Equity or Externalities: The Redistributive Debates in Norway and the United States*

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

Click Here for latest version

Abstract

We use natural language processing to classify all redistributive speeches in the U.S. Congress and Norwegian parliament from 2015 to 2022. In Norway, 68% of redistributive excerpts are pro-redistributive, compared with 44% in the United States. The U.S. debate largely follows the classic equity-efficiency divide: pro-redistributive arguments emphasize fairness, while anti-redistributive arguments emphasize efficiency. In Norway, pro-redistributive arguments also frequently invoke the negative societal consequences of inequality (inequality externalities), which appear in 37% of pro-redistributive excerpts, compared with 13% in the United States. Fairness-based arguments are markedly more emotional in both settings, especially expressing anger and compassion, while externality-based arguments are more likely to appeal to logic and consensus. Fairness arguments induce more outrage but are not more convincing than externality arguments in a U.S. survey experiment. Our results document two distinct ways to argue for redistribution, challenging the classic equity-efficiency framing, and provide a general framework for comparing the content and evaluations of arguments across domains.

JEL Codes: D63, D72, P16, C55

*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 ⓡ, using the AEA Author Randomization Tool; the resulting randomization can be verified here. Version: March 25, 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.

1 Introduction

Economic inequality has long been a central concern in political and economic life. Yet despite historical conflicts over taxation and redistribution having played a central role in shaping policy (Martin et al., 2009), we know surprisingly little about how redistribution is debated today. Modern research on redistributive preferences has largely focused on agreement with pre-selected survey items, the causal effect of information treatments, or behavior in experimental games (e.g. Kuziemko et al., 2015; Stantcheva, 2021; Hvidberg et al., 2023)—approaches that rely on narrow and stylized stimuli, abstracting from how redistribution is debated in practice. The broader discourse on redistribution and its influence on societal outcomes remains poorly understood.

In this paper, we provide the first large-scale empirical analysis of the contemporary redistributive debate. Using GPT-4o Mini and recent legislative speech databases (Aroyehun et al., 2024; Fiva et al., 2025), we identify all pro- and anti-redistributive arguments in the U.S. Congress and the Norwegian parliament (the *Storting*) from 2015 to 2022. We then focus on a distinction closely related to the classic equity–efficiency divide in public finance—whether arguments evaluate redistribution in terms of justice and desert or in terms of its broader consequences for society. On the pro-redistributive side, this distinguishes between appeals to fairness (Konow, 2000) and appeals to the broader societal consequences of inequality, which we refer to as inequality externalities—that is, effects of inequality on outcomes such as social cohesion, democratic institutions, or productivity (Støstad and Cowell, 2024). On the anti-redistributive side, we distinguish between appeals to fairness—for example claims that individuals are entitled to the income they earn (Mankiw, 2010)—and appeals to efficiency (Okun, 1975). In practice, this division is more substantively informative for pro-redistributive excerpts, which are our main focus; for these excerpts we also analyze additional content dimensions such as emotional tone, evidence used, and framing. We remain agnostic on specific redistributive policies or institutional designs.

In the United States, 44% of redistributive excerpts are pro-redistributive, compared with 68% in Norway. Behind these headline numbers we observe significant political polarization in both countries. In the U.S. Congress, 92% of redistributive excerpts by Democrats are pro-redistributive, whereas only 7% by Republicans are—a pronounced form of the party polarization documented in Gentzkow et al. (2019). In Norway, the corresponding shares are 92% for left-leaning parties and 33% for right-leaning parties. While polarization is still high, the more pro-redistributive Norwegian debate is thus almost entirely driven by how often redistribution is argued for by the political right. Every Norwegian political party has a higher share of pro-redistributive arguments than the U.S. Republicans.

Moving to how legislators justify redistribution, the United States debate largely follows the familiar equity-efficiency divide: pro-redistributive arguments emphasize fairness, while anti-redistributive arguments emphasize efficiency. In Norway, the anti-redistributive side also relies heavily on efficiency, but pro-redistributive arguments rely far more heavily on externality reasoning. In the Norwegian *Storting*, 37% of pro-redistributive excerpts contain an externality argument, nearly three times the 13% share in the U.S. Congress. This pattern holds across the political spectrum: conditional on making a redistributive argument, every Norwegian party has a higher externality-to-fairness ratio than either major U.S. party.

Given the cross-country differences in the use of fairness and externality arguments, a natural question is whether the two argument types differ in their content. Through a variety of independent LLM-based classifications, 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 (p.p.) more likely than externality arguments to include an explicit emotional appeal, but 23 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 also more common in fairness arguments, whereas consensus-seeking language and references to empirical evidence appear more frequently in externality arguments.

Two examples illustrate the contrast. Jens Stoltenberg, arguing for redistribution in the Norwegian *Storting*, appeals to collective returns in a speech our methodology classifies as containing a logical appeal and consensus-seeking:

"*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.*"

While Elizabeth Warren, arguing for redistribution in the U.S. Congress, appeals to fairness and desert in a speech our methodology classifies as containing anger and compassion:

"*This just isn't fair. Corporations are wallowing in profits while hard-pressed families are picking up the bill.*"

This contrast raises a broader question about how economists conceptualize redistributive politics. The standard framework in public economics describes redistribution as a trade-off between equity and efficiency. In the United States, the structure of the debate closely resembles this model: arguments for redistribution appeal primarily to fairness, while arguments against emphasize efficiency. In Norway, however, a substantial share of pro-redistributive arguments across the political spectrum are themselves efficiency arguments, appealing to the societal consequences of inequality. This implies a disagreement not between equity and efficiency, but between competing views about the societal consequences of inequality and redistribution – which in turn implies that the equity–efficiency framing may not be the only natural way to discuss inequality.

If different redistributive debates rely on different types of arguments, a natural question is whether citizens receive these argument types differently. A high prevalence of fairness arguments could increase affective polarization, for example, while externality arguments may be rare in the United States simply because citizens find them unconvincing. To study these questions, we conduct a U.S.-based survey experiment in which we first elicit fairness and externality arguments from one set of respondents, and then collect 32,680 evaluations of those arguments from a separate panel. We begin by showing that the content differences between fairness and externality arguments in the legislative data replicate in this controlled setting, strongly suggesting that these are stable properties of the argument types themselves.

We then show that fairness arguments provoke significantly more self-reported outrage in the controlled setting—an effect fully explained by their higher anger content. As such, we hypothesize that a focus on fairness may contribute to the high anger content (Algan et al., 2025) and affective polarization (Boxell et al., 2024) in the U.S. debate. While respondents do not find externality arguments generally less convincing, this masks meaningful educational heterogeneity; non-college-educated respondents find fairness arguments significantly more convincing, a difference we do not detect among college-educated respondents. On the supply side, legislators from more educated districts use more externality arguments in both countries. These patterns—consistent with externality arguments' greater reliance on logical appeals and empirical evidence—point to an educational divide that runs through both the production and reception of redistributive arguments.

In sum, our results document two distinct ways to argue for redistribution. Arguments based on inequality externalities, which are more logical and consensus-seeking, are much more common in the Norwegian Storting than in the U.S. Congress, where emotional fairness arguments dominate. Fairness arguments elicit substantially more outrage, and are especially convincing to less-educated respondents. The choice between these two types of arguments may thus carry consequences not only for who is persuaded, but for the emotional tenor of the redistributive debate itself.

While our evidence does not identify any causal direction between the redistributive debate and policy outcomes, a natural interpretation of our results is that the mix of argument types may influence redistributive policy. The emergence of large-scale redistributive systems may be predicated on the emotional intensity of fairness arguments, for example, as argued by Scheve and Stasavage (2016). By contrast, the consensus-seeking nature of externality arguments may be particularly effective at preserving existing redistributive institutions. 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 differences in fairness views are often suggested as the key driver of variation in contemporary redistributive institutions (e.g. Alesina and Angeletos, 2005; Almås et al., 2020), our results raise the possibility that differences in the prevalence of externality arguments—or the underlying beliefs they reflect—may also play a significant role. We emphasize that this interpretation is speculative: the primary contribution of this paper is to document that two distinct modes of redistributive argument exist, differ systematically in content and reception, and vary markedly in prevalence across political systems.

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 modern economic debates. We complement recent work that quantifies the register of, or extracts narratives from, legislative speech (Gentzkow et al., 2019; 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.

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 accuracy (Le Mens and Gallego, 2025; Soria, 2025; Asirvatham et al., 2026). 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 approximately 5 million API calls.

The analysis speaks to the literature on 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., Fong, 2001; Cappelen et al., 2007; Haidt and Graham, 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 (Jones and McBeth, 2010; Shiller, 2017; Roth et al., 2020; Andre et al., 2023) and framing theory (Chong and Druckman, 2007). Most closely, we build on the small literature that

[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.

studies redistributive rhetoric directly (Scheve and Stasavage, 2016; Hilmar and Sachweh, 2022). Compared with these works, our approach offers much broader coverage: Scheve and Stasavage (2016) hand-code 428 U.K. parliamentary speeches on the income tax across six years, while Hilmar and Sachweh (2022) studies one tax issue in one parliament (the wealth tax in Germany). We instead classify the full redistributive debate in two legislatures, and then compare argument type across countries, parties, content dimensions, and citizen evaluations.

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 cross-country pattern we detect resonates with prior literature on the broader Nordic tradition of defining equality as not only a matter of justice, but of sustaining social cohesion, trust, and economic coordination (Rothstein, 1998; Sejersted, 2011); our contribution is to document the contemporary prevalence of this mode of justification systematically, and to show its comparative rarity in the United States.

The paper is structured as follows. Section 2 defines the key argument types. 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 Equity, efficiency, and externalities

We focus on a distinction related to the classic equity–efficiency divide in public finance—whether arguments evaluate redistribution in terms of justice and desert or in terms of its broader consequences for society. Here we expand on this distinction.

On the pro-redistributive side, we distinguish between *fairness arguments* and *inequality externality arguments*. Fairness arguments justify redistribution on moral grounds, evaluating the incomes individuals receive in themselves (Konow, 2000; Cappelen et al., 2007). Inequality externality arguments justify redistribution by appealing to the societal consequences of inequality—for example, that high inequality undermines civic institutions, social cohesion, or economic performance (Wilkinson and Pickett, 2009; Lobeck and Støstad, 2023). The distinction is related to the familiar normative-positive divide in economics, though the mapping is not exact.[2] While both types of arguments support the same policy direction, they invoke different mechanisms and have different policy implications. To illustrate, consider the optimal marginal income tax rate $\tau(z)$ at income z in a nonlinear income tax framework (Saez, 2001; Støstad and Cowell, 2024):

$$\tau(z) = \frac{1 - \overline{G}(z) + \eta\Omega(z)}{1 - \overline{G}(z) + \alpha(z)\epsilon(z) + \eta\Omega(z)}$$

Here $\alpha(z)$ is the local Pareto parameter (a measure of distributional thickness at z), and $\epsilon(z)$ is the elasticity of taxable income at z. Fairness concerns are represented by $\overline{G}(z) > 0$,

[2]Externality arguments rely on positive claims about the causal effects of inequality, but they still require a normative premise, namely that the outcomes in question are undesirable.

which captures the social planner's normative fairness concerns (the average social welfare weight above income z). Externality concerns are represented by $\eta\Omega(z) = \eta\kappa(z)\alpha(z)\epsilon(z) + \eta\overline{\kappa}(z)$, which depends on the externality magnitude η, the inequality metric $\kappa(z)$,[3] and individuals' behavioral responses $\alpha(z)\epsilon(z)$.

The optimal tax consequences of $-\overline{G}(z)$ and $\eta\Omega(z)$ differ, notably in how they affect top tax rates. Consider how both affect $\tau(z)$ when z is high. The maximal value of $-\overline{G}(z)$ is zero, as $\overline{G}$ is bounded below by 0 under the assumption that the social planner avoids intentionally harming any one individual absent other benefits. The externality channel, however, can be large and positive, illustrating that social welfare may be improved if top income-earners substitute from labor to leisure—reducing income inequality and thus externalities.[4] Intuitively, the welfare effects of agents' behavioral responses to taxation, which are treated as entirely welfare-negative under a standard fairness framework, may be positive. This reflects a difference in objectives: fairness concerns evaluate the incomes individuals receive, while externality concerns treat inequality itself as an outcome to be minimized. This calls for Pigouvian correction even absent any fairness-based motive. We discuss the conceptual distinction further in Appendix A.

On the anti-redistributive side, the analogous distinction is between fairness arguments and efficiency arguments. Fairness arguments against redistribution hold that people are entitled to their market outcomes, that redistributive taxation is unjust, or that redistribution violates principles such as desert, property rights, or equal treatment (e.g. Nozick, 1974; Mankiw, 2010). Efficiency arguments hold that redistribution distorts incentives and harms growth—that is, redistribution is costly for society (e.g., Okun, 1975). Again, the policy implications of these two channels differ: entitlement-based fairness concerns ($\overline{G}(z)$) and incentive-based efficiency concerns ($\alpha(z)\epsilon(z)$ or positive inequality externalities $\eta\Omega(z)$) yield distinct optimal tax schedules (Saez and Stantcheva, 2016; Støstad and Cowell, 2024).

3 Legislative Speeches

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

Data: U.S. Congressional Speeches 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.

Data: Norwegian *Storting* speeches We use the database of Fiva et al. (2025), modified for comparability with the Aroyehun et al. (2024) database by dividing all speeches into the same 150-word excerpts and restricting the analysis to between 2015 and 2022. The dataset contains 258,490 excerpts.

[3] $\kappa(z)$ denotes the weight of the agent's income z in the inequality metric. $\overline{\kappa}(z)$ is the average weight above z.

[4] Policy implications from fairness and externality concerns also differ under negative social welfare weights. Under such weights, the social planner's aim is to harm the top agent; under a negative inequality externality, the social planner's aim is to lower income inequality, ideally through incentivizing top income-earners to substitute into leisure.

Unit of analysis Following Aroyehun et al. (2024), our unit of analysis is the 150-word speech excerpt. This fixed-length segmentation offers several advantages. First, it ensures comparability both across countries—where speech norms and formats differ—and across excerpts within a country, as each observation covers a similar amount of text. Second, it avoids the need to define where one argument begins and another ends, a boundary that is inherently ambiguous in flowing parliamentary speech. Third, using full speeches as the unit of analysis would produce very long and heterogeneous texts—U.S. Congressional speeches regularly exceed 1,000 words—making content classification excessively polluted while possibly exceeding LLM context windows. The main cost of our approach is that a 150-word excerpt may contain more than one argument or topic, introducing within-excerpt noise. We address this concern through a controlled survey experiment and a range of robustness checks described below.

3.1 Methodology: Classification with large language models

Using GPT-4o Mini with a Python API wrapper, we classify various properties of each speech. The classification methodology is identical across the two datasets. The prompt for each classification is long and listed in Appendix B, along with further methodological details; we summarize the prompts here. Section 4 shows that the method accurately classifies survey-elicited arguments whose type is known by design, and Section 5 shows that the main results are robust across a wide range of alternative classification approaches, including alternative prompts, open-source LLMs, keyword-based methods, and pre-translation of the input text.

We first independently classify whether each speech excerpt contains an argument in favor or against 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*". Prompt in Appendix B.1.
- **Anti-redistributive arguments:** If the speech excerpt contains an argument "*in favor of decreasing economic redistribution*", giving the examples "*advocating for lower taxes on the wealthy, supporting tax breaks on big corporations, restricting social programs, decreasing the minimum wage, redistributing income or wealth less, etc*". Prompt in Appendix B.2.

We then classify each redistributive excerpt according to the type of argument it contains, following the distinctions introduced in Section 2. For each excerpt classified as pro-redistributive, we independently classify whether it contains a fairness-based argument and/or an inequality externality-based argument:

- **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*". Prompt in Appendix B.3.
- **Inequality externality-including excerpt:** If the speech includes an argument based on "*inequality causing negative externalities at the societal level*". 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)*". Prompt in Appendix B.4.

For each excerpt classified as anti-redistributive, we independently classify whether it contains a fairness-based argument and/or whether it contains an efficiency-based argument:

- **Fairness-including excerpt:** If the speech includes an argument based on "*fairness or justice – emphasizing that people are entitled to their outcomes, that redistribution is unfair, or otherwise appealing to justice or moral principles to argue against redistribution*". Prompt in Appendix B.5.
- **Efficiency-including excerpt:** If the speech includes an argument based on "*redistribution causing economic inefficiencies or harmful incentive effects — emphasizing that a redistributive policy harms growth or other socially valuable outcomes, or that higher economic inequality is efficient.*" Prompt in Appendix B.6.

As each 150-word excerpt may touch on multiple themes, the fairness and externality/efficiency classifications are made independently. An excerpt may therefore contain only one, both, or neither type of argument. This yields four categories for each side of the debate: pro-redistributive excerpts are classified as *Fairness only*, *Externality only*, *Both*, or *Neither*; anti-redistributive excerpts are analogously classified as *Fairness only*, *Efficiency only*, *Both*, or *Neither*. For results on the overall prevalence of each argument type—when analyzing how often each is used by different legislators, for example—we use all excerpts tagged with that classification, regardless of overlap. When comparing the content of different argument types—when analyzing whether a fairness-based excerpt contains more anger than an externality-based excerpt, for example—our main specification compares excerpts containing only one type of argument (comparing e.g. the anger content of *Fairness only* excerpts to the anger content of *Externality only* excerpts). This avoids any spillovers from excerpts that blend both types of arguments. For conciseness we define these as *fairness-based excerpts* and *externality-based excerpts*. As we return to in Section 5.2, our main results are robust to other definitions.

Our main focus is on pro-redistributive excerpts, for which we make the following additional content classifications:

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.7.
- **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.8.

Emotions: We classify four types of emotions:

- **Anger:** If the speech expresses "*anger, frustration, resentment, irritation, or outrage.*" Prompt in Appendix B.9.
- **Compassion:** If the speech expresses "*compassion, empathy, care, or concern for others.*" Prompt in Appendix B.10.
- **Fear:** If the speech expresses "*fear, anxiety, worry, concern, dread, or a sense of threat or vulnerability.*" Prompt in Appendix B.11.
- **Pessimism:** If the speech expresses "*pessimism, doubt, hopelessness, or negative expectations.*" Prompt in Appendix B.12.

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.13.
- **Narrative:** If the speech "*uses a personal anecdote, individual case, vivid story, or selective example.*" Prompt in Appendix B.14.

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.15.
- **Villainization:** If the speech "*villainizes a person, group, or entity through its wording and tone.*" Prompt in Appendix B.16.
- **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.17.

