"Us vs Them": Salient Conflict and Belief Polarization*

Nicola Gennaioli Frederik Schwerter Guido Tabellini

February 19, 2026

Abstract

In an online experiment with a representative US sample (N=12,960) we show that increasing the salience of an economic or a cultural conflict *without providing any news* boosts disagreement on a range of political issues by 8-35%. The data support two key predictions of the Bonomi et al. (2021) identity theory of political beliefs. First, polarization amplifies – through stereotypes – latent disagreement among the economic or cultural groups standing in salient conflict. Second, there is a belief realignment away from no-longer salient groups, causing some people to move across the conservative-progressive divide. These results can throw light on real world political conflicts and propaganda.

*We thank Pedro Bordalo, Chris Roth, Moses Shayo, Andrei Shleifer and Florian Zimmermann for helpful comments. Gennaioli: Università Bocconi, Department of Finance and IGIER (email: nicola.gennaioli@unibocconi.it); Schwerter: Frankfurt School of Finance and Management, Economics and Law Department (email: f.schwerter@fs.de); Tabellini: Università Bocconi, Department of Economics and IGIER (email: guido.tabellini@unibocconi.it). We thank Ginevra Casini for excellent research assistance. Gennaioli thanks the European Research Commission for an Advanced ERC Grant (GA 101097578). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

1 Introduction

Polarization is a key contemporary issue, in the US and other advanced democracies. Voters of different parties sharply disagree on facts and policies, especially in cultural domains such as religion, race, immigration (e.g., Achen and Bartels, 2017; Alesina et al., 2023). Disagreement in turn fuels political conflict, which many see as a main social problem (Klein, 2020; Graham and Svolik, 2020; Levitsky and Ziblatt, 2018; Baker et al., 2016). But where does disagreement come from? And why does it occur in the presence of verifiable data or even scientific consensus (Kahan, 2015)?

In rational choice, disagreement reflects differences in priors, data and stable tastes. Growing evidence challenges this approach, showing distortions in the processing of news. One body of work emphasizes heuristics causing under-reaction, neglect of correlation, selection in "echo chambers", or overconfidence (Tversky and Kahneman, 1974; Enke and Zimmermann, 2019; Enke, 2020; Levy and Razin, 2019; Ortoleva and Snowberg, 2015). Another body of work stresses motivated beliefs: people under-weight contrary news to stick to their preferences (Bénabou and Tirole, 2016; Thaler, 2024). In both cases, disagreement is founded on distorted use of news.

We evaluate a different mechanism: people form beliefs on disputed social issues based on salient ingroups and outgroups. This approach rests on a major social psychology paradigm, Social Identity Theory (Tajfel and Turner, 1979), and has one key implication: the mere contextual salience of conflict between groups fosters disagreement among them, even in the absence of news.

We formalize this force using the Bonomi et al. (2021) model of identity and beliefs (henceforth BGT). There are two policy domains: economics, comprising material issues like taxation, inequality, etc., and culture, comprising issues such as race, religion, etc. A person's undistorted belief on an economic (cultural) issue depends on issue-specific tastes or information and on her "primitive" economic (cultural) belief. Primitive beliefs capture broad personal attitudes shaped by experiences, social background, etc. A person may trust markets or rich people more than another and, due to this primitive economic belief, be more conservative on a range of economic issues. A person may also cherish traditional values more than another and, due to this primitive cultural belief, be more conservative on a range of cultural issues. Primitive beliefs are correlated: many people tend to be "pure" conservative

or progressive in economics and culture. But correlation is not full: some people are "mixed", with incongruent economic and cultural beliefs. They may distrust markets but cherish traditional values, or vice-versa.

Primitive beliefs create two latent conflicts: an economic one, say between the supporters of free markets and those of government intervention, and a cultural one, say between religious and secular people. Based on primitive beliefs, a person belongs to an economic and a cultural group (conservative or progressive). She can then identify with one of them or remain unidentified. As shown in BGT, a person is more likely to identify with her most salient ingroup, and identity causes her beliefs to move toward ingroup stereotypes. The key consequence of this process is to amplify disagreement along primitive beliefs, dampening the role of issue specific tastes or prior information.

In this model, higher salience of economic (cultural) conflict causes two effects. First, it increases the share of people who are economically (culturally) identified versus not identified at all, polarizing disagreement among economic (cultural) groups. Second, it realigns some people, reducing identity with no longer salient groups compared to salient ones. Due to these effects, changes in salience cause: i) changes in disagreement among different groups, ii) belief instability, with spillovers across issues including across economic and culture, and iii) stronger instability in issues where salient groups disagree more.

