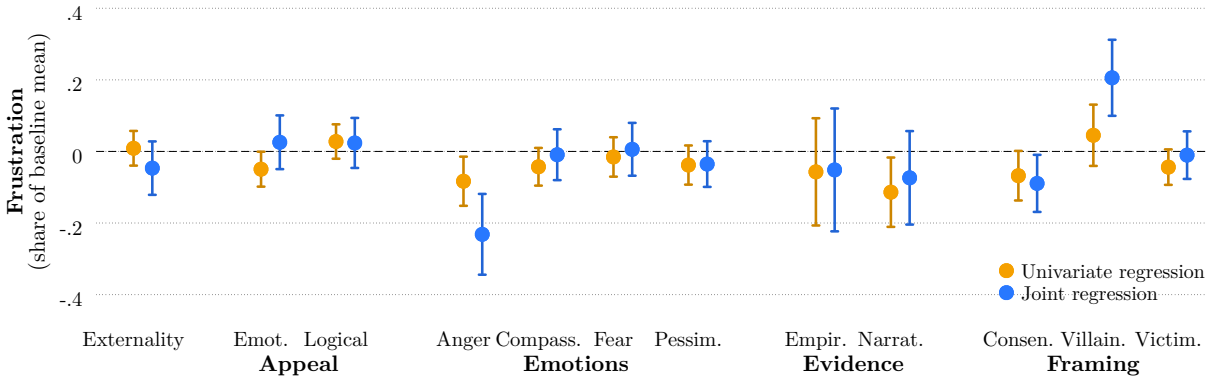
Note. This figure shows the result of regressing convincingness on argument characteristics. The sample is restricted to participants who passed the attention check (72.3% of the 4,089 participants). The yellow dots correspond to a regression of convincingness on the indicated variable, as well as person fixed effects, a dummy for above-median argument length, and the average word length in characters. The blue dots correspond to the joint regression of Convincingness on all characteristics, with the same controls. Standard errors are clustered at the argument level. Back to Figure 9.

Figure F13: Association between outrage and argument characteristics, attentive respondents only



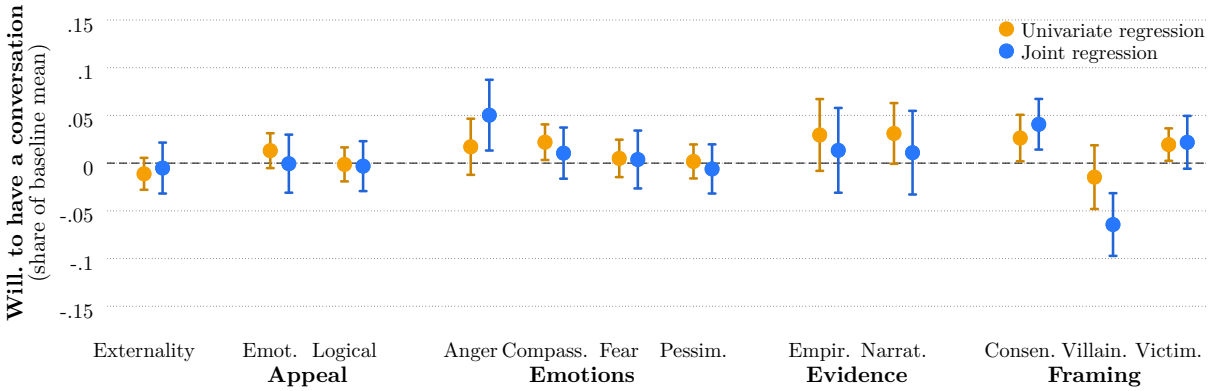
Note. This figure shows the result of regressing outrage on argument characteristics. The sample is restricted to participants who passed the attention check (72.3% of the 4,089 participants). The yellow dots correspond to a regression of outrage on the indicated variable, as well as person fixed effects, a dummy for above-median argument length, and the average word length in characters. The blue dots correspond to the joint regression of outrage on all characteristics, with the same controls. Standard errors are clustered at the argument level. 95% confidence intervals are shown. Back to Figure 10.