In total, we classify all speech excerpts as pro-redistributive or anti-redistributive. We then apply the fairness and efficiency classifications to anti-redistributive excerpts, and the fairness, externality, and 11 content classifications to pro-redistributive excerpts.

We now present results in three steps: first, the overall prevalence of pro- and anti-redistributive arguments; second, the breakdown into fairness, externality, and efficiency arguments; and third, how fairness and externality arguments differ in content.

3.2 Results: Pro- and anti-redistributive arguments

In Figure 1a we show the share of redistributive excerpts that are pro-redistributive across countries and political parties. In the United States, 44% of redistributive excerpts argue in favor of redistribution (25,366 pro vs. 32,331 anti); in Norway, 68% of redistributive excerpts argue in favor of redistribution (17,109 pro vs. 8,660 anti).[5] As such, the Norwegian redistributive

[5]These classifications are independent, but only 0.12% of all excerpts are classified as containing both pro- and anti-redistributive arguments, providing a useful sanity test for our methodology. We note that such excerpts may rationally exist, as speakers can argue for one type of redistribution and against another within the same speech excerpt. These excerpts are included in the remaining analysis; excluding them does not affect our results.

debate is markedly more pro-redistributive than its U.S. counterpart. This establishes the first main result of the paper:

Result #1

In Norway, 68% of redistributive excerpts are pro-redistributive, compared with 44% in the United States.

In a stylized model where redistributive policy reflects the balance of arguments on each side, one might expect the pro- and anti-redistributive shares to converge toward parity in equilibrium. The fact that they do not—and that they differ so markedly across countries—suggests that the relationship between redistributive arguments and redistributive policy is more complex than a simple equilibrium model would predict.

We then examine how the balance between pro- and anti-redistributive arguments varies across political parties. In the United States, redistributive excerpts by Democrats are overwhelmingly pro-redistributive, with 92% coded as pro-redistributive. By contrast, 94% of redistributive excerpts by Republicans are coded as *anti*-redistributive.[6] This sharp sorting is consistent with recent work finding increasing polarization across topics in the U.S. Congress (Gentzkow et al., 2019).

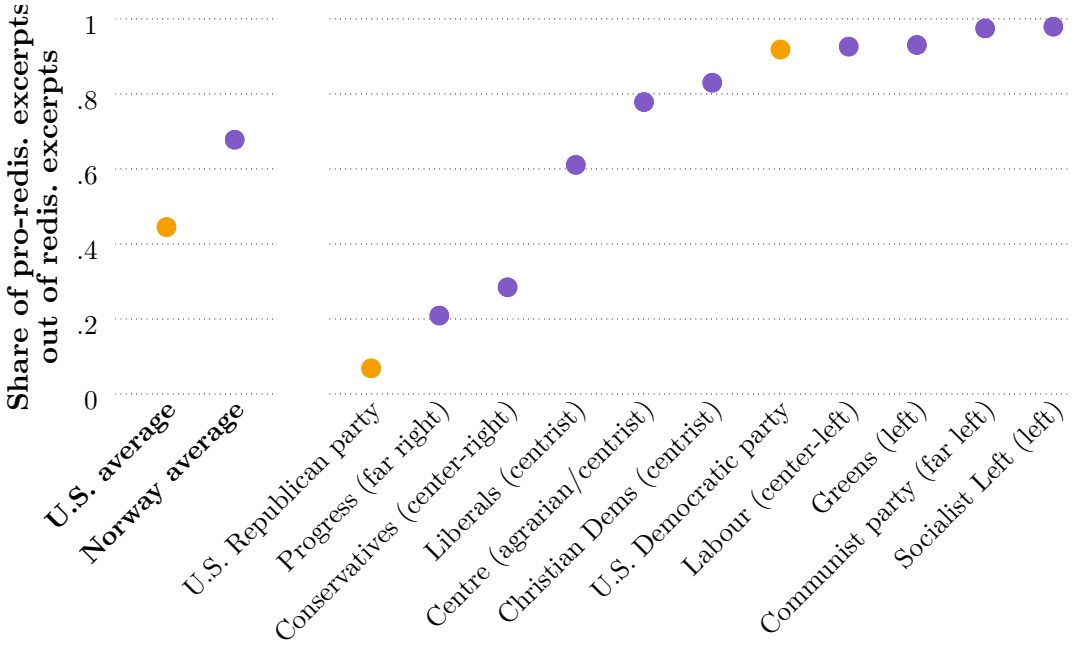
In Norway, pro-redistributive arguments are relatively more frequently made across the political spectrum. Every Norwegian right-leaning party has a higher share of their redistributive excerpts being pro-redistributive than the U.S. Republican Party: the far-right Progress Party (*FrP*) has the lowest share at 20.9%, while the center-right Conservative party (*Høyre*) – one of Norway's two "ruling parties" – has 28.5%. We find high shares of pro-redistributive excerpts among the libertarian and environmentally-friendly Liberals (*Venstre*, 61.1%) and the moderate Christian Democrats (*KrF*, 83.0%). Among left-leaning parties in both countries, the pro-redistributive share is generally high (90%+), e.g. 91.8% for the U.S. Democrats, 92.6% for the Norwegian center-left Labour party (*AP*) – Norway's second "ruling party" – and 97.9% for the Norwegian Socialist Left (*SV*).

Redistributive topics are also discussed somewhat more frequently in the Norwegian *Storting*, where 10% of all speech excerpts contain a redistributive argument (pro- or anti), compared to 8% in the U.S. Congress. Figure 1b shows the share of pro- and anti-redistributive excerpts among all excerpts across countries and political parties. This figure illustrates that the main difference across left-leaning parties is how often they discuss redistribution-related topics. In Norway, the share of excerpts that contain pro-redistributive arguments exceeds one-quarter for the far-left Communist Party (26.5%, *Rødt*), then declines to 16.2% for the Socialist Left, 9.5% for the Labour Party, and 7.0% for the Greens (*MdG*). The equivalent number for the U.S. Democrats is 6.0%. These shares are more comparable across countries on the right. In the U.S. Republican party, 8.1% of all excerpts are anti-redistributive, compared to 9.7% for Norway's Progress Party and 5.2% for Norway's Conservatives.

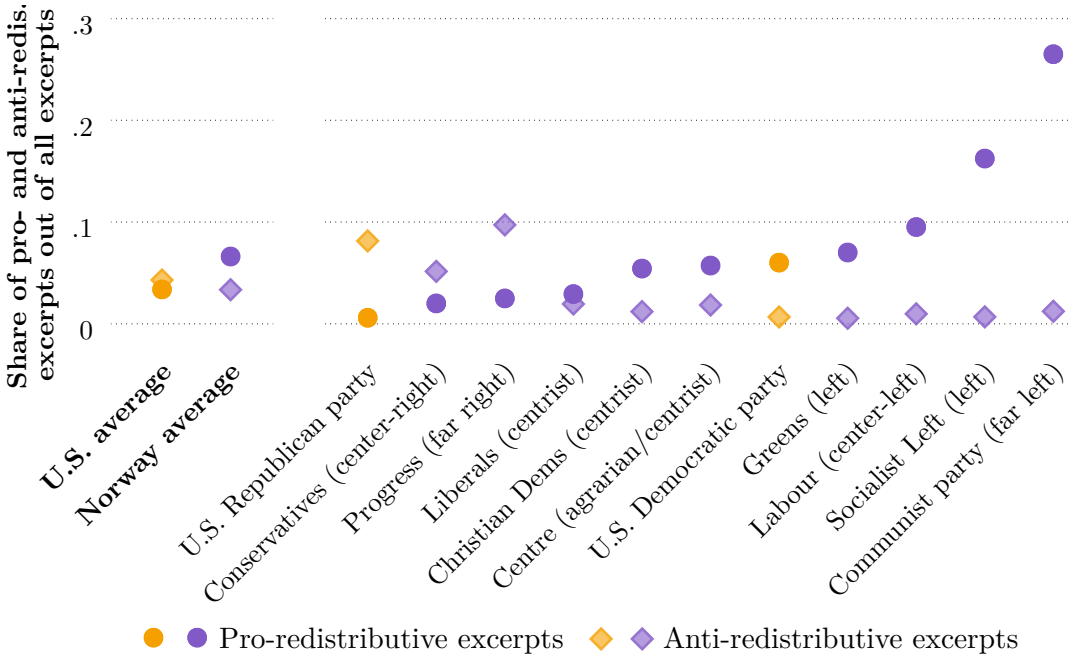
[6] 10% of redistributive excerpts by Democrats are anti-redistributive, and 7% of redistributive excerpts by Republicans are pro-redistributive. As mentioned in Footnote 5, these shares do not sum exactly to 100% because the pro- and anti-redistributive classifications are assigned independently.

Figure 1: Redistributive arguments across political parties

(a) Share of pro-redistributive excerpts in the redistributive debate in the U.S. Congress and Norwegian *Storting*



(b) Share of redistributive excerpts in all debate in the U.S. Congress and Norwegian *Storting*



Note. This figure reports (a) the share of pro-redistributive excerpts among all redistributive (pro or anti) excerpts given across countries and political parties in the U.S. Congress and Norwegian *Storting*, and (b) the share of pro-redistributive and anti-redistributive excerpts among all speeches given across countries and political parties. In (a), the numerator includes all excerpts classified as containing a pro-redistributive argument, while the denominator contains all the excerpts classified as containing a pro- or anti-redistributive argument. In (b), the numerator includes all excerpts classified as containing a pro-redistributive (resp. anti-redistributive) argument, while the denominator contains all speech excerpts. Appendix B shows the full prompt texts.

3.3 Results: Types of redistributive arguments

3.3.1 Pro-redistributive arguments

Having established the overall prevalence of pro- and anti-redistributive arguments, we now turn to the results of the sub-classification into fairness and externality arguments for pro-redistributive excerpts, and fairness and efficiency arguments for anti-redistributive excerpts. In Figure 2a we show these classifications across countries.

In the U.S. Congress, 52.3% of speeches containing arguments in favor of redistribution are classified as including a fairness argument and 12.8% are classified as including an externality argument. Of these, 40.6% of speech excerpts contain *only* fairness arguments. As an example, one excerpt has Ron Wyden arguing that "*we also want everybody to pay their fair share*", and that "*we have been coming together all in the name of fairness*".[7] By comparison, 2.1% of excerpts contain *only* inequality externality arguments. In one example, Ro Khanna describes the benefits of giving money to citizens earning $50-$60,000 as "*they are going to spend that money in their local community; [...] that money is going to go into the local economy, and it is going to actually create jobs*".

10.5% of excerpts contain both types of arguments. This reflects both our excerpt length and the fact that the two logics can be combined even within short statements, for example when a fairness claim is directly linked to broader social consequences. In one excerpt, Darren Soto argues that "*when you have the rich control all capital, all political power, that is when they control us, and we don't have the innovation*" – in the same excerpt as he argues that "*if you don't have a fair market, you can't have successful capitalism, and part of a fair market includes having a strong voice for our middle class*". Reasonable observers may disagree about whether such an excerpt is best understood as fairness-based or externality-based. Our methodology avoids forcing a single classification in these cases. We include these excerpts when analyzing prevalence, but not argument content.

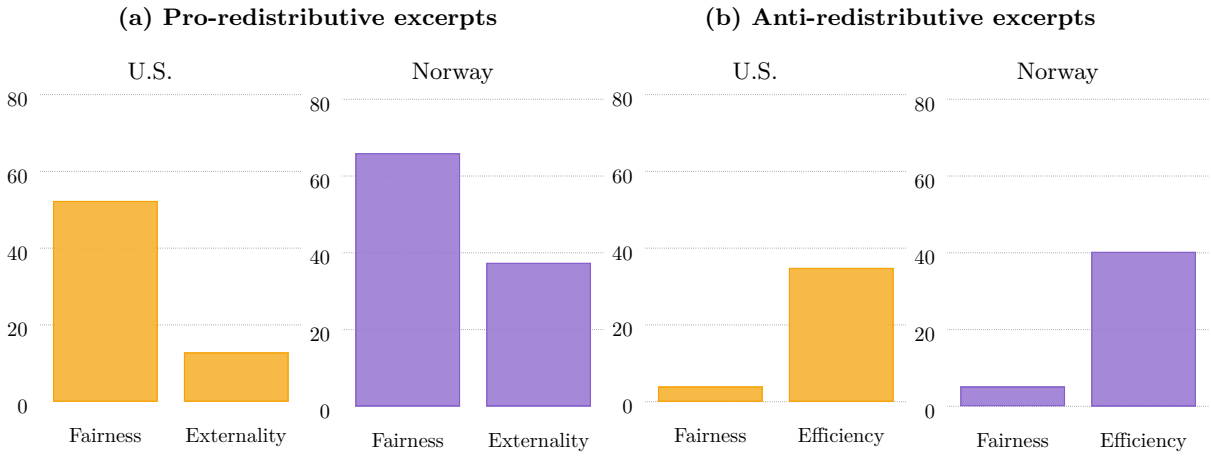
Excerpts classified as *Neither* do not articulate an explicit fairness- or externality-based justification for redistribution. Instead, they usually consist of unelaborated position statements, descriptions of policy provisions or financing, partisan positioning, or claims about helping particular groups without a broader redistributive rationale. Manual inspection suggests that this is a natural residual category in legislative debate. Many legislators simply state that a policy should be expanded, cut, or supported, list what a bill would do, or describe whom it would affect, without further explaining their position. We show examples in Appendix C.3.

In the Norwegian *Storting*, 66.0% of pro-redistributive speech excerpts are classified as including a fairness argument and 37.4% are classified as including an externality argument. Of these excerpts, 32.6% and 4.0% include *only* fairness or externality arguments, respectively.

The shares of redistributive excerpts containing externality arguments in the *Storting* and the U.S. Congress are thus 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. We discuss alternate ways to define this ratio in Appendix

[7] See Appendix C.3 for the full example excerpts.

Figure 2: Types of redistributive arguments in the U.S. Congress and Norwegian *Storting*



Note. This figure represents the share of pro- and anti-redistributive excerpts tagged as fairness- or externality-based, in the U.S. Congress and Norwegian *Storting* data respectively. The shares uses the inclusive definition of arguments types, Fair (All) and Ext. (All) arguments, that is, including arguments that were tagged as both fairness and externality-based. The denominator includes all the arguments tagged as pro- or anti-redistributive, respectively.

C.1—across all choices, externality arguments are much more frequently used in Norway, being between 1.8-5.7 times as common.

3.3.2 Anti-redistributive arguments

As we show in Figure 2b, efficiency arguments dominate in both countries on the anti-redistributive side. Of the anti-redistributive excerpts in the U.S. Congress, 34.9% contain efficiency arguments and 3.8% contain fairness arguments. In the Norwegian *Storting*, 40.3% contain efficiency arguments and 5.1% contain fairness arguments. We note that relatively more of the anti-redistributive arguments are classified as *Neither*, and thus propose few specific justifications; this is consistent across different prompt specifications and validated in manual checks. The relative shares of fairness and efficiency arguments also remain roughly consistent across these different prompts. Given the relative uniformity of anti-redistributive argument types, the remainder of the analysis focuses on the pro-redistributive side.

The two countries thus do not simply differ in the prevalence of pro- and anti-redistributive arguments—they also argue for redistribution in different ways. This establishes the second main result of the paper:

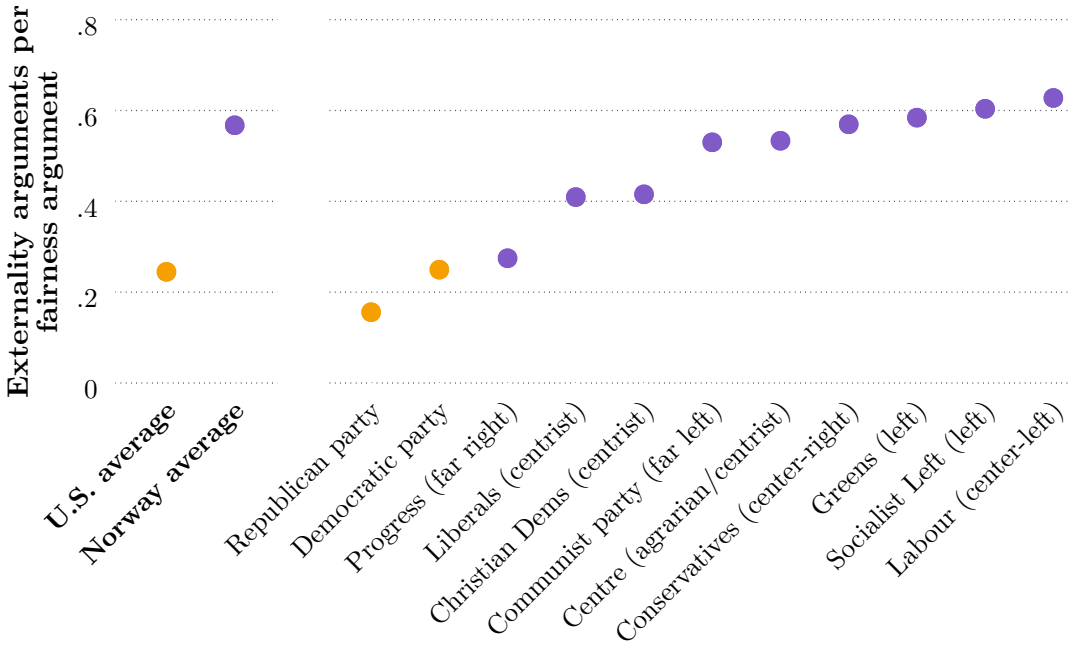
Result #2

The U.S. debate largely follows the classic equity-efficiency divide. In Norway, many arguments for redistribution are themselves efficiency-based, emphasizing the societal consequences of inequality.