We test the model in a pre-registered online experiment on a large representative US sample (N=12,960). We elicit beliefs on 10 issues, 5 in economics and 5 in culture, 2 of which are factual (i.e. how the world works) and 3 normative (i.e. policies). We measure membership in an economic group by asking respondents to select "which enemy they or people like them find most threatening" from a list of conservative and progressive groups such as billionaires, trade unions, etc. We measure cultural membership using a list comprising opposing cultural groups such as white pride, LGBTQ, etc. In our Control condition, memberships are measured after beliefs. Our treatments render one conflict salient by eliciting one enemy before beliefs: the economic one in the Economic Treatment (ET) and the cultural one in the Cultural Treatment (CT). To measure identity, in ET and CT we further ask people to describe their ingroup and outgroup and to report expected disagreement among them.

Our protocol: i) does not provide news, it increases the salience of a conflict during belief formation by changing the sequence of questions, and ii) elicits both economic and cultural memberships. It thus allows us to measure dis-

agreement among specific groups when these are latent vs salient, which is key to our model. It also minimizes experimenter demand: the enemy questions are similar to other policy questions we measure. It is unlikely that people notice to be in a treatment and, to the extent they do, that they infer from it our many model predictions. Consistent with this notion, the distribution of people among economic and cultural groups is very stable across treatments.

We find that salience causally increases disagreement: both ET and CT increase the standard deviation of economic *and* cultural beliefs in the population. Effects are large and occur on both facts and normative views. In principle, these effects may arise because some people are scared or angry due to salient conflict, so they become more conservative/authoritarian (Jost et al. 2003), or they give noisier answers.

Belief polarization however ties to groups, consistent with identity. First, increasing the salience of conflict among two groups amplifies their latent disagreement: ET polarizes the average economic conservative versus progressive, CT the average cultural conservative versus progressive. Effects are sizable: disagreement in different issues increases from 8% to 35% relative to Control. Second, changes in salience matter via "kernel of truth" stereotypes (Bordalo et al., 2016; Bonomi et al., 2021): ET (CT) causes individual beliefs to change more sharply in factual or normative issues where disagreement among the salient economic (cultural) groups is larger in Control.

Third, making one conflict salient realigns conflict away from no longer salient groups. We see a change in the locus of conflict: by reducing the relative salience of economic conflict, a switch from CT to ET reduces conflict among cultural groups relative to economic ones. The change comes from mixed types, consistent with a re-categorization of self from a conservative to a progressive ingroup and vice-versa. We also see belief de-polarization: people who identify – in ET or CT – with groups they expect to disagree little, are less polarized than comparable people in Control.

Our results are robust to controlling for a rich set of demographics, suggesting that the changing salience of economic and cultural groups is a robust force. They also hold after controlling for partisanship and when restricting to independent voters, showing that our measured economic and cultural groups shape polarization within and across parties, also suggesting that our treatments do not operate by priming party identity.

We unveil a new mechanism of political belief formation, based on the

salience of group conflict. It differs from existing accounts, which stress distortion of news based on heuristics, selective exposure (Levy and Razin, 2019), stable motives (Bénabou and Tirole, 2016), or partisanship (Kunda, 1990; Druckman et al., 2013; Klar, 2014). It predicts that polarization amplifies a latent disagreement among groups standing in salient conflict, which also implies that beliefs de-polarize if salient groups are non-conflicting. We also differ from standard accounts in which political persuasion polarizes voters based on party cues (Zaller, 1992) or hatred (Glaeser et al., 2005; Iyengar and Westwood, 2015) with little external constraint. In our theory, persuasion is constrained in important ways by voters' self categorization into social groups. For instance, a right-wing politician increasing the salience of cultural conflict can move pro redistribution religious voters to the right but exert the opposite effect on libertarian voters, creating a tradeoff.

In Section 6 we discuss real world implications of our theory. We argue that it can help explain and unify three forces often linked to the surge of populism and political conflict in the last two decades: i) economic and social shocks (Autor et al., 2020), ii) social media (Levy, 2021), iii) stereotyping of outgroups and affective polarization. We show that it also implies, different from news-distortion theories, that provision of unbiased information may fail to correct beliefs, and in fact may even backfire, if conflict among divided groups remains highly salient.