Figure F14: Association between frustration and argument characteristics



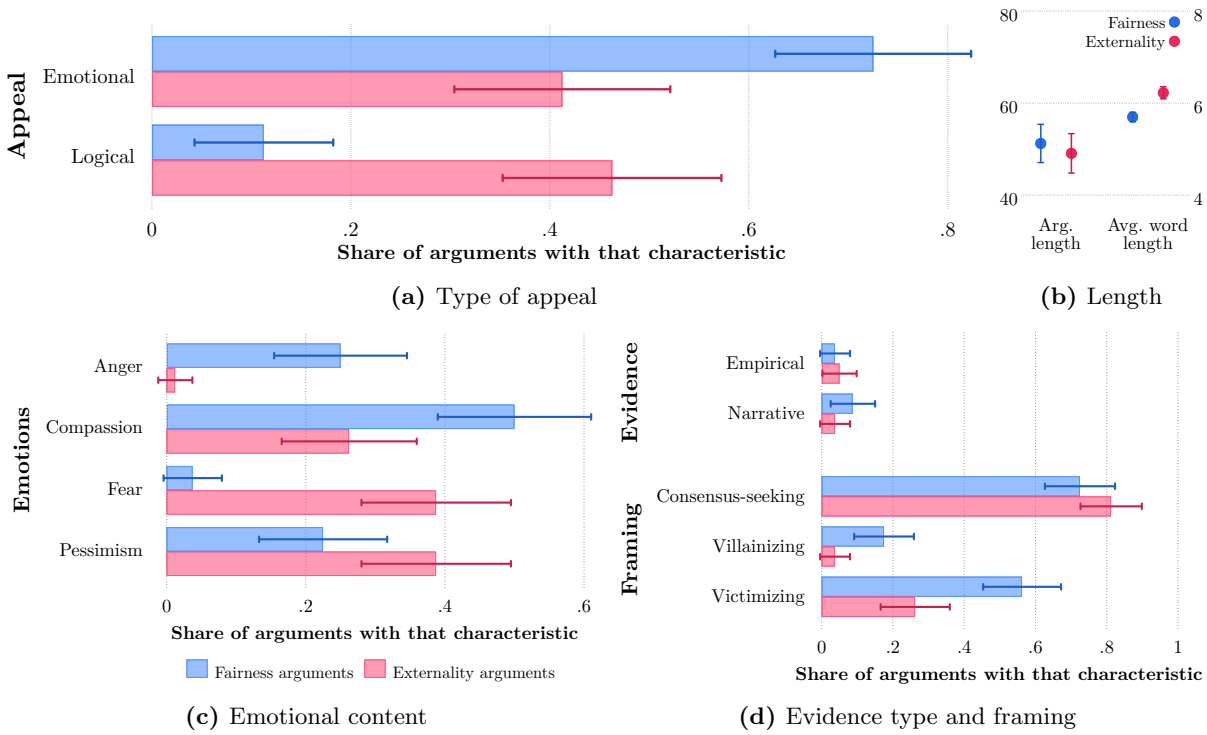
Note. This figure shows the result of regressing frustration on argument characteristics. The yellow dots correspond to a regression of frustration on the indicated variable, as well as person fixed effects, a dummy for above-median argument length, and the average word length in characters. The blue dots correspond to the joint regression of Frustration on all characteristics, with the same controls. Standard errors are clustered at the argument level, and 95% confidence intervals are shown. Back to Section 4.4.4.

Figure F15: Association between willingness to have a longer conversation and argument characteristics



Note. This figure complements Appendix E.2, showing the result of regressing willingness to have a longer conversation on argument characteristics. The yellow dots correspond to a regression of willingness to have a longer conversation on the indicated variable, as well as person fixed effects, a dummy for above-median argument length, and the average word length in characters. The blue dots correspond to the joint regression of willingness to have a longer conversation on all characteristics, with the same controls. Standard errors are clustered at the argument level, 95% CIs.

Figure G16: Content of arguments across argument type, experiment data



(a) Type of appeal

(b) Length

(c) Emotional content

(d) Evidence type and framing

Note. This figure complements Figure F9, showing the net amount of different content classifications across fairness and externality arguments (as defined in Survey 1) in the experimental data. Appendix B shows the full prompt texts.