We next examine argument type differences across political parties within countries. We show the externality-to-fairness ratio across parties, conditional on making a pro-redistributive argument, in Figure 3.[8] Republicans use 0.16 externality arguments per fairness argument,

[8]In both countries, our sample of pro-redistributive arguments is predominantly made up by parties on the political left (91.4% of excerpts are from Democratic politicians, whereas in Norway, 80.3% are from left-leaning

Figure 3: 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.3 (resp. B.4) shows the prompt classifying the excerpts as containing a fairness (resp. an externality) argument.

compared to 0.25 among Democrats. 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. 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.[9]

3.4 Results: Content differences between fairness and externality arguments

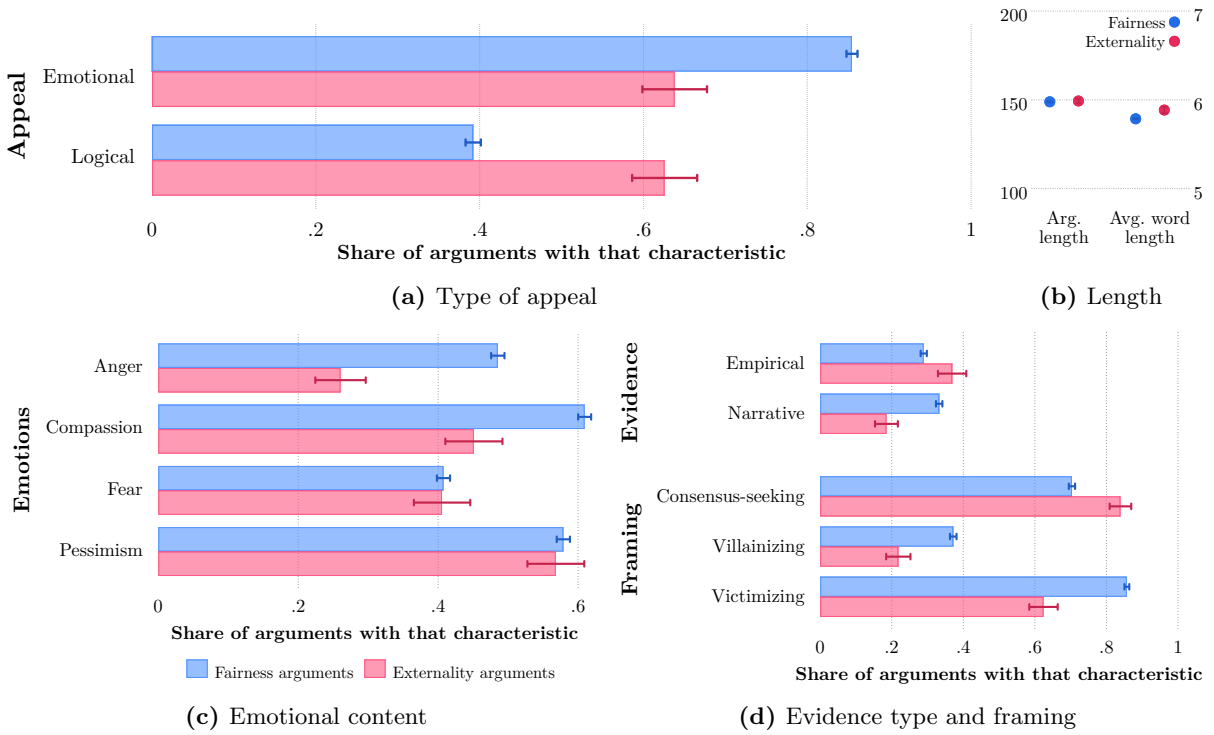
We have shown that the mix of fairness and externality arguments differs sharply across countries. A natural question is whether these two types of arguments also differ in their content — whether they appeal to different emotions, use different types of evidence, and frame the debate differently. We examine this using the 11 content classifications defined in Section 3.1, beginning with the U.S. Congress. Before discussing the content differences, we note that the average excerpt length and average word length within speech excerpts are similar.[10]

parties).

[9]The ratio for the Conservatives is significantly higher than that of any other right-of-center (the Liberals, the Christian Democrats, and the Progress Party) under standard t-tests. The Labour Party has the highest ratio in our sample, significantly higher than that of every party but the Socialist Left and the Greens under standard t-tests. 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).

[10]The average word length is slightly lower in fairness-based excerpts than externality-based excerpts (4.79 vs. 4.89).

Figure 4: 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 In the U.S. Congress, an average of 72% of excerpts contain appeals to emotions, while 48% contain appeals to logic.

In Figure 4a 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 In the U.S. Congress, an average of 38% of excerpts express anger, 53% express compassion, 37% express fear, and 53% express pessimism.

In Figure 4c 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 In the U.S. Congress, an average of 30% of excerpts include empirical evidence, while 26% include narrative-based evidence.

In Figure 4d 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 In the U.S. Congress, 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 4d 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%, $p<0.001$).

Cross-country differences To summarize, fairness arguments in the U.S. Congress are consistently more emotional, adversarial, and narrative-driven, while externality arguments are more analytical, evidence-based, and consensus-seeking. A natural question is whether these patterns also appear in the Norwegian *Storting*.

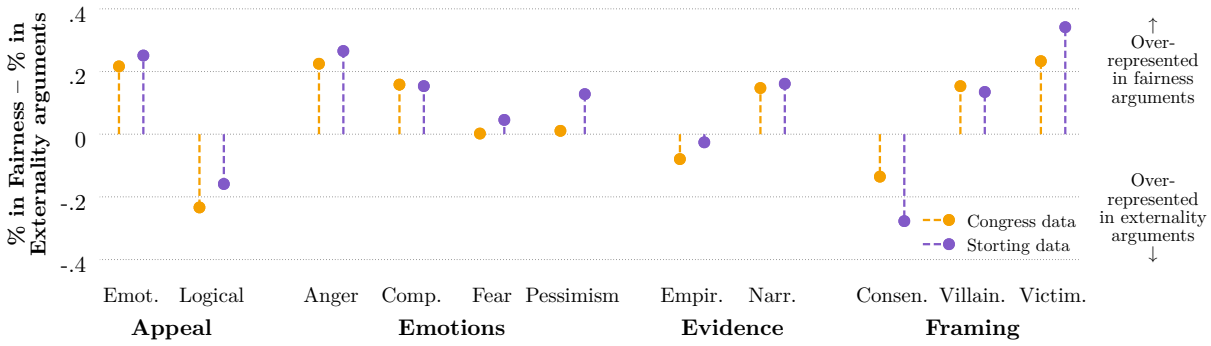
Figure 5 compares content differences between fairness and externality arguments in the U.S. Congressional data (repeating the data from Figure 4) to those in the Norwegian *Storting*. The direction of the comparison (e.g., fairness arguments containing more anger) is consistent across all 11 classifications, and we generally find strikingly small differences across legislatures. The content differences between fairness and externality arguments thus appear to be consistent across countries, ruling out that they are purely an artifact of the U.S. political context.

Beyond the fairness-externality gap, the U.S. and Norwegian redistributive debates also differ in their overall content. U.S. pro-redistributive speeches are, on the whole, more emotionally charged, more likely to express anger, and more likely to victimize and villainize, while Norwegian pro-redistributive speeches are more likely to appeal to logic, cite empirical evidence, and seek consensus (Appendix C.2). Since Norway has a much larger share of externality arguments — which tend to be more analytical — a natural question is how much of this overall difference is compositional. In Appendix C.4, we show that reweighting each country's speeches by the other's mix of fairness and externality arguments closes approximately 10–30% of cross-country content gaps, suggesting that composition contributes to these differences, though the exercise is sensitive to specification and similar cross-country content differences also appear in non-redistributive speech.

3.5 Results: Speaker identity and education

We have documented that Norwegian legislators use substantially more externality arguments than their U.S. counterparts. What might account for this difference? In Table 1, we show associations between Congress members' propensity to use fairness and externality arguments with district characteristics. 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, which is also measured at the state level; and local education level, measured at the

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 4 for the *Storting* data). Figure F4 adds 95% confidence intervals to the content differences. Appendix B shows the full prompt texts.

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 G3 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 female gender. Table G1 includes the coefficients on the controls. *Significance levels*: *10%, **5%, ***1%.

district 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 are 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 G3 in Appendix shows that the education result is robust to different definitions of fairness and externality arguments.

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)	0.242			0.032	0.553			-0.727
Log average county income		0.072		0.025		0.187*		0.034
Avg. county educ. level (years)			0.018	0.011			0.057**	0.094**
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.4. The dependent variable corresponds to the *All* definition, that is, including arguments that were classified as both a fairness- and externality-based argument). Table G4 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. Table G2 shows the coefficients on the controls. *Significance levels*: *10%, **5%, ***1%.

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 legislator in 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 G4 shows that this result is largely robust across different definitions of externality arguments. We return to this pattern in the survey experiment.

3.6 Results: Types of fairness and externality arguments

We finally divide fairness and externality arguments into subtypes: fairness arguments into needs-based, rights-based, and compensatory (merit-based); externality arguments into top-income-based and bottom-income-based. In the U.S. Congress, 61% of fairness excerpts are classified as needs-based, 25% are classified as rights-based, and 13% are classified as compensatory (merit-based). Across inequality externality arguments, 18% are classified as top income-based and 81% are classified as bottom income-based. Shares are very similar in the Norwegian *Storting*. Content is similar across fairness subtypes, while top-income based externality arguments are closer to fairness arguments in style. We discuss these subtypes further in Appendix C.5.

Overall, the legislative analysis has established several findings. The Norwegian redistributive

debate is substantially more pro-redistributive, driven largely by right-leaning parties arguing in favor of redistribution more often than their U.S. counterparts. Externality arguments are two to five times more common in the Norwegian *Storting*, a difference that holds across the political spectrum. Fairness arguments are consistently more emotional, containing more anger, compassion, and narrative evidence, while externality arguments are more logical, containing more consensus-seeking and empirical evidence—patterns that are strikingly consistent across both legislatures. Finally, in both countries, legislators from more educated regions use more externality arguments and fewer fairness arguments.

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

4 Survey Experiments

The legislative analysis documents that fairness and externality arguments differ in content and prevalence, but leaves open whether these differences matter to citizens. The dominance of fairness arguments in the U.S. could reflect Americans finding externality arguments unpersuasive, for example, or that fairness arguments are more effective at mobilizing outrage. While we cannot identify such channels, we can measure whether the two argument types differ in how they are received. To do so, and to create a controlled benchmark for our classification methodology, we conduct survey experiments in which U.S. respondents generate and evaluate redistributive arguments.

4.1 Methodology

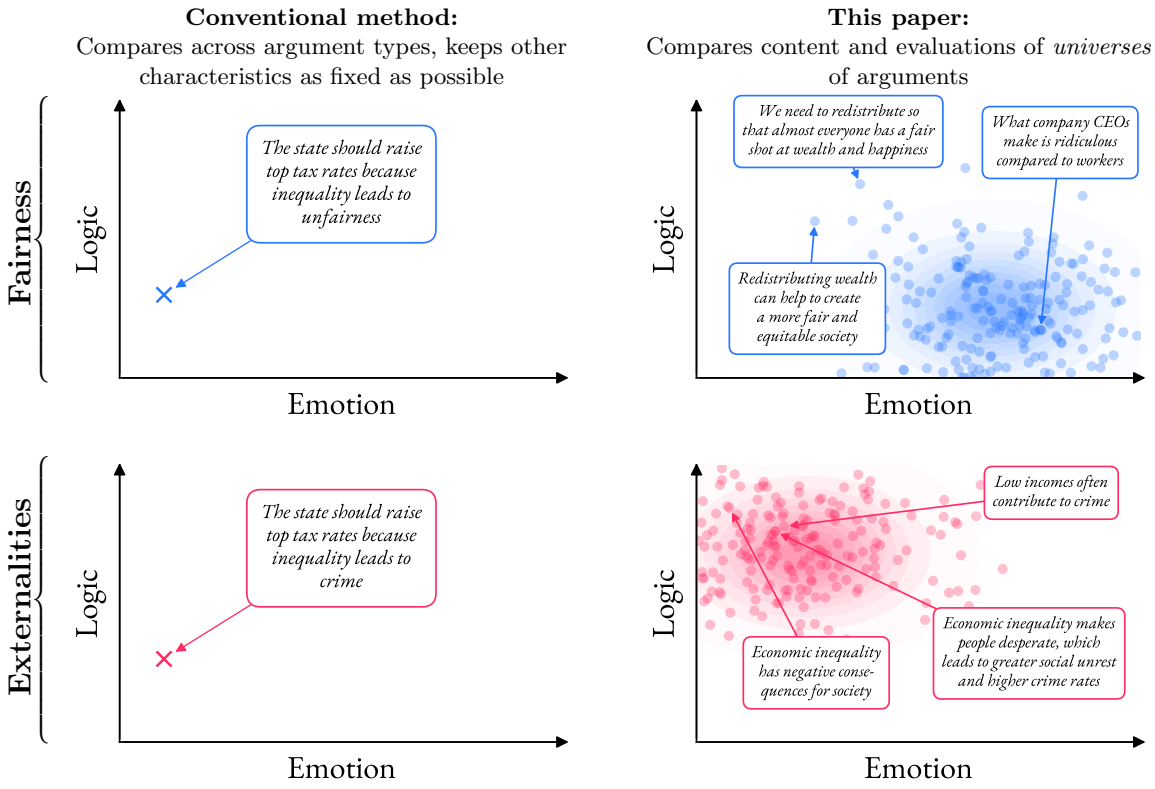
Most survey experiments use tightly controlled stimuli: a single argument is varied in one dimension while keeping all else fixed. This ensures internal validity for comparing the specific stimuli used, but the resulting estimates are stimulus-specific. In practice, it is often a research goal to compare *types* of arguments—asking, for example, whether fairness arguments are more convincing than externality arguments. Such an extrapolation requires assuming that the chosen stimuli are representative of their respective types, however. This assumption is unverifiable by design.

We take a different approach. Rather than selecting a single stimulus per type, we randomly sample arguments from each type as generated by survey respondents. This replaces an unverifiable assumption about the choice of stimuli with a verifiable assumption about the sampling of stimuli. Specifically, the assumption that allows extrapolation is that the sampling matches the external setting of interest. The variation across arguments within a type is not a threat to identification — it is statistical noise that averages out. The result is an estimate of how argument types compare, not how two particular arguments compare. As an additional benefit, the sampled arguments themselves become an object of study, allowing us to characterize the content of each argument type.

4.1.1 Theoretical Framework: Comparing Universes of Arguments

Suppose an argument is characterized by a vector X that includes its observable and unobservable properties in a multi-dimensional space. These properties can represent anything that

Figure 6: 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. Simulated data for illustrative purposes. The argument texts are from our Survey 1 sample.

changes the nature of the argument, for example the broader topic, use of evidence, emotional content, or any other property. We separate this content into the type of the argument X_{type}, defined as the main topic of analysis, and all other properties $\mathbf{X}$, such that $X = (X_{type}, \mathbf{X})$.

The traditional approach is to keep $\mathbf{X}$ 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:

$(X_{Fair}, \mathbf{X}_1) =$ "*The state should raise top tax rates because inequality leads to unfairness*",
$(X_{Ext}, \mathbf{X}_1) =$ "*The state should raise top tax rates because inequality leads to crime*",

where $\mathbf{X}_1$ 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 $\mathbf{X}$ 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 $\mathbf{X}$ constant means that the chosen stimuli are unlikely to be representative of their respective argument types.

The core difference in our approach is that we allow all other dimensions of the argument,

$\mathbf{X}$, 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 $X_{type} = X_T$, where T denotes the specific type. The assumption that allows extrapolation is that this sample is an unbiased draw from the distribution of arguments of type T.

We then let individuals j evaluate the argument to decide on an individual-specific outcome $Y_j = \phi_j(X)$, for example whether the argument is convincing to the individual; the evaluation is done through an individual-specific evaluation function ϕ_j. The second main assumption required for external validity is that the evaluation function ϕ_j captures the outcome of substantive interest—for example, that survey-measured convincingness reflects how persuasive an argument would be in a natural setting.[11] 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 $X_{type} = X_T$. Formally, we assess $\mathcal{G}(Y|X_{type} = X_T)$ for each X_T. This can be used to test hypotheses for $\mathcal{G}(Y|X_{type} = X_T)$ – whether one type of argument is more convincing than the other, for example.

We thus explore both (i) the content of each type of argument, $(\mathbf{X}|X_{type} = X_T)$, and (ii) individuals' reactions to each type of argument, $\mathcal{G}(Y|X_{type} = X_T)$. Any statistically significant effects from the resulting analysis can be interpreted as reflecting the argumentative environment, provided the sampling and evaluation assumptions hold.

We illustrate the sampling methodology in Figure 6. There 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. These two dimensions are chosen for illustration; in practice, arguments vary along many observable and unobservable dimensions simultaneously. In the left panels of Figure 6, we show the traditional approach. In the right panels of Figure 6 we show the approach in this paper, which compares randomly sampled arguments from an underlying distribution.

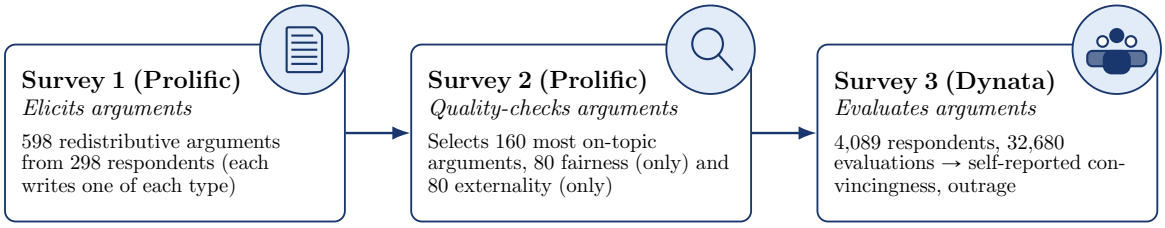
The main limitation of our approach is that randomly sampled arguments vary widely in content, which increases statistical noise relative to tightly controlled designs. This noise can be modeled empirically, however, so that any statistically significant differences across types – whether in content or evaluation – can be given an externally valid interpretation under the aforementioned assumptions. The methodology is also flexible, and can be adapted to any setting where arguments (or statements) vary within types. One could, for example, contrast economic versus cultural frames in immigration debates, or compare individual-risk arguments and population-spillover arguments on vaccination policy.

4.1.2 Data collection

The survey methodology follows a three-step process, which we illustrate in Figure 7. 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

[11] This is a standard assumption for all similar experiments; we mention it here due to the focus on external validity.

Figure 7: Overview of the Three-Survey Design



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$^{\text{rd}}$ 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.[12] We show the full question text in Figure 8; we note that the survey was conducted before the proliferation of LLMs (ChatGPT first launched November 30$^{\text{th}}$ 2022).[13] 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 discuss the methodology further 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. This reduces the sample to 160 pro-redistributive arguments, 80 of each type, on pre-specified criteria.[14] 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—one of fairness arguments and one of externality arguments—which correspond to $\hat{\mathcal{F}}(Y|X_{type})$ for each X_{Type}. We show the criteria and discuss the methodology in Appendix D.2.

Survey 3: Evaluating arguments Finally we collect *evaluations* of the Survey 1 arguments with a distinct sample of 4,089 respondents, collected with the professional survey company

[12] 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.

[13] 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.

[14] The study as a whole was pre-registered; we discuss differences from the pre-registration in Appendix D.5

Figure 8: 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 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 text and [brackets]. Differences in text across pro- and anti-redistributive argument elicitation are denoted by {brackets}. Figure D6 shows the information respondents were shown before the argument elicitation.

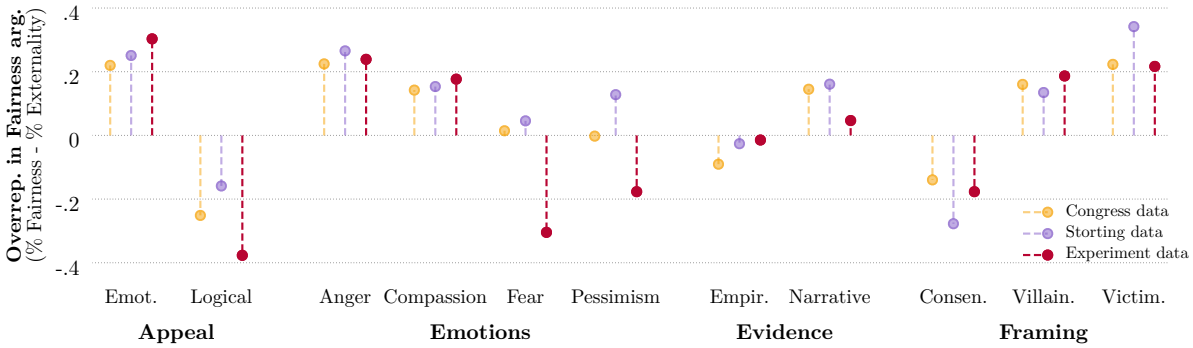
Dynata (following other similar studies, e.g. Lobeck and Støstad, 2023; Andre et al., 2023) and broadly representative on observables. Our main outcomes are self-reported measures of whether the respondent was convinced by the argument (which we define as "convincingness") and whether the respondent 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 of pro-redistributive arguments (80 of each type). We collect a total of 32,680 evaluations for these arguments from 4,089 respondents, at an average of 202 evaluations per argument. Each respondent evaluated 10 arguments, including both pro- and anti-redistributive arguments; the inclusion of anti-redistributive arguments avoids signaling to respondents that all arguments favor redistribution. We restrict our analysis to the pro-redistributive evaluations.

Representativity The Survey 1 composition is largely Democratic; we intentionally oversampled Democrats given the pro-redistributive focus of the study.[15] We did not enforce representativity in Survey 2 given the quality-checking motive of the survey. Survey 3, which collects the argument evaluations, was conducted with the professional survey company *Dynata* using a sample that is broadly representative of the U.S. population on observables, with pre-specified quotas on gender, age, political affiliation, income, region, and race (details in Appendix D.4). Participation in online access panels is inherently selective, which prevents an interpretation of full representativity; still, related studies using online panels report broadly comparable results to nationally representative benchmarks (Stantcheva, 2023; Haaland et al., 2023).

[15]Importantly, there is no significant difference in political affiliation across argument types, so partisan composition does not confound the fairness-externality comparison.

Figure 9: 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 4 for the experimental data). Appendix B shows the full prompt texts.

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).

4.2 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.[16]

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 9. 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. 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. 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. We consider this strong evidence that the content differences between argument types are not artifacts of a particular setting, but stable properties likely to generalize beyond the contexts we study. This establishes

[16] We discuss the average content in the sample in Appendix E.1.

the third main result of the paper:

Result #3

Fairness arguments and inequality externality arguments are two distinct ways to argue for redistribution. Across legislative and experimental settings, fairness arguments are more likely to use emotional appeals, anger, compassion, narrative-based evidence, victimization, and villainization. Externality arguments are more likely to use logical appeals, empirical evidence, and consensus-seeking language.

Having established that fairness and externality arguments systematically differ in cross-country prevalence and content, we now turn to how citizens receive them.

4.3 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 through a sample of 4,089 U.S. survey respondents.

4.3.1 Descriptive statistics

Overall, 54% of evaluations find a given pro-redistributive argument convincing. Across political parties, 40% of Republicans find a given pro-redistributive argument convincing, as compared to 68% of Democrats. 34% report that a conversation with the author could provoke anger because the respondent agrees with the argument ("outrage"), with corresponding shares of 26% among Republicans and 43% among Democrats. 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. 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 G5, and the three most and least outrage-inducing arguments in Supplementary Table G6.

4.3.2 Empirical specification

Our coefficient of interest is β in the following regression:

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

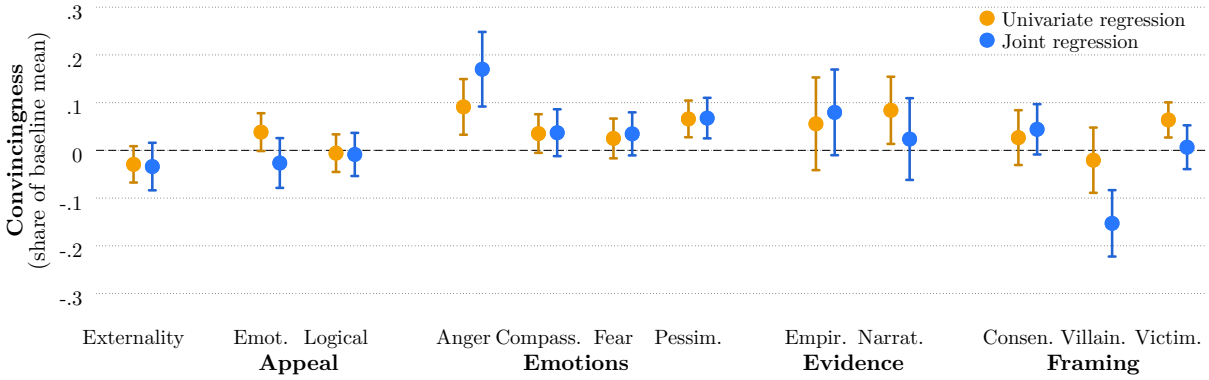
where $y_{i,j}$ is the outcome of interest (e.g. convincingness) for individual i and argument j, γ_i is a person fixed effect, α_j is the argument characteristic of interest (e.g. a dummy equal to one if the argument is an externality argument), and standard errors $\varepsilon_{i,j}$ are clustered 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.

4.3.3 Convincingness

We show how different argument characteristics relate to convincingness in Figure 10. 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). These confidence bands are relatively narrow; as a point of comparison, many short information treatments find effect sizes on redistributive preferences of 0.1-0.2 standard deviations (Stantcheva, 2021; Lobeck and Støstad, 2023).

We can also examine how convincingness varies with other argument characteristics, although these associations should not be given a causal interpretation, since characteristics such as anger, pessimism, narrative evidence, and villainization are not randomly assigned across arguments. 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 10: Association Between Convincingness and Argument Characteristics

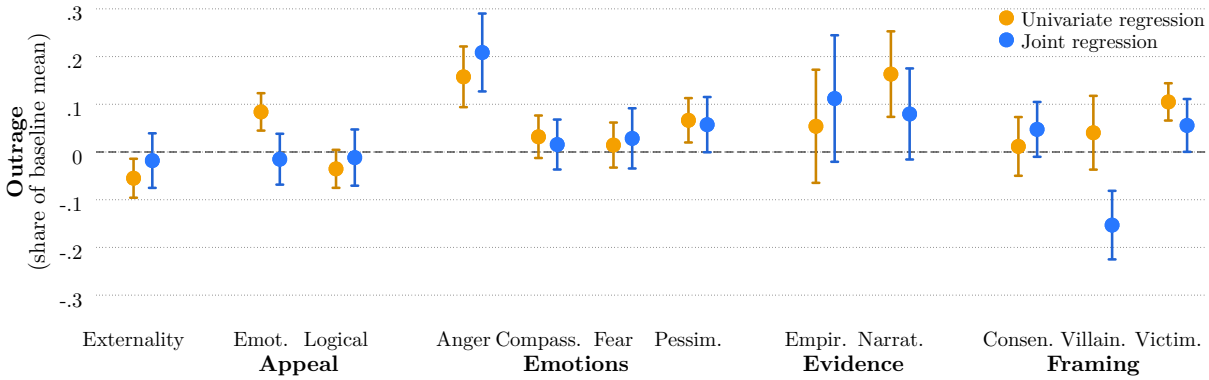


Note. This figure shows the result of regressing convincingness on argument characteristics. The Convincingness variable is first divided by its mean (.537), so coefficients should be interpreted as a share of the mean. The yellow dots correspond to a regression of convincingness on the indicated variable, 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.

4.3.4 Outrage

We show how different argument characteristics relate to outrage in Figure 11. 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*

Figure 11: 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.

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 appears to be driven by the underlying anger content of fairness arguments; the association is not significant in the joint regression, nor when only controlling for anger content.[17] This illustrates a key advantage of our current approach: in a conventional experiment that holds content fixed across types, this effect would be undetectable.

This establishes our fourth main result:

> Result #4
>
> In a controlled survey experiment, we cannot reject the null that fairness and externality arguments are equally convincing. Fairness arguments are more likely to elicit outrage.

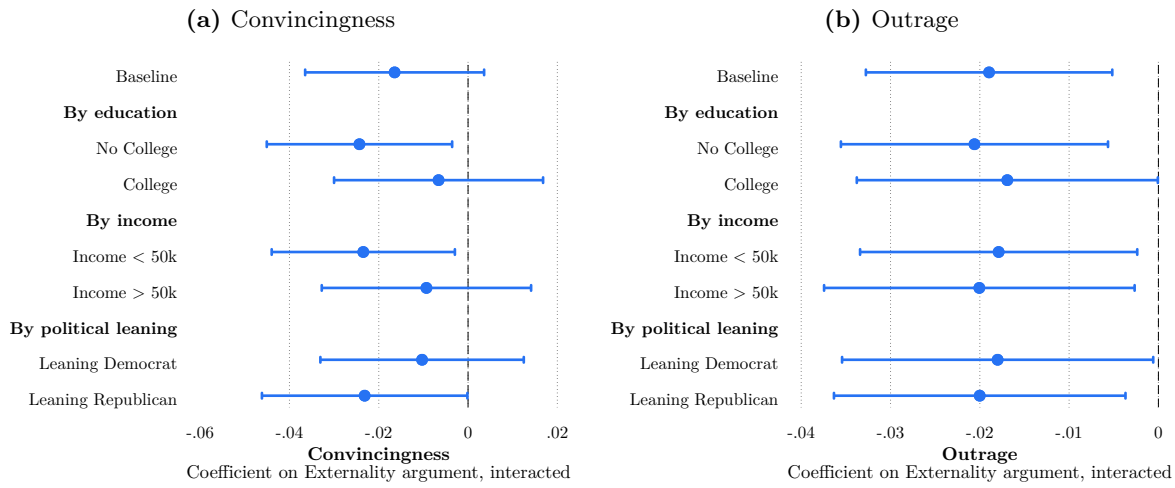
The equal convincingness of both argument types is notable: it suggests that the dominance of fairness arguments in the U.S. Congress does not reflect externality arguments being inherently unpersuasive. The outrage difference, however, points to a potentially important consequence of the argument mix, as a debate dominated by fairness arguments may generate substantially more affective polarization.

4.3.5 Heterogeneity across groups

Are fairness or inequality externality arguments more convincing for some groups than others? Figure 12a 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. Because of limited statistical power, we can-

[17] As expected, anger content is strongly associated with outrage in both separate and joint regressions. Emotional appeals, victimization, and narrative evidence are also positively associated with outrage in the bivariate specification, while villainization is negatively associated with outrage in the joint regression (likely reflecting that villainization reduces agreement with the argument).

Figure 12: 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 12a, 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.

not cleanly disentangle these two patterns. Still, the education gradient is consistent with the legislative evidence – where representatives from more highly educated districts are more likely to use externality arguments – and with the greater reliance of externality arguments on logical appeals and empirical evidence. Taken together, the results point to an educational divide. This establishes our fifth and final main result:

Result #5

We document an educational gradient: across analyses, fairness arguments are relatively more prominent and persuasive in lower-education settings.

This finding suggests an educational sorting in redistributive debate, both in argument supply and in argument demand: more educated actors are more likely to use inequality externality arguments, while less educated audiences are more convinced by fairness arguments.

We also find suggestive heterogeneity along other dimensions. Fairness arguments are relatively more convincing for respondents with annual incomes below $50,000 and for Republican-leaning respondents. These patterns 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 respondents. We show the limited heterogeneity across groups for outrage in Figure 12b.

4.4 Robustness: Validation of self-reported measures

Our main experimental outcomes are self-reported convincingness and outrage.[18] If these survey responses are themselves our outcomes of interest—rather than proxies for deeper preferences—the evaluation assumption (Section 4.1.1) is trivially satisfied: we are directly observing how respondents evaluate arguments. To allow for a stronger interpretation, we examine whether our measures predict movement in a separate post-treatment measure of support for redistribution. Exposure to more pro-redistributive arguments increases support for redistribution across the survey, with positive effects for both argument types and no significant difference between them. Arguments rated as more convincing generate larger shifts in redistributive preferences, while arguments rated as more outrage-inducing do not. A leave-one-out specification using other respondents' evaluations yields similar magnitudes but is imprecisely estimated. We report the full analysis, which validates our use of self-reported measures, in Appendix E.5.

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, consistent with the broader educational gradient documented above, fairness arguments are relatively more persuasive among lower-education respondents.

We now turn to the robustness of the LLM methodology.

5 Robustness of the LLM methodology

Recent work shows that modern LLMs can classify open-ended text with accuracy that approaches, and sometimes exceeds, human benchmarks (Heseltine and Clemm von Hohenberg, 2024; Le Mens and Gallego, 2025; Soria, 2025). Following the best-practice guidelines of Yang et al. (2025), we subject our LLM classifications to a broad set of validation and robustness tests.

5.1 Survey Experiment: Robustness of LLM classification

Our primary robustness check asks whether the LLM classification procedure used for the Congressional speech excerpts also recovers the argument types of the elicited and human-screened arguments from Surveys 1 and 2. Using the exact prompts described in Section 3.1, we classify each of the final 160 arguments with GPT-4o Mini in two independent runs. Each argument is then assigned to one of four categories: fairness, inequality externality, both, or neither.

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. 89% of the fairness arguments are correctly classified as fairness and *not* externality arguments, and 82% of the 80 externality arguments are correctly classified as externality and *not* fairness arguments. Overall, 86% are correctly classified twice.

[18] Self-reported measures are common in survey experiments on redistribution and tax policy, as such measures consistently track patterns in incentive-based and causal validation exercises (e.g. Haaland et al., 2023; Stantcheva, 2023).

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, however, they likely to attenuate the measured differences in content between fairness and externality arguments, implying that our difference estimates should be viewed as conservative.

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) employing an alternative prompt that defines excerpts as including primarily fairness arguments, primarily externality arguments, or neither (Appendix B.18).

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 closer to the experimental findings discussed in Section 4.

5.3 Alternative prompts

To limit researcher discretion in prompt selection, we report our main results using the first prompt for which the full analysis was completed. We then assess robustness to prompt wording by reclassifying fairness-based and externality-based arguments in the legislative data using three alternative prompts for each argument type. These prompts vary the wording while keeping the underlying conceptual content fixed. The full text of the three additional prompts, which were also the first three alternatives we tested, is reported in Appendix B.19–B.24.

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. Content analysis also 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

Throughout the paper, we use GPT-4o Mini for all classifications, as it was the first model with which we completed the full analysis. In this section we use two alternative LLMs: GPT OSS 120B, an open-weight model with 117 billion parameters that OpenAI describes as its most powerful open-weight model; and the full DeepSeek R1 model, a mixture-of-experts model with 671 billion parameters. Both models use the exact same prompts as our main specification. We also report a more conservative specification that restricts the sample to observations on which all three LLMs agree.

The stronger prevalence of externality arguments in Norway relative to the U.S. is stable across LLMs. The Norwegian externality-to-fairness ratio is 2.3 times higher than that of the U.S. in our main specification using GPT-4o Mini; it is 1.3 times higher with GPT OSS 120B, and 2 times higher with DeepSeek R1. It is 2.7 higher if we only include arguments on which all three LLMs agree. Content patterns are very similar across LLMs (Figure F8); for almost all characteristics, the sign of the fairness-externality gap is unchanged.

5.5 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).

5.6 Word search

As a simple benchmark, Appendix C.7 complements the LLM classifications with keyword searches. First, the keywords support the basic classification: across samples, fairness arguments use language of distributive justice and deservingness, while externality arguments use language about social and economic consequences. Second, even these simple word counts recover the main content contrast within the classified arguments, as fairness arguments are more likely to contain anger-related language. At the same time, the limited coverage of keyword searches underscores the value of the LLM-based approach.

5.7 Language of input text

We test the robustness of our main result to the language of the input text. We start by translating a subset of U.S. Congress excerpts from English to Norwegian, and of Norwegian *Storting* excerpts from Norwegian to English with GPT OSS 120B[19]. We then run the main fairness and externality prompts on the translated arguments, using GPT 4o-Mini.

We compare arguments within language. Comparing all U.S. Congress excerpts to the translated subset of Norwegian *Storting* excerpts yields a 2.2 larger externality-to-fairness ratio in

[19]Translation is a very time-consuming procedure, which motivates our choice of running this robustness check on a subset of arguments.

the *Storting data*, compared to a 2.3 larger ratio in the original language in that specific subsample. Comparing all Norwegian *Storting* excerpts to the translated Congress subset yields a 1.5 larger externality-to-fairness ratio in the *Storting* data, compared with a 2.3 larger ratio in that subsample. We attribute this slightly larger discrepancy to a likely worse performance of the LLM when translating into Norwegian than into English.

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

6 Conclusion

This paper has examined how citizens and legislators argue about redistribution, and how these arguments are received. We first established that Norway's redistributive debate is more pro-redistributive than that of the United States, while anti-redistributive arguments in both countries rely primarily on efficiency claims. On the pro-redistributive side, fairness arguments are more likely to be emotional, particularly containing anger and compassion, while externality arguments appeal more to logic and seek consensus. Externality arguments are much more common in the Norwegian *Storting* than in the U.S. Congress, and while both argument types 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 large-scale empirical analysis of the contemporary redistributive debate.

Overall, these findings establish two distinct ways to argue for redistribution. While the fairness register is well-known, the pragmatic externality approach may appear more novel, especially in countries where the fairness angle is prominent. The externality angle shifts the redistributive debate from its usual equity–efficiency framing toward a more strictly consequentialist register. Their relative prevalence is likely to affect societal outcomes; it is natural to think that fairness arguments lead to more affective polarization, for example. Externality-based debates, on the other hand, may lend themselves to empirical evidence and scientific disagreements. The balance of these rhetorical modes may have shaped not only the tone of public debate, but also the long-run trajectory of redistributive policy.