G Appendix Tables

Table G1: Effect of district-level education level on U.S. Congress members' use of fairness and externality arguments, robustness

	Fair. (All)		**Fair. (Only)**		**Fair. (Primarily)**	
	(1)	(2)	(3)	(4)	(5)	(6)
District educ. level	-0.031***	-0.032***	-0.035***	-0.036***	-0.014	-0.014
State top 10% inc. share		0.216		0.264		0.081
Log. avg. state inc.		0.022		0.038		0.003
Member's educ. level		0.008		0.010		0.002
Observations	1186	1182	1178	1174	1059	1056
Year FE, Member controls	Yes	Yes	Yes	Yes	Yes	Yes

	Ext. (All)		**Ext. (Only)**		**Ext. (Primarily)**	
	(1)	(2)	(3)	(4)	(5)	(6)
District educ. level	0.016**	0.017**	0.012***	0.013***	0.014	0.014
State top 10% inc. share		-0.048		-0.046		-0.081
Log. avg. state inc.		-0.052		-0.049*		-0.003
Member's educ. level		0.002		0.001		-0.002
Observations	1186	1182	1178	1174	1059	1056
Year FE, Member controls	Yes	Yes	Yes	Yes	Yes	Yes

Note. This table complements Table 1, showing the relationship between district education level and use of fairness and externality arguments by U.S. Congress members. We use four definitions of the dependent variable: *Fair/Ext. (All)* is equal to 1 if the speech contains an externality- or fairness-based argument, regardless of whether it also contains an argument of the other type. In *Fair/Ext. (Only)*, we exclude speeches that contain both a fairness and an externality argument; the variable is equal to 1 if the speech contains a fairness or an externality argument. *Fair/Ext. (primarily)* corresponds to the exclusive classification described in Section C.1: each argument is classified as either a fairness or an externality argument. *Fair/Ext. (Only, alt.)* is similar to *Fair/Ext. (Only)*, but uses an alternative prompt. Controls include state-level top 10% income share, log state income per capita, Congress member's education level in 4 categories (less than Bachelor's, Bachelor's, Master's, Professional degree or PhD), year fixed effects, a dummy for being a Republican, and a dummy for male gender. *Significance levels*: *10%, **5%, ***1%.

Table G2: Effect of district-level education level on Norwegian *Storting* members' use of externality arguments, robustness

	Fair. (All)		**Fair. (Only)**		**Fair. (Primarily)**	
	(1)	(2)	(3)	(4)	(5)	(6)
County educ. level	0.005	-0.010	-0.016	-0.067	-0.031	-0.130*
County-level Gini		0.854		1.173		1.865*
Log. avg. county inc.		-0.122		-0.140		-0.242
Member's educ. level		-0.031		-0.006		-0.002
Observations	1705	906	1604	866	1306	702
Year FE, Member controls	Yes	Yes	Yes	Yes	Yes	Yes

	Ext. (All)		**Ext. (Only)**		**Ext. (Primarily)**	
	(1)	(2)	(3)	(4)	(5)	(6)
County educ. level	0.040***	0.088*	0.017	0.036	0.031	0.130*
County-level Gini		-0.494		-0.457		-1.865*
Log. avg. county inc.		-0.018		-0.006		0.242
Member's educ. level		-0.024		0.000		0.002
Observations	1705	906	1604	866	1306	702
Year FE, Member controls	Yes	Yes	Yes	Yes	Yes	Yes

Note. This table complements Table 2, showing the relationship between district education level and use of fairness and externality arguments by Norwegian *Storting* members. We use three definitions of the dependent variable: *Fair/Ext. (All)* is equal to 1 if the speech contains an externality- or fairness-based argument, regardless of whether it also contains an argument of the other type. In *Fair/Ext. (Only)*, we exclude speeches that contain both a fairness and an externality argument; the variable is equal to 1 if the speech contains a fairness or an externality argument. *Fair/Ext. (primarily)* corresponds to the exclusive classification described in Section C.1: each argument is classified as either a fairness or an externality argument. Controls include district-level income Gini, log district income per capita, *Storting* member's education level in 4 categories (less than Bachelor's, Bachelor's, Master's, PhD), year fixed effects, a dummy for each political party, and a dummy for male gender. *Significance levels*: *10%, **5%, ***1%.