An open question is whether these differences help explain cross-country variation in redistributive policy. Differences in fairness views are often suggested as the key driver for such variation (e.g. Alesina and Angeletos, 2005; Almås et al., 2020), but our findings point to a complementary channel: countries may differ not only in how fair citizens consider the economic system, but in how prominently the societal consequences of inequality feature in public debate. Whether the prevalence of externality arguments reflects, reinforces, or independently shapes redistributive institutions is a question we cannot answer here — but the magnitude of the cross-country differences we document suggests it deserves further investigation.

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 and George-Marios Angeletos. Fairness and Redistribution: US vs. Europe. *American Economic Review*, 95(5):960–980, 2005.

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.

Asirvatham, Hemanth, Elliott Mokski, and Andrei Shleifer. GPT as a Measurement Tool. Working Paper 34834, National Bureau of Economic Research, 2026.

Boxell, Levi, Matthew Gentzkow, and Jesse M Shapiro. Cross-country Trends in Affective Polarization. *Review of Economics and Statistics*, 106(2):557–565, 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.

Chong, Dennis and James N Druckman. Framing Public Opinion in Competitive Democracies. *American Political Science Review*, 101(4):637–655, 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.

Fong, Christina. Social Preferences, Self-Interest, and the Demand for Redistribution. *Journal of Public Economics*, 82(2):225–246, 2001.

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

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

Gentzkow, Matthew, Jesse M. Shapiro, and Matt Taddy. Measuring group differences in high-dimensional choices: Method and application to congressional speech. *Econometrica : journal of the Econometric Society*, 87(4):1307–1340, 2019.

Haaland, Ingar, Christopher Roth, and Johannes Wohlfart. Designing Information Provision Experiments. *Journal of Economic Literature*, 61(1):3–40, 2023.

Haidt, Jonathan and Jesse Graham. When Morality Opposes Justice: Conservatives have Moral Intuitions that Liberals may not Recognize. *Social Justice Research*, 20(1):98–116, 2007.

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.

Hilmar, Till and Patrick Sachweh. "Poison to the Economy":(Un-) Taxing the Wealthy in the German Federal Parliament from 1996 to 2016. *Social Justice Research*, 35(4):462–489, 2022.

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.

Jones, Michael D and Mark K McBeth. A Narrative Policy Framework: Clear Enough to be Wrong? *Policy Studies Journal*, 38(2):329–353, 2010.

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.

Mankiw, N Gregory. Spreading the Wealth Around: Reflections Inspired by Joe the Plumber. Technical report, National Bureau of Economic Research, 2010.

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.

Nozick, Robert. *Anarchy, State, and Utopia*. Basic Books, New York, 1974.

Okun, Arthur M. *Equality and Efficiency: The Big Tradeoff*. Brookings Institution Press, 1975.

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.

Rothstein, Bo. *Just Institutions Matter: The Moral and Political Logic of the Universal Welfare State*. Cambridge university press, 1998.

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. Using Elasticities to Derive Optimal Income Tax Rates. *The Review of Economic Studies*, 68(1):205–229, 2001.

Saez, Emmanuel and Stefanie Stantcheva. Generalized Social Marginal Welfare Weights for Optimal Tax Theory. *American Economic Review*, 106(1):24–45, 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.

Sejersted, Francis. *The Age of Social Democracy: Norway and Sweden in the Twentieth Century*. Princeton University Press, 2011.

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. How to Run Surveys: A Guide to Creating Your Own Identifying Variation and Revealing the Invisible. *Annual Review of Economics*, 15(1):205–234, 2023.

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.

Weinzierl, Matthew. The Promise of Positive Optimal Taxation: Normative Diversity and a Role for Equal Sacrifice. *Journal of Public Economics*, 118:128–142, 2014.

Wilkinson, Richard and Kate Pickett. *The Spirit Level: Why More Equal Societies Almost Always Do Better*. Allen Lane, 2009.

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

A main focus in this paper is the distinction between 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. 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,

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

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 first 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.

As we discuss in Section 2, these distinctions have concrete policy implications. In standard fairness-based optimal taxation, higher taxes on top earners involve a trade-off between raising revenue and discouraging effort (Saez, 2001). Under an externality rationale, the relevant trade-off differs: if high incomes generate negative externalities, then the behavioral response of top earners reducing their income is itself partly beneficial, as it shifts top-income agents to leisure

and thus reduces the associated externality of their incomes (Støstad and Cowell, 2024). As such, both the behavioral and mechanical effect have positive welfare effects under an externality motivation, but not under a fairness motivation. We make one further note, which is that negative social welfare weights (a fairness idea) also lead to benefits from reducing top incomes. While such negative weights are generally avoided in optimal policy – they are assumed away in Saez and Stantcheva (2016), for example – such weights also imply different policy goals given that negative social welfare weights aim to lower utility directly, while externality-based motivations aim to shift top-income agents towards leisure while being indifferent about their net utility.

A large share of pro-redistributive arguments can be cast in one of these forms. 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*"). A similar asymmetry arises on the anti-redistributive side, where fairness-based objections to redistribution and efficiency-based objections imply different optimal policy responses (Weinzierl, 2014; Saez and Stantcheva, 2016).

Fairness subtypes 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.

Externality subtypes 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.

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 approximately 5 million separate API calls. 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. The LLM has no further context or memory of previous classifications.

The following outcomes were classified:

1. **In favor of redistribution:** Whether the excerpt includes an argument in favor of increasing economic redistribution.
2. **Against redistribution:** Whether the excerpt includes an argument in favor of decreasing economic redistribution.

These two classifications were made on the full sample of 750,544 (U.S. Congress) and 258,490 (Norwegian *Storting*) speech excerpts; all other classifications were made on the subsample of pro- or anti-redistributive excerpts.

3. **Fairness argument (pro-redistributive):** Whether the excerpt includes an argument that frames economic inequality in terms of fairness or justice
4. **Inequality externality argument (pro-redistributive):** Whether the excerpt includes an argument that frames economic inequality as a harmful externality
5. **Fairness argument (anti-redistributive):** Whether the excerpt includes an argument that frames economic inequality in terms of fairness or justice
6. **Efficiency argument (anti-redistributive):** Whether the excerpt includes an argument that frames economic redistribution as inefficient
7. **Emotional appeal:** Whether the excerpt appeals to emotion
8. **Logical appeal:** Whether the excerpt appeals to logic, reasoning, or factual evidence
9. **Anger:** Whether the excerpt contains anger
10. **Compassion:** Whether the excerpt contains compassion
11. **Fear:** Whether the excerpt contains fear or concern
12. **Pessimism:** Whether the excerpt contains pessimism
13. **Empirical evidence:** Whether the excerpt contains empirical evidence
14. **Narrative evidence:** Whether the excerpt contains narrative evidence
15. **Consensus-seeking:** Whether the excerpt is consensus-seeking
16. **Villainizing:** Whether the excerpt villainizes any specific group
17. **Victimizing:** Whether the excerpt victimizes any specific group
18. **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
19. **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 substance but make the language vary
20. **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.5)
21. **Externality argument subtypes:** Whether the excerpt includes a top-income or a bottom-income externality argument (see Appendix C.5)

Below we list all full prompts.

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

```
You are a classifier.

You are analyzing a speech in [language] 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: Against redistribution

```
You are analyzing a speech in [language] to determine whether
    it includes an argument **against the idea of increasing
   economic redistribution**, or **in favor of decreasing
   economic redistribution**.
Give this text a score of either:
- 1 = The text **does** make an argument against economic
   redistribution/in favor of decreasing redistribution (e.g.,
    advocating for lower taxes on the wealthy, supporting tax
   breaks on big corporations, restricting social programs,
   decreasing the minimum wage, redistributing income or
   wealth less, etc.)
- 0 = The text **does NOT** make such an argument (e.g., it's
    neutral, unrelated, takes a stance *against* policies such
    as tax breaks that benefit the wealthy, corporations, or
   groups such as 'the top 1%', criticises policies that hurt
   the poor or decrease the budget of public programs such as
   education and health, or argues in favor of more
   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".
```

B.3 Prompt to GPT-4o Mini: Fairness argument (pro)

User prompt:

```
[Insert argument]
```

System prompt:

```
You are analyzing a speech in [language] 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.4 Prompt to GPT-4o Mini: Inequality externality argument (pro)

User prompt:

```
[Insert argument]
```

System prompt:

```
You are analyzing a speech in [language] 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.5 Prompt to GPT-4o Mini: Fairness argument (anti)

User prompt:

```
[Insert argument]
```

System prompt:

```
You are analyzing a speech in [language] to determine whether it
    includes an argument against reducing inequality, or against
    redistribution, that is framed around fairness or justice --
    specifically, discussing redistribution, taxation, or transfers
     as unfair, unjust, undeserved, coercive, discriminatory, or
    morally wrong, or invoking related ideas such as procedural
    fairness, property rights, what people have earned, equal
    treatment, duties to workers or taxpayers, or the claim that
    government rewards the undeserving at the expense of the
    deserving. These arguments must go beyond discussing the
    effects on broader societal outcomes (e.g., democracy, trust,
    stability, capitalism, growth, cohesion, the integrity of
    public institutions, innovation), and instead emphasize or
    imply principles of fairness, justice, rights, legitimacy, or
    morality.

Give this text a score of either:
1 = The text does make an argument against redistribution based
    on fairness or justice -- emphasizing that people are entitled
    to their outcomes, that redistribution or state action is
    unfair, or otherwise appealing to justice, rights, legitimacy,
    or moral principles to argue against redistribution.
0 = The text does NOT make such an argument (e.g., it is
    unrelated, or focuses only on the societal consequences,
    efficiency, practical outcomes, without invoking fairness).

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.6 Prompt to GPT-4o Mini: Efficiency argument (anti)

User prompt:

```
[Insert argument]
```

System prompt:

```
You are analyzing a speech in [language] to determine whether it
   includes an argument against reducing inequality, or against
   redistribution, that frames redistribution, taxation, transfers
   , or associated government expansion as economically harmful or
    inefficient -- specifically, discussing such policies as
   reducing growth, jobs, wages, investment, productivity,
   competitiveness, consumer welfare, or fiscal sustainability, or
    as creating broader harms such as inflation, debt, deficits,
   dependency, waste, bureaucracy, red tape, misallocation, or
   burdens on businesses, workers, consumers, or taxpayers. These
   arguments must go beyond discussing fairness, and instead
   emphasize or imply the effects on broader societal outcomes (e.
   g., growth, investment, productivity, employment, wages,
   competitiveness, fiscal sustainability, government efficiency,
   or economic dynamism).

Give this text a score of either:
1 = The text does make an argument based on redistribution
   causing economic inefficiencies or harmful incentive effects --
    emphasizing that a redistributive policy harms growth or other
    socially valuable outcomes, or that limiting redistribution
   improves broader societal outcomes.
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 broader economic
    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.7 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.8 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.9 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.10 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.11 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.12 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.13 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.14 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.15 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.16 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.17 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.18 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.19 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.20 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.21 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.22 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.23 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.24 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.25 Prompt to GPT-4o Mini: Fairness argument subtypes (Needs-based, Rights-based, Compensatory)

User prompt:

```
[Insert argument]
```

System prompt:

```
You are analyzing a speech in [language] 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.26 Prompt to GPT-4o Mini: Externality argument subtypes (Top-income, Bottom-income)

User prompt:

```
[Insert argument]
```

System prompt:

```
You are analyzing a speech in [language] 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.3-B.4). 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.3-B.4). 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 F9.

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.3). Our main specification to explore the amount of fairness arguments.

Ext. (All) Classified as including an externality argument (prompt in Appendix B.4). 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.18 to classify each pro-redistributive argument as either:

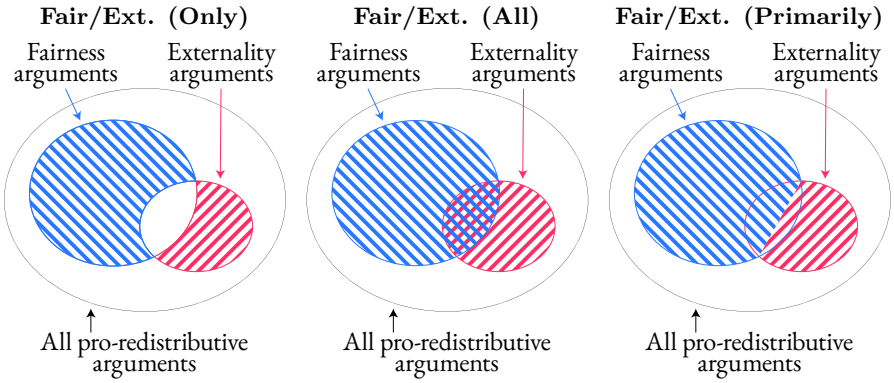
Fair. (Primarily) Classified as being primarily a fairness argument.

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 externality-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	4.8
Ext. (Only)	Fair. (Only)	2.3
Ext. (Only)	All pro-redistributive speeches	1.8
Ext. (Only)	All speeches	2.9
Ext. (Primarily)	Fair. (Primarily)	3.2
Ext. (Primarily)	All pro-redistributive speeches	2.9
Ext. (Primarily)	All speeches	4.7

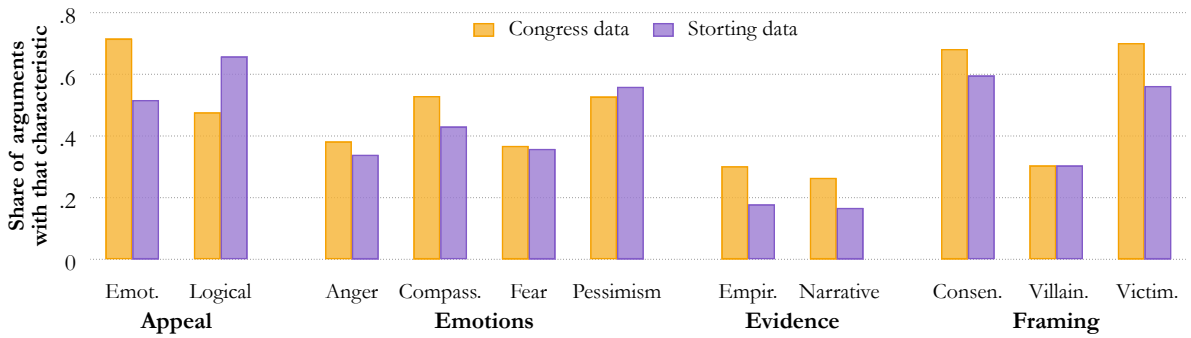
Note. This table shows the relative use of externality-based argument in Norway compared to the U.S., varying the definition of externality-based argument use. In each row, the NO/US ratio is computed as $(\text{num}_{NO}/\text{denom}_{NO})/(\text{num}_{US}/\text{denom}_{US})$. The definitions of Fair./Ext. 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.

Figure C2: Average Argument Characteristic in the U.S. Congress and Norwegian *Storting*



Note. This figure displays, for each content characteristic, the share of pro-redistributive excerpts classified as containing this characteristic in the U.S. Congress and Norwegian *Storting* samples. All excerpts classified as containing a pro-redistributive argument are included. Figure C3 compares the content of pro-redistributive and non-redistributive speech excerpts. Appendix B shows the full prompt texts.

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 Cross-country content differences in the pro-redistributive debate as a whole

Figure C2 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 C3). Argument length is similar by construction, while average word length is slightly higher in the *Storting*.

The cross-country content differences raise the question of whether the redistributive argument mix explains part of the difference. We show the effect of constructing counterfactuals—reweighting each country's redistributive debates using the other's mix of fairness and externality arguments—in Appendix C.4. 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.

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).

C.3 Congress excerpts: Examples (supplement to Section 3.3)

Fairness Only: Ron Wyden, Democrat My goodness, that is what America is about. But we also want everybody to pay their fair share. And I am really pleased that the President has been talking about a billionaire minimum income tax. We have been coming together all in the name of fairness. And Mike looks specifically at how to do that and address the parliamentary and legal concerns. I would be on the floor until breakfast time tomorrow and I see our friend from Mississippi has something that is important to him and I just want to close with one last point. You know, I think if you walked into a coffee shop in Maine or Oregon or Mississippi and you said: "Well, let me talk to you a little bit about reconciliation," people would be a little baffled with you, but maybe they would say, I hope...

Externality Only: Ro Khanna, Democrat Is that money going to create jobs in the United States? Or could that money be spent anywhere in the world, sheltered anywhere in the world? As opposed to if you give that money to someone making $50,000, $60,000, they are going to spend that money in their local community; they are going to buy more groceries; they are going to buy more things for their house; they are going to get more education; that money is going to go into the local economy, and it is going to actually create jobs. That is what is going to grow the economy. That is our only shot of getting 3 percent economic growth if we invest in actual workers, the people doing the work.

Both: Darren Soto, Democrat ...that aren't making a big difference in this world because when you have the rich control all capital, all political power, that is when they control us, and we don't have the innovation. We don't have the incentive. We don't have the progress that is so critical to capitalism, which I think is being missed on this. If you don't have a fair market, you can't have successful capitalism, and part of a fair market includes having a strong voice for our middle class, for our working folks, through our unions. It is no surprise that a rise to greatness in this country was tied to the rise of the middle class. Think about it: GIs returning home from World War II, fanning out to the suburbs. Even before that, around World War I and before that, when you had all of these major milestones that we talked about a...

Neither 1: Terri Sewell, Democrat Mr. Speaker, I rise today in support of H.R. 2901, to make permanent the IRS Volunteer Income Tax Assistance (VITA) matching grant program. The Volunteer Income Tax Assistance (VITA) program assists low-to- moderate income, elderly, disabled, and limited English speaking taxpayers in preparing and filing their federal income tax returns. Since fiscal year 2008, Congress has used the appropriations process to provide funding for the Community VITA matching grants program for tax return preparation assistance. But Congress has not authorized the VITA program to receive grants through the tax-writing process. This bill puts the VITA grant program on more solid, permanent, and predictable footing.

Neither 2: Richard E. Neal, Democrat Madam Speaker, I yield 2 minutes to the gentlewoman from Wisconsin (Ms. Moore), who is always the great champion of expanding the earned income tax credit.

Neither 3: Timothy Kaine, Democrat ...the environment such as the American Battlefield Protection Act, Chesapeake Bay Program, and the Army Corps programs in Norfolk; strong funding for the National Park Service and for NASA's programs at Wallops Island; and strong funding for Plan Central America. This bill also includes critically important programs on the revenue side. Three critical low- and middle-income tax programs the child tax credit, earned income tax credit, and American opportunity tax credit have been made permanent in this bill, so has the research and development tax credit, along with an expansion in this credit for startups championed by Senator Coons that I have cosponsored. Also made permanent are tax programs for teachers, for conservation, and for military families. We have made other programs last for another 5 years. And others will be extended for 2 years, a step forward for these programs we have been extending for only 1 year at...

C.4 Applying counterfactual 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.[20]

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.

[20]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: Counterfactuals 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 Norwegian *Storting* data, as detailed in Section C.4. Results are based on speeches containing only fairness or only externality arguments in each dataset.

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.

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 more 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 C3). 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.

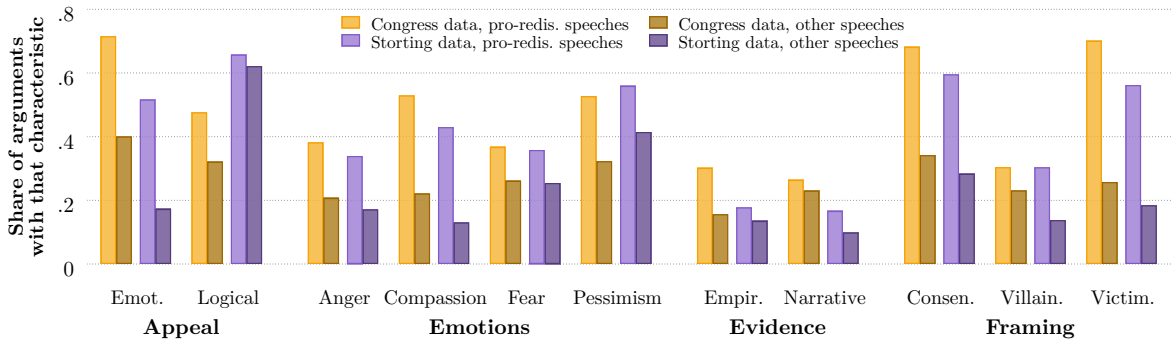
Overall, the counterfactuals confirm that differences in argument-type composition contribute to the observed divergence between Norwegian and U.S. redistributive debates, but they do not account for the majority of it—which is unsurprising, given that legislative speech across the countries differ across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

Table C3: Counterfactuals 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.4 supra.

Figure C3: 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.5 Types of fairness and inequality externality arguments

Here we discuss the classification of different types of fairness and inequality externality arguments. The theoretical background of these categories is discussed in Appendix A.

Methodology: Types of fairness arguments With a single prompt we classify each fairness argument into one of those subtypes (prompt in Appendix B.25):

- **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 has none of the above themes.

Methodology: Types of inequality externality arguments With a single prompt we classify each externality argument into one of those subtypes (prompt in Appendix B.26):

- **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 has none of the above themes.

Results We show the resulting classifications in Figure C4. 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.

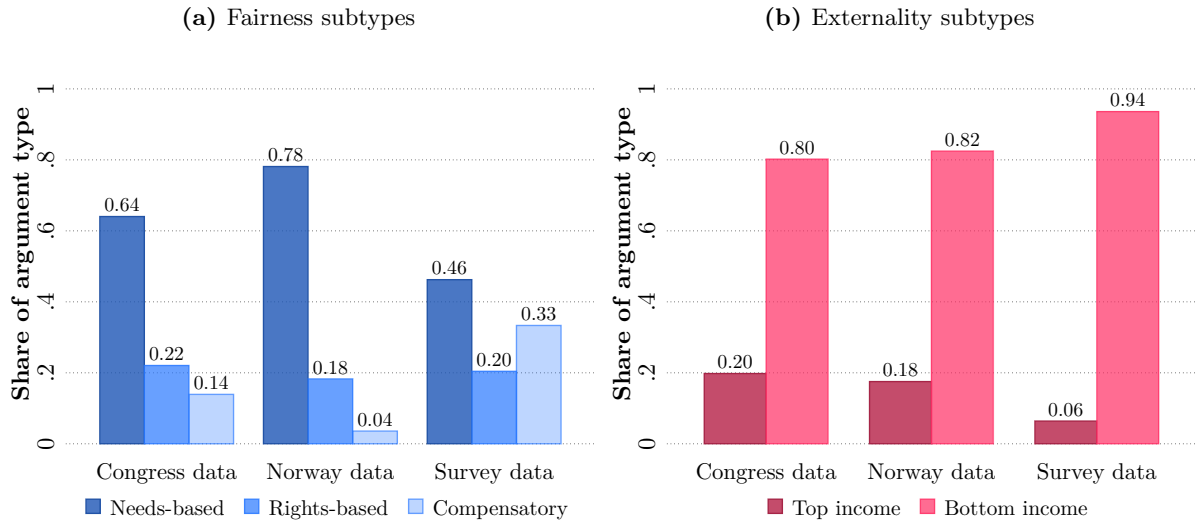
Figure C5 compares content in the U.S. Congress across these subtypes. The three fairness types are broadly similar, though needs-based arguments exhibit somewhat stronger emotional elements—especially anger, compassion, and villainizing language. The two externality types differ more sharply: most of 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.6 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

Figure C4: 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.5 supra. The prompts used for this classification are in Appendices B.26-B.25. Standard errors are robust, 95% CIs.

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 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

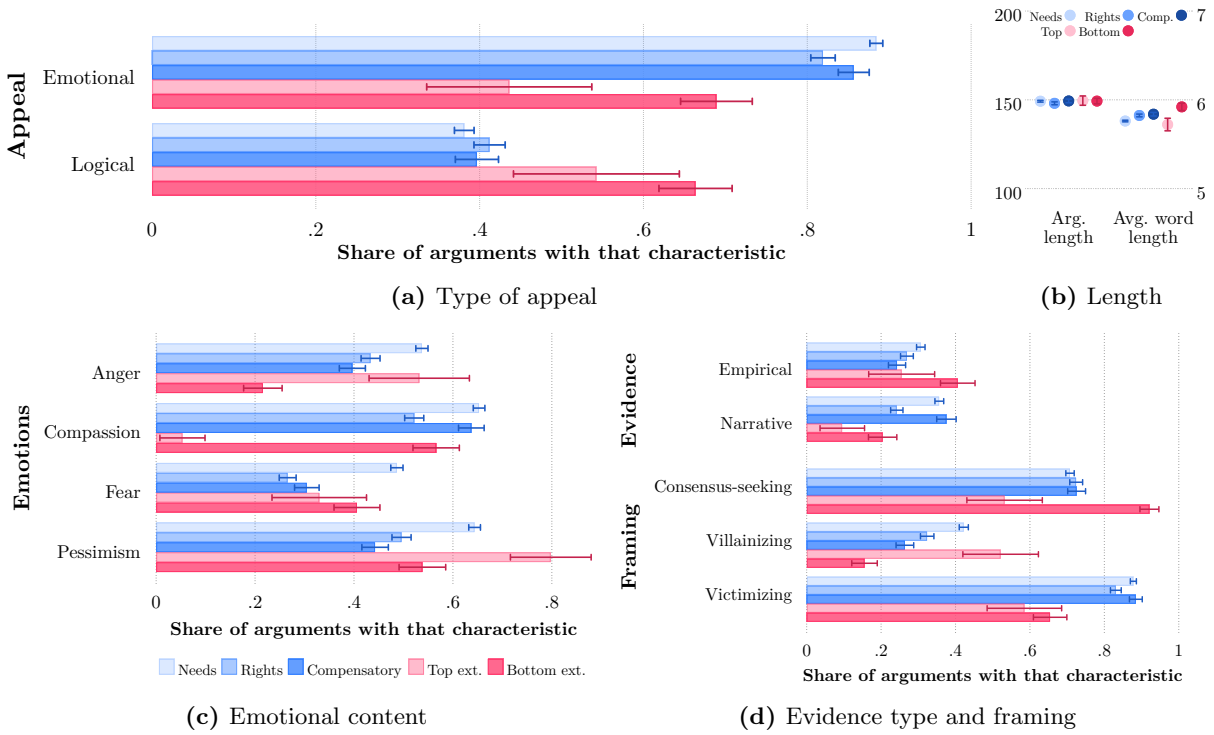
$$r'_j = \log\left(\frac{n_j^F + \alpha}{N^F - n_j^F + (W-1)\alpha}\right) - \log\left(\frac{n_j^E + \alpha}{N^E - n_j^E + (W-1)\alpha}\right).$$

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 $z_j = r'_j/\sigma_j$.

C.7 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

Figure C5: 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 4, 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.5 supra. The prompts used for this classification are in Appendices B.26-B.25. Standard errors are robust, 95% CIs.