Table G3: Most and least convincing arguments

3 most convincing arguments	**3 least convincing arguments**
– The workers themselves should get raises. How many corperations where the CEOs are making millions meanwhile the people that do the actual grunt work to make that money arent even getting a living wage? How would you feel if you worked your fingers to the bone for pennies meanwhile some suit sits there, does basically nothing and just rakes in the cash? [Fair][Anger, Compassion]	– Wealth should be redistributed away from the richest individuals because they did nothing in particular to deserve so much wealth. While it is true they usually took some extra risk in business ventures, for example, nothing can justify being a millionaire or billionaire when the average worker (especially for the wealthy person) lives paycheck to paycheck. [Fair]
– Lower pay wages/jobs should make more money, a living wage. The distribution of wealth has steadily grown for the rich and lessened for the workers. CEO's make way too much money. [Fair]	– When people have more money there is less crime. Less crime means there are less victims. Less victims means you hear less complaining. [Ext][Hope]
– Everyone deserves to be able to afford basic necessities such as food, shelter, and healthcare. Every employee should be paid a living wage that is adjusted as the cost of living either increases or decreases. [Fair][Hope, Compassion]	– Economic equality flattens culture. When only a small portion of specific kinds of people can afford to create, publish and market art and content, those people's ideas over time begin to overcome and replace the ideas of everyone else [Ext][Fear]

Note. This table shows the three arguments with the highest and lowest convincingness fixed effect in the experiment data. The fixed effect is obtained by regressing convincingness on both person and argument fixed effects.

Table G4: Most and least outrage-inducing arguments

3 most outrage-inducing arguments	**3 most frustration-inducing arguments**
– Workers are not paid the true value of their labor in relation to the overall profit of their employers and shareholders. The pay of CEO's and executives is so unfair in relation to that of the labor force that it is obscene. The system of putting profits to shareholders above the basic needs of the workers is inherently unfair. [Fair]	– There are not enough high paying jobs to go around, so some people are going to end up unemployed or underemployed through no fault of their own. We can afford to compensate these economic losers more than we already do. Because we do so little redistribution now we can afford a modest increase without significantly affecting peoples' motivation to work and earn money. [Fair][Hope]
– People making low/minimum wages struggle to make ends meet - often leading to unhappiness, frustrations, broken marriages and inability raise kids well (due to multiple jobs). This has a huge downstream impact on overall society. [Ext][Fear]	– Wealth should be redistributed away from the richest individuals because they did nothing in particular to deserve so much wealth. While it is true they usually took some extra risk in business ventures, for example, nothing can justify being a millionaire or billionaire when the average worker (especially for the wealthy person) lives paycheck to paycheck. [Fair]
– I am an educator and I can tell you without a doubt that inequality affects children greatly. The kids in my first grade class who come from families in poverty always have a more difficult time accessing their learning than those who don't. This makes sense because, in the hierarchy of needs, you can't effectively concentrate about how to spell said if you don't know that you will have food to eat or be in a warm bed at night. [Ext][Compassion]	– It would cause more people to get the money that they deserve instead of it only going to a select group. Mostly to poor people though, as they are the ones who need it the most. [Fair][Hope]

Note. This table shows the three arguments with the highest and lowest outrage fixed effect in the experiment data. The fixed effect is obtained by regressing outrage on both person and argument fixed effects.

Table G5: Effect of arguments and their characteristics on redistributive preferences: Only respondents who succeeded on an attention check

	Post-evaluation redistributive preferences					
	(1)	(2)	(3)	(4)	(5)	(6)
Pre-evaluation RP	0.303***	0.304***	0.145***	0.303***	0.146***	0.304***
	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)
N pro-redistr. arguments	0.019***		0.017***	0.019***		
	(0.007)		(0.007)	(0.007)		
N pro-redistr. arguments (fair)		0.024***			0.021***	0.024***
		(0.008)			(0.007)	(0.008)
N pro-redistr. arguments (ext)		0.014*			0.014**	0.014*
		(0.008)			(0.007)	(0.008)
Avg. convincingness, own			0.654***		0.654***	
			(0.029)		(0.029)	
Avg. outrage, own			0.038		0.037	
			(0.025)		(0.025)	
Avg. convincingness, others				0.607		0.668
				(0.416)		(0.417)
Avg. outrage, others				-0.250		-0.409
				(0.605)		(0.611)
Adjusted R^2	0.293	0.294	0.436	0.294	0.436	0.294
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2930	2930	2930	2930	2930	2930