characteristic of fairness and externality arguments in the U.S. data: that is, the words most overrepresented in fairness (resp. externality) arguments, compared to their frequency in the corpus overall. The text processing methodology is detailed in Appendix C.6. 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

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

(a) Fairness arguments

	Frequency in: Fairness arguments	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	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 5.6, 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.6. The log odds ratio z-score measures the significance of the relative overrepresentation of that word among fairness (resp. externality) arguments.

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.

C.8 Robustness: Language of text

We test the robustness of out results to translating the parliamentary excerpts into the other country's language. The translation and classification are done with two different LLMs, to make sure our results aren't due to a specific model's features.

For translation, we use GPT OSS 120B and run the following prompt: "Translate the following text into [English/Norwegian], keeping the substance of the text unchanged. Return only the [English/Norwegian] translation." Translation is a much more time-consuming process than classification, as the output is more complex—we therefore limit this robustness check to a subsample of 2,409 arguments from the U.S. Congress and 10,547 arguments from the Norwegian *Storting* data.

We then run the main Fairness and Externality prompt on the translated arguments, and compare externality-to-fairness ratios within language. Comparing all U.S. Congress excerpts to the translated subset of Norwegian *Storting* excerpts yields a 2.2 larger externality-to-fairness ratio in the *Storting data*, compared to a 2.3 larger ratio in the original language in that specific subsample.

The externality-to-fairness ratio is .24 in the original U.S. Congress data; it is .55 for the

Table C5: Words most characteristic of each argument type in *Storting* 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 5.6, showing the 20 words most characteristic of *Fair. (Only)* and *Ext. (Only)* arguments in the Norwegian *Storting* data, according to their log odds ratio z-score. The methodology is detailed in Section C.6. The log odds ratio z-score measures the significance of the relative overrepresentation of that word among fairness (resp. externality) arguments.

Norwegian excerpts translated into English, and .57 for those same arguments in their original language. This yields a 2.2 larger externality-to-fairness ratio in the *Storting* data, compared with a 2.3 larger ratio in the untranslated subsample.

The externality-to-fairness ratio is .57 in the original Norwegian *Storting* data; it is .37 for the U.S. excerpts translated into Norwegian, and .24 for those same arguments in their original language. This yields a 1.5 larger externality-to-fairness ratio in the *Storting* data, compared with a 2.3 larger ratio in the untranslated subsample. We attribute this slightly larger discrepancy to a likely worse performance of the LLM when translating into Norwegian than into English.

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 5.6, 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.6. 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 D6.

Figure D6: 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 8 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.[21] 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

[21] 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 Appendix E.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 Appendix E.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.

D.4 Representativity

We show the representativity of each survey in Table D7.

In Survey 1, we show the data for the 127 unique respondents whose pro-redistributive arguments were retained in the final sample of 160 arguments (each respondent wrote one fairness and one externality argument, but the quality filter in Survey 2 rarely kept both). Respondents chose 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

Table D7: Representativity: Survey experiment

	Survey 1 (Writers)	Survey 2 (Screeners)	Survey 3 (Evaluators)	U.S. Population
Male (%)	—	—	50%	49%
Median age	—	—	45	47
Democrat (%)	77%	—	34%	28%
Republican (%)	9%	—	31%	28%
Independent (%)	14%	—	36%	41%
College degree (%)	—	—	44%	33%
White (%)	—	—	60%	59%
Black (%)	—	—	12%	13%
Hispanic (%)	—	—	18%	19%
Income $0-25k (%)	—	—	25%	30%
Income $25-50k (%)	—	—	26%	18%
Income $50-100k (%)	—	—	29%	22%
Income $100k+ (%)	—	—	21%	30%
South (%)	—	—	40%	38%
West (%)	—	—	18%	24%
Midwest (%)	—	—	22%	21%
Northeast (%)	—	—	18%	17%
N	127	210	4089	—

Note. This table shows the demographic composition of respondents across the three surveys, compared to the U.S. adult population. Survey 1 collected only political affiliation; Survey 2 collected no demographic information. U.S. population benchmarks are from the 2022 American Community Survey (gender, age, education, race, region, income), Gallup 2022 annual average (party identification, excluding leaners), and the 2023 Current Population Survey (2022 household income). "—" indicates the variable was not collected.

study, we intentionally over-sampled Democrats. 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. The raw self-reported party affiliation of argument authors is 77% Democrat, 9% Republican, and 14% Independent; after reclassifying independents who lean toward a party, the composition is 88.8% Democrat-leaning and 11.3% Republican-leaning, similar to that in the Congressional pro-redistributive speech data (91.4% Democrat, 8.6% Republican). We did not collect data on other demographic dimensions, 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, was conducted with the professional survey company *Dynata* using pre-specified quotas on gender, age, political affiliation, income, region, and race. As shown in Table D7, the resulting sample closely matches the U.S. adult population on most dimensions. Gender, race, and regional composition are well-aligned with population benchmarks. The sample is slightly more educated than the general population (44% with a college degree, compared to 33% nationally) and has slightly lower incomes (21% over $100k annual income, compared to 30% for U.S. adults). Political affiliation is broadly representative, though Independents are somewhat underrepresented relative to Gallup benchmarks. Overall, the coverage is similar to comparable studies (e.g. Stantcheva, 2021; Lobeck and Støstad, 2023) with better coverage of racial minorities. While we cannot fully guarantee representativeness,

Table D8: Attrition: Survey 3

	Final sample	Removed	Left voluntarily
Republican %	31%	32%	31%
Democrat %	34%	31%	33%
Independent %	36%	37%	36%
Male %	50%	47%	36%***
White %	60%	50%**	54%**
Black %	12%	17%	14%
Neither black or white %	28%	33%	32%*
Income $0-25k %	25%	35%***	34%***
Income $25-50k %	26%	24%	26%
Income $50-100k %	29%	27%	25%*
Income $100k+ %	21%	14%*	15%**
Age 18-29 %	20%	48%***	21%
Age 30-39 %	19%	31%***	20%
Age 40-49 %	16%	16%	19%
Age 50-59 %	17%	4%***	18%
Age 60-69 %	17%	1%***	13%**
Age 70 and above %	11%	0%***	10%
4-year college degree or more %	44%	32%***	36%***
Employed %	57%	76%***	55%
Unemployed %	9%	12%	11%
Outside the labor force %	34%	12%***	34%
South %	41%	44%	43%
West %	19%	13%	18%
Northeast %	18%	21%	18%
Midwest %	22%	21%	20%
Prior externality beliefs %	53%	38%***	49%
Prior fairness views %	56%	80%***	57%
Sample size	4,089	216	724

Note. This table shows the demographic composition of Survey 3 respondents across three groups: the final sample, those who started but did not complete the survey, and those who were removed on the pre-specified criterion of being among the fastest 5% to complete the survey. Stars indicate significant differences relative to the final sample (*Significance levels*: *5%, **1%, ***0.1%.). "Prior externality beliefs" is the share of respondents who believe that an increase in economic inequality would change society for the worse, based on the question: "*Generally speaking, do you think an increase in economic inequality would change society for the better, for the worse, or not at all?*". "Prior fairness views" is the share who do not disagree with the statement: "*The distribution of money and wealth in the United States is basically fair, because everybody has an equal opportunity to succeed.*"

as participation in online access panels is inherently selective, related studies report broadly comparable results to nationally representative benchmarks (Stantcheva, 2023; Haaland et al., 2023).

Attrition In terms of attrition, Survey 1 recruited 360 Prolific workers, of whom 298 completed the survey (83% completion rate). Survey 2 recruited 242 Prolific workers, of whom 210 completed the survey (87% completion rate). Since Surveys 1 and 2 serve only to elicit and screen arguments and capture limited demographic information, we do not report attrition tables for them.

Survey 3 recruited 5,029 respondents from the Dynata panel who finished the demographic section of the survey. Of these, 4,305 completed the survey (86% completion rate). We had also pre-specified dropping the fastest 5% survey-takers to exclude inattentive respondents, yielding a final sample of 4,089 respondents.

We show attrition checks in Table D8. The fastest 5% of respondents differ significantly from the rest of the sample in ways consistent with inattention. For example, they rarely select older ages in the age drop-down menu, and their responses across questions are shifted toward a more uniform distribution. Respondents who dropped voluntarily are more likely to be female,

non-white, lower-income, and less educated, but we find no evidence of selective attrition by political affiliation, prior externality beliefs, or prior fairness views—the variables most relevant to our experimental outcomes. Our main findings are unchanged when we include these groups in the analysis.[22] We thus find no evidence that attrition is a concern for our experimental estimates.

D.5 Differences to pre-specification

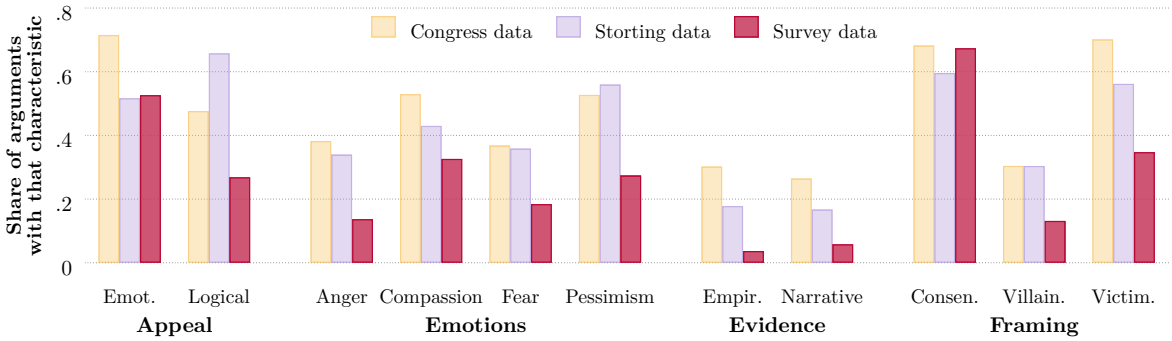
When launching the survey experiment, our objective was purely to collect *evaluations* of different types of arguments. In the final paper we conduct various additional analyses that were not pre-specified, largely due to the proliferation of LLMs (the experiment was done before the initial launch of 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 (Appendix E.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 10-11), 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 4,089 respondents rather than the pre-registered 4,000. Restricting to N=4,000 has no effect on results.
- 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 Appendix E.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.2-E.3.

[22] For those who left voluntarily, this means including the data up until they left the survey.

- 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.2); as these two types of anger differ substantially both empirically and conceptually, we do not present the evidence of the merged sample.

Figure E7: 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.

E Survey Experiment: Further data and analysis

E.1 Content classification: Comparison to legislative excerpts

Nearly all content classifications appear more frequently in the legislative speeches than in the experimental data (Figure E7). 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.

E.2 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.3 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.4 Examples: Argument content

Here we show illustrative examples of each content classification from the arguments discussed in Section 4.3.

E.4.1 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.4.2 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.4.3 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.4.4 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.*

E.5 Robustness: Effect of arguments on redistributive preferences (Supplement to Section 4.4)

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.[23] 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.[24] 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 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

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

[24] 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.

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 G7).

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 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.

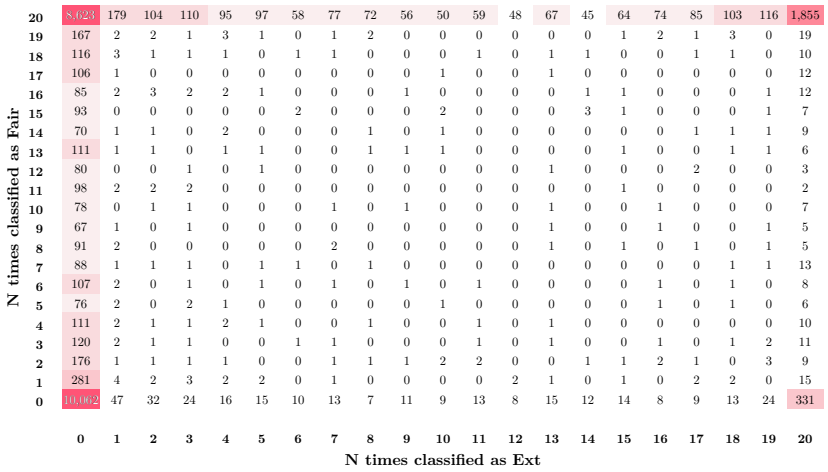
Table E9: Effect of Arguments and their Characteristics on Redistributive Preferences

	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 G7 shows that the findings are virtually identical when restricting the sample to respondents who passed the check. Table G8 shows the distribution of pro-redistributive arguments seen. *Significance levels*: *10%, **5%, ***1%.

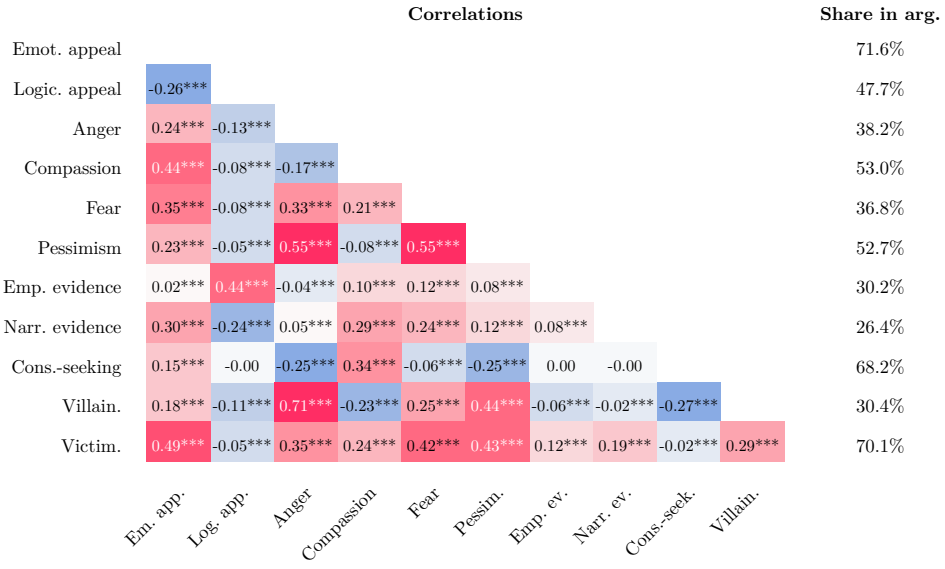
F Appendix Graphs

Figure F1: Consistency of classification as Fairness or Externality type



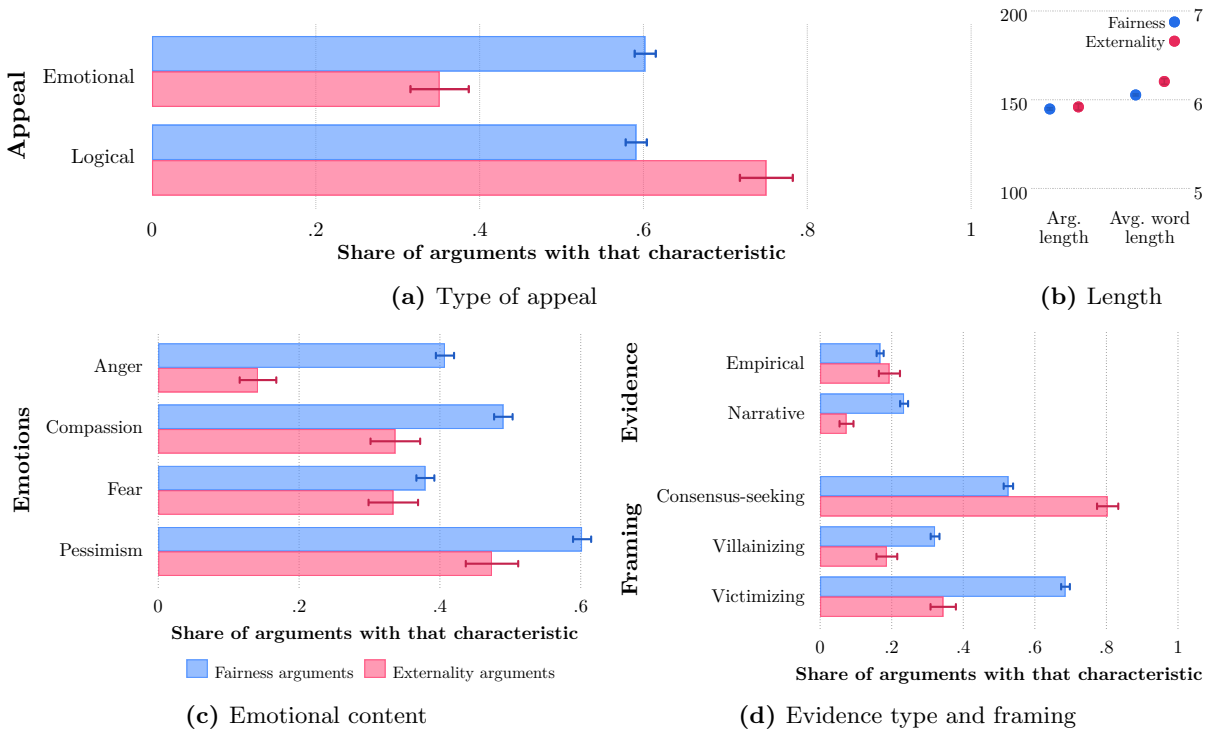
Note. This figure complements Section 3.1, reporting the results from 40 API calls, 20 for the fairness classification and 20 for the externality classification, in the U.S. Congressional data. An excerpt is classified as an externality or fairness argument if it is classified as such in at least 90% 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 C2. 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 4 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.

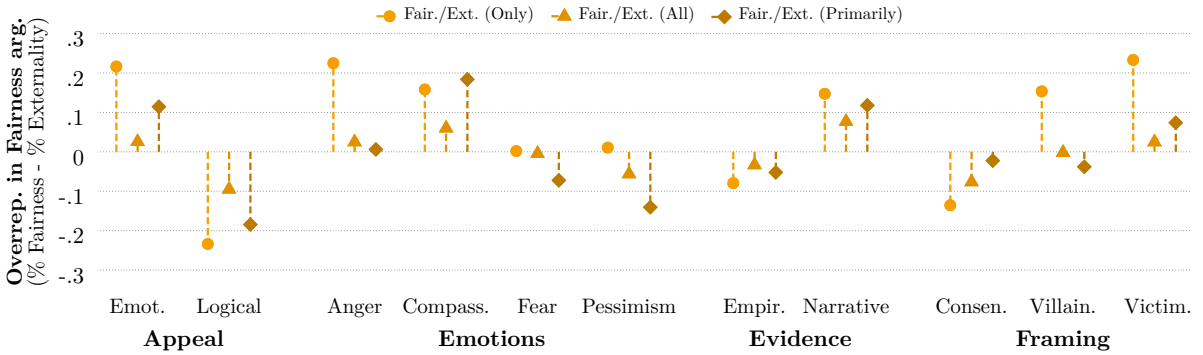
Figure F4: Content gaps between fairness and externality excerpts, with confidence intervals



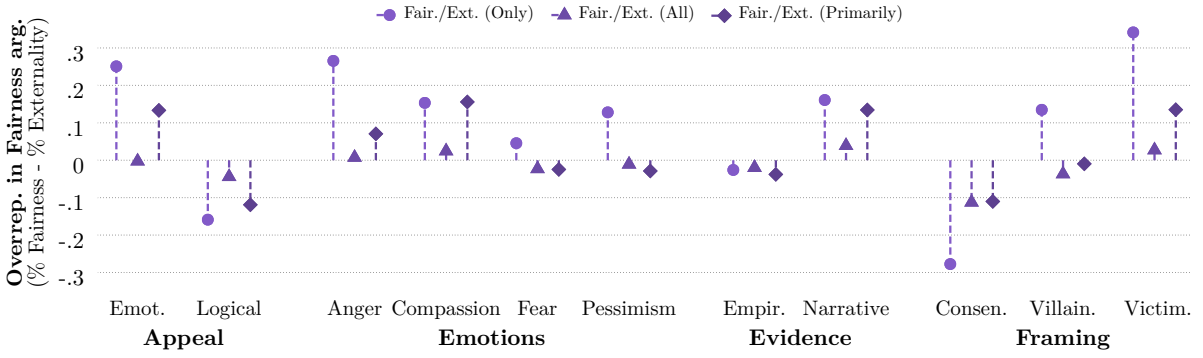
Note. This figure complements Figure 5 by adding 95% confidence intervals to the content differences between fairness and externality arguments across the two legislative settings (U.S. Congress and Norwegian *Storting*).

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

(a) U.S. Congress speeches

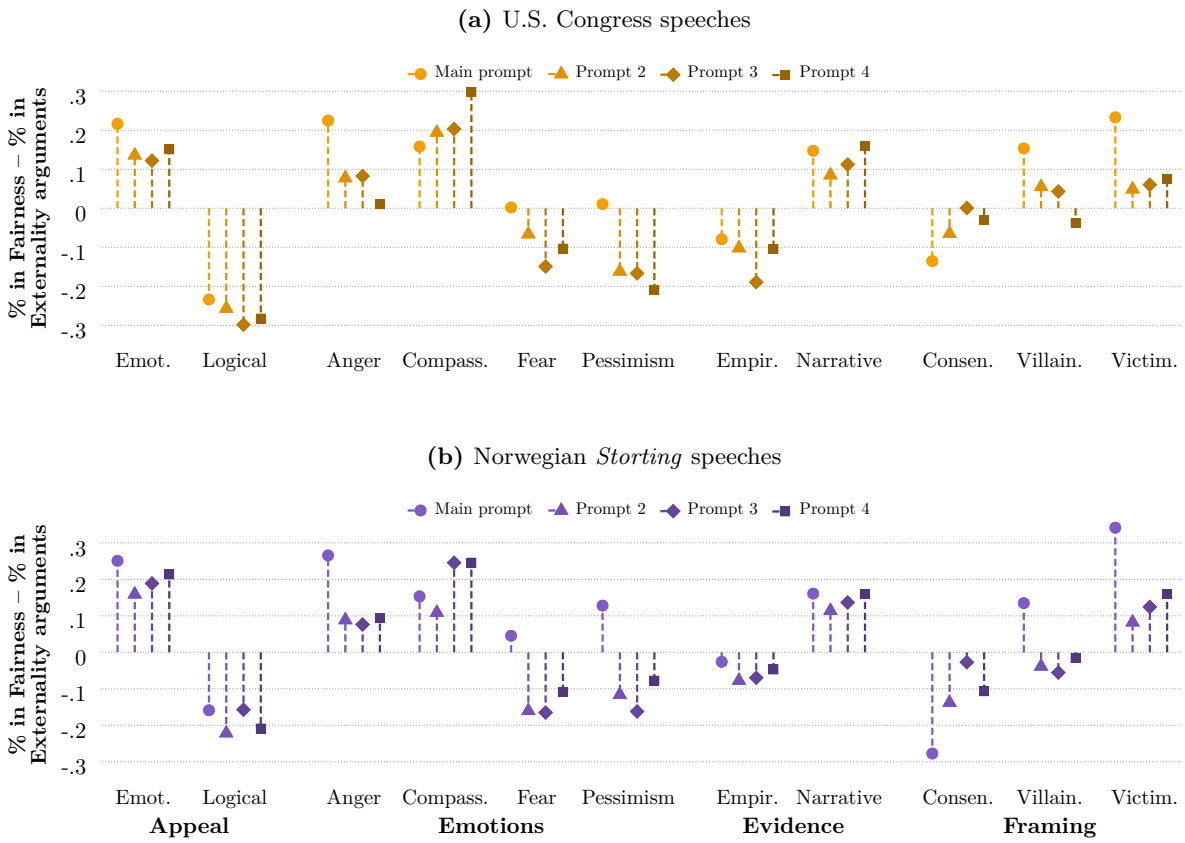


(b) Norwegian *Storting* speeches



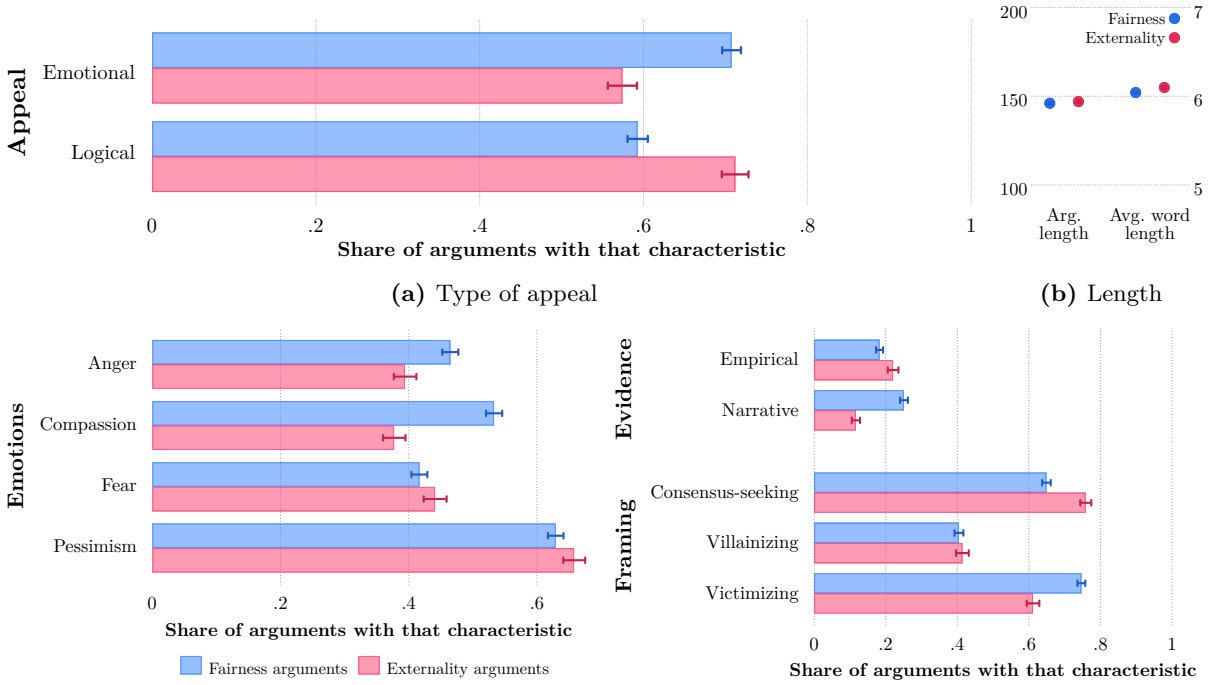
Note. This figure complements Section 5 and Appendix C.1, reporting content differences across fairness- and externality-based excerpts across the U.S. Congress and the Norwegian *Storting*. In each dataset, we compare (i) the main specification (*Fair. Only* and *Ext. Only*), (ii) the *Fair. All* and *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.19–B.24. 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

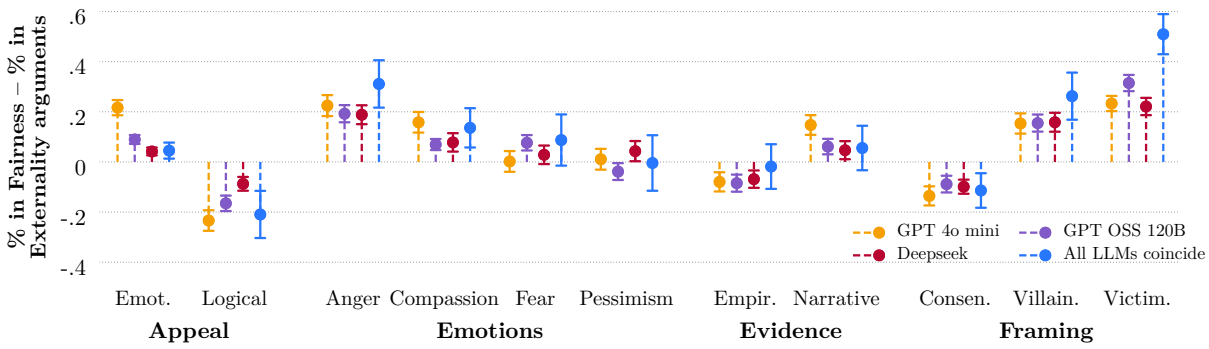


(a) Type of appeal **(b)** Length

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

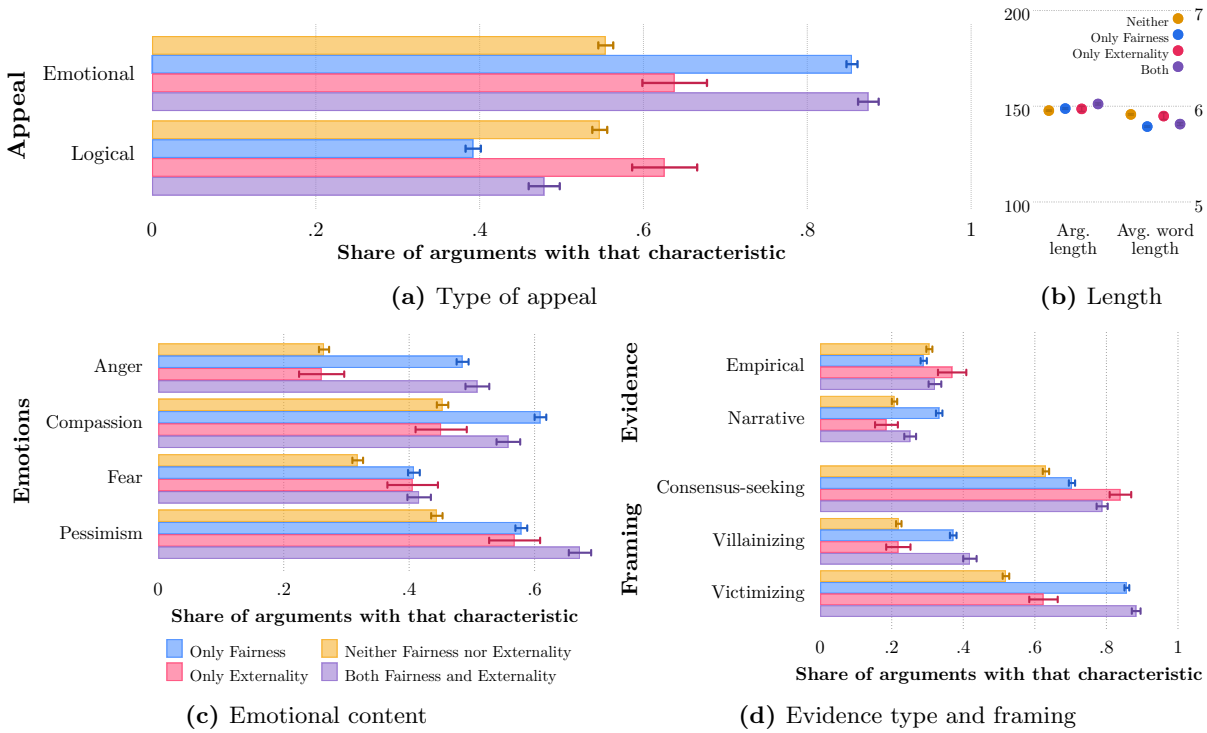
Note. This figure complements Figure F3, showing the content of fairness and externality-based excerpts in the Norwegian *Storting* using the "primarily" classification from Appendix B.18. The classification method is discussed further in Appendix C.1; each argument is classified as primarily making 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: 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.

Figure F9: U.S. Congressional Speeches: Content of arguments across argument type, all four types

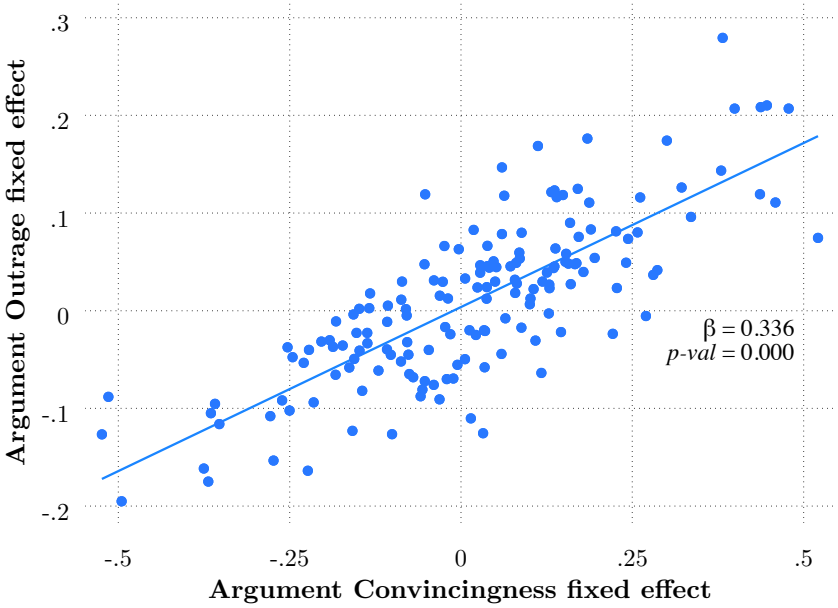


(a) Type of appeal **(b)** Length

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

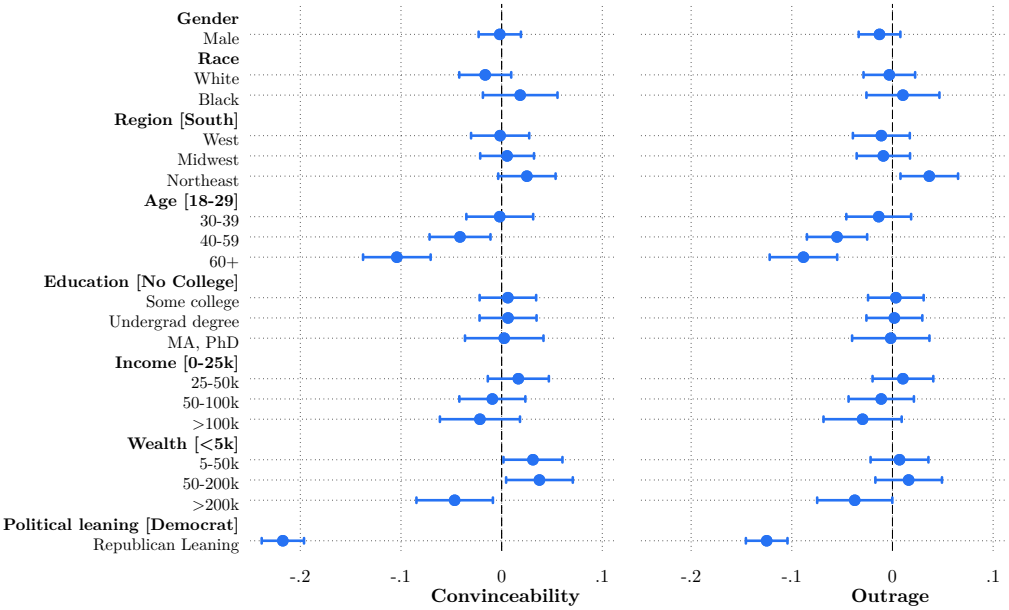
Note. This figure complements Figure 4, showing content classifications for speeches that were tagged as containing *both* a fairness and an externality argument, or *neither* a fairness nor an externality argument. Results are based on the full sample, including speeches that contain both fairness and 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 F10: Association between convincingness and outrage



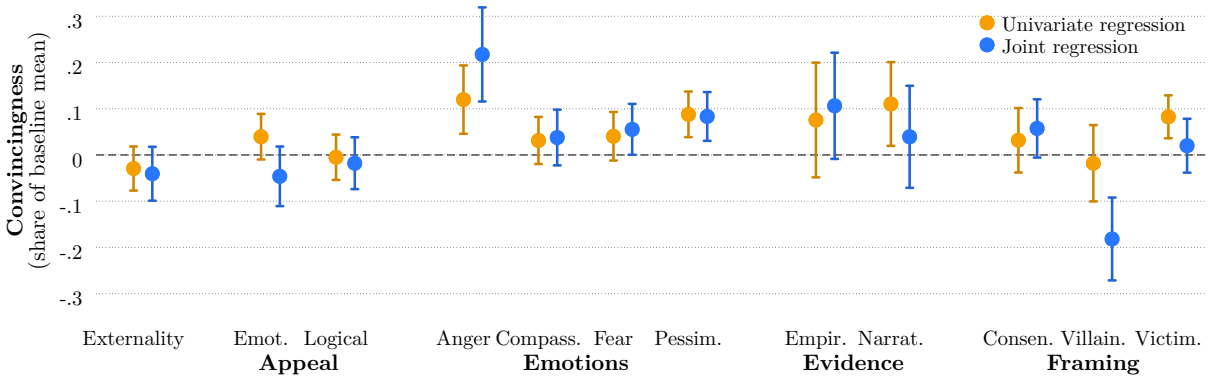
Note. This figure complements Figures 10-11, showing the association between convincingness and outrage for each of the 160 arguments in the survey experiment.

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



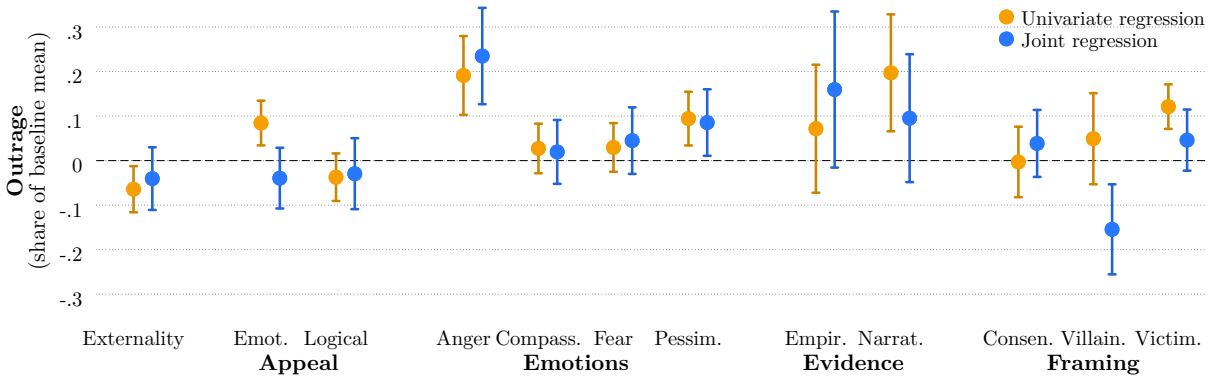
Note. This figure complements Section 4.3, reporting which demographic variables correlate with the respondent being (i) convinced and (ii) outraged by a given argument. In both cases, we show regression coefficients in p.p. from a joint regression on the outcome variable. 95% confidence intervals shown.

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



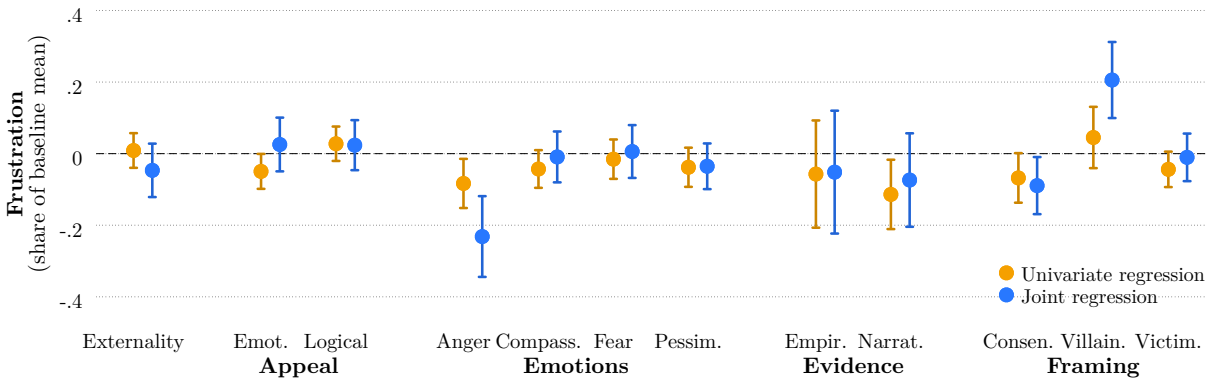
Note. This figure shows the result of regressing convincingness on argument characteristics. The sample is restricted to participants who passed the attention check (72.3% of the 4,089 participants). The yellow dots correspond to a regression of convincingness on the indicated variable, as well as person 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 10.

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



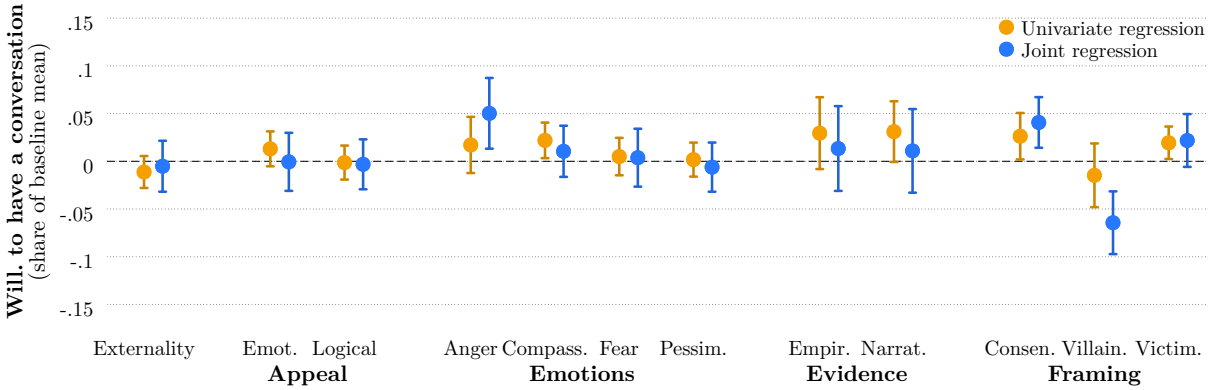
Note. This figure shows the result of regressing outrage on argument characteristics. The sample is restricted to participants who passed the attention check (72.3% of the 4,089 participants). The yellow dots correspond to a regression of outrage on the indicated variable, as well as person fixed effects, 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 11.

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 4.3.4.

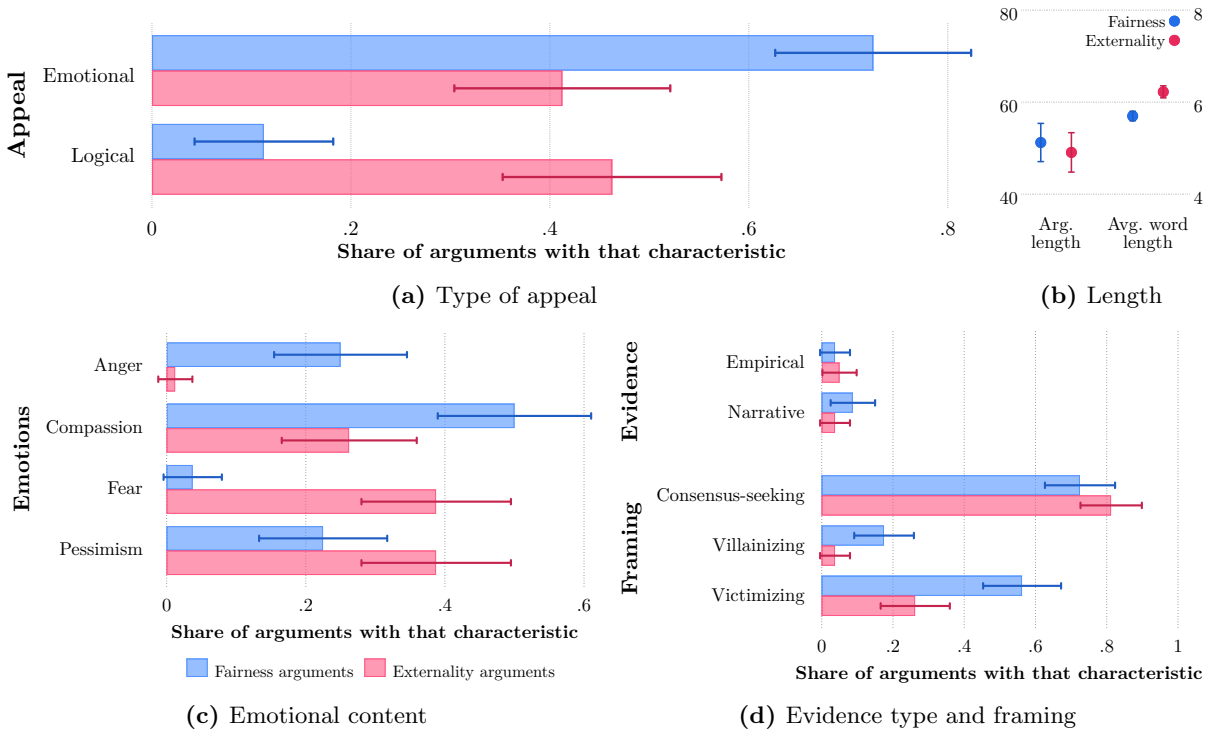
Figure F15: Association between willingness to have a longer conversation and argument characteristics



Note. This figure complements Appendix E.3, 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.

G Appendix Tables

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 E7, 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.

Table G1: 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**
Republican	0.027	0.021	0.047	0.058*	-0.059***	-0.061***	-0.049**	-0.053**
Congress member's educ. level	-0.004	-0.004	0.007	0.009	0.002	0.002	0.003	0.002
Female	0.016	0.019	0.015	0.016	-0.028***	-0.028***	-0.030**	-0.029**
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 complements Table 1 and 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 G3 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%.

Table G2: 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)	0.242			0.032	0.553			-0.727
Log average county income		0.072		0.025		0.187*		0.034
Avg. county educ. level (years)			0.018	0.011			0.057**	0.094**
Female	0.033	0.033	0.032	0.032	-0.001	-0.001	-0.004	-0.005
Storting member's educ. level	-0.029	-0.029	-0.029	-0.029	-0.022	-0.022	-0.025	-0.024
Progress (far right)	-0.197***	-0.196***	-0.197***	-0.197***	-0.360***	-0.359***	-0.367***	-0.366***
Conservatives (center-right)	-0.289***	-0.289***	-0.289***	-0.290***	-0.204***	-0.206***	-0.208***	-0.208***
Christian Dems (centrist)	-0.171***	-0.171***	-0.171***	-0.171***	-0.227***	-0.225***	-0.228***	-0.228***
Greens (left)	-0.157*	-0.151*	-0.158*	-0.157*	-0.193***	-0.183***	-0.213***	-0.202***
Communist party (far left)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Socialist Left (left)	0.133***	0.133***	0.133***	0.133***	0.052	0.051	0.050	0.050
Centre (agrarian/centrist)	-0.113***	-0.113***	-0.112***	-0.113***	-0.161***	-0.162***	-0.158***	-0.158***
Liberals (centrist)	0.000	0.002	-0.001	-0.001	-0.177***	-0.175***	-0.190***	-0.189***
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 complements Table 2 and 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.4. The dependent variable corresponds to the *All* definition, that is, including arguments that were classified as both a fairness- and externality-based argument). Table G4 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%.

Table G3: 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 G4: 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 G5: 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 G6: 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 G7: 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 E9, 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 G8: 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 E9, 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