Note. This table supplements Table 3, reporting the effect of arguments and their characteristics on redistributive preferences among respondents who succeeded on an attention check immediately after the evaluations. The attention check was framed as a normal argument evaluation, with the "argument text" as follows: *Hello! This is from the people writing the survey. We just want to check that you're reading the arguments. Please click "Very convinced", "No", and "No, not at all" below. This is the last "argument". Thank you for reading carefully!*. 72% of respondents answered this attention check correctly. Each respondent evaluates ten redistributive arguments, of which 85% are pro-redistribution and 15% anti-redistribution on average. "Post-evaluation redistributive preferences" is a dummy for the respondent agreeing that "The government should take measures to reduce differences in income levels". *Pre-evaluation RP* is the respondent's pre-evaluation preference for redistribution, a dummy for answering above a 4 to "*How much redistribution of income do you prefer across citizens in the U.S.?*", where 0 is "*No redistribution*" and 7 is "*Full redistribution*" (results are robust to changing this threshold). *N pro-redistr. arguments* is the number of pro-redistributive arguments seen, overall and by type (fairness or externality). *Avg. convincingness, own* and *Avg. outrage, own* are the respondent's average self-reported evaluations of convincingness and outrage for the arguments they rated. *Avg. convincingness, others* and *Avg. outrage, others* are the corresponding leave-one-out averages across all other respondents evaluating the same arguments. These measures are used to test whether exposure to more, or more convincing, pro-redistributive arguments increases stated redistributive preferences. Included controls are standard demographic and socioeconomic characteristics: political leaning (Republican Leaning), gender (Male), race (Black, Other Race), income brackets ($25–50k, $50–100k, $100k+), age groups (30–39, 40–49, 50–59, 60–69, 70+), education (College or more), employment status (Unemployed, Not in workforce), and region (West, Northeast, Midwest). *Significance levels*: *10%, **5%, ***1%.

Table G6: Distribution of pro-redistributive arguments seen

N pro-redistr. arguments	Number of respondents
4	1
5	16
6	102
7	341
8	718
9	821
10	390

Note. This table supplements Table 3, reporting number of pro-redistributive arguments seen per respondent. Each respondent evaluates ten redistributive arguments, of which 85% are pro-redistribution and 15% anti-redistribution on average.

H Variables

H.1 Survey 1 (Elicitation)

H.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?

H.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?

H.2 Survey 2 (Quality check)

H.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.)

H.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

H.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

H.3 Survey 3 (Argument evaluations)

H.3.1 Redistributive preferences, pre-argument evaluation

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 (so without the taxes and laws we have today)
Full redistribution means that all citizens should have the same income
[*1-7 slider, where 1 = no redistribution, 7 = full redistribution*]

H.3.2 Agreement with positive inequality externality

Do you agree or disagree with the following statement:
Large differences in income are necessary for America's prosperity.

· Agree strongly
· Agree
· Neither agree nor disagree
· Disagree
· Disagree strongly

H.3.3 Introduction, argument evaluation

You are now beginning Part 2 of the survey. Here we will show you **10 different arguments** either **for or against reducing the economic differences between people**. This is the largest part of the survey.

Almost all of the arguments were **written by Americans who, like you, answered an online survey**. Some of them want to decrease the differences in incomes and wealth between people, while others oppose such a decrease. We have not altered what was written in any substantial way. This means that there may be grammatical mistakes, for example.

Please read the arguments closely and think about whether you find them convincing — your truthful answers here are important to us.

H.3.4 Convincingness

Another survey respondent is trying to convince you with the following argument for **more redistribution** of income and wealth:
[*Argument text*]
Are you personally convinced by this argument or statement?

· Very convinced
· Somewhat convinced
· Neither convinced nor unconvinced
· Somewhat unconvinced
· Very unconvinced

H.3.5 Willingness to have a conversation

[*displayed on the same page as the previous question*]

Would you be willing to have a longer conversation with this person about these ideas?

· Yes
· No

H.3.6 Agitation

[*displayed on the same page as the two previous questions*]

Imagine someone said this to you in person.

Do you think a discussion about this argument could provoke an emotional reaction like anger or agitation in you?

· 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

H.3.7 Redistributive preferences, post-argument evaluation

To what extent do you agree or disagree with the following statement:

The government should take measures to ***reduce*** *differences in income levels.*

· Agree strongly
· Agree
· Neither agree nor disagree
· Disagree
· Disagree strongly