We build on work on populism and identity politics by offering an experimental test of Bonomi et al. (2021). Shayo (2009) presents a utility-based model in which identity affects the demand for redistribution but not factual beliefs, Grossman and Helpman (2021) study trade policy. In our approach identity shapes stereotypes, distorting also factual not just normative beliefs. Shayo (2020) surveys other related work on social identity.

We proceed as follows. Section 2 presents our experiment and reports key facts from our control condition. Section 3 presents the model, showing the polarizing and realignment effects of salience. Section 4 shows that ET and CT boost polarization within and across domains by amplifying conflict among salient groups, producing surprising belief spillovers. Section 5 shows that the same treatments create belief realignments. Section 6 concludes.

2 The Experiment

2.1 Conceptual Framework

A person's undistorted belief on a generic economic issue i (income inequality, redistribution etc.) or cultural issue j (discrimination, immigration, etc.) is respectively equal to:

$$e_i = e + \upsilon_i, \tag{1}$$

$$c_j = c + \upsilon_j, \tag{2}$$

where higher values stand for more progressive views. Terms e and c capture the person's general economic and cultural outlooks. Higher e implies more progressivity on all economic issues i, e.g. due to a belief that governments are more trustworthy or fair than markets. Higher c implies more progressivity on all cultural issues j, e.g. due to a belief that modern values are more reliable or fair than traditional ones. We call (e, c) *primitive beliefs* because they are rooted in a person's social background or experiences.

Terms υ_i and υ_j reflect issue-specific tastes or information. They are on average zero across issues and people, uncorrelated with each other and with (e,c): one may *know* that income inequality has increased (high υ_i on this issue) but view market outcomes as fair (low e_i on this issue).[1] (e, c) is Gaussian in the population, with zero means, unit variances and correlation ρ.

Primitive beliefs shape latent social conflicts. Economic conflict arises between the progressive group E of people having $e > 0$ and its conservative opponent $-E$ of people having $e < 0$. Cultural conflict arises between the progressive group C of people having $c > 0$ and its conservative opponent $-C$ of people having $c < 0$. If e and c are positively correlated, $\rho > 0$, groups E and $-E$ also disagree on culture and groups C and $-C$ also disagree on economics. We will see that this is indeed the case in the data.

A person (e, c) can in turn belong to one of four social types. Two types are "pure": they have congruent progressive or conservative ingroups, (E, C) or $(-E, -C)$. Two types are instead "mixed": economically progressive and culturally conservative, $(E, -C)$, or vice-versa, $(-E, C)$. The distinction between pure and mixed types is important to dissect our mechanism.

Our broad hypothesis is that increasing the salience of economic (cultural)

[1] In a Bayesian model, e can capture the prior belief on economic issues, c on cultural ones. In non-Bayesian models, they can reflect selective neglect of previously received data.

conflict can mobilize conflicting economic (cultural) groups, thereby amplifying their latent disagreement in primitive beliefs on a range of issues. In contrast to existing approaches, this occurs even if *no news* is provided.

We test our theory using a survey described in Section 2.2. Section 2.3 documents latent conflicts in the Control sample and shows, using text analysis, that our treatments primed the intended groups for most subjects.

2.2 Experimental Design

We fielded our survey in three waves (March 2023, June 2023, and October 2024) through Bilendi. The sample is representative of the US voting population in terms of gender, age, income, and region, but it is whiter and more educated than the benchmark, see Table B.1 in the Appendix. We measure beliefs, demographics, and perform two salience treatments described below. Survey details (precise questions, screenshots, etc.) are in Appendix B.2.

Outcomes: Political beliefs We consider five economic and five cultural issues, on verifiable events ("factual") and policy preferences ("normative"). Responses are on a 0-10 scale. Higher value means more progressive.

Economic facts concern: (i) income inequality in the last 30 years (progressive responses state greater increase), (ii) whether poverty is mainly caused by low effort (conservative response) or bad luck (progressive response). Economic policy questions elicit approval for (i) the federal estate tax; (ii) government support in finding jobs and in the standard of living; (iii) government provision of health and education (stronger approval is more progressive). In Equation (1), each of these beliefs reflects both e and v_i. Averaging across them for a respondent yields a proxy for her primitive economic belief e.

Cultural facts concern (i) whether Black Americans are discriminated against in hiring decisions (more discrimination means more progressive), (ii) the frequency with which an immigrant commits a crime (smaller frequency is more progressive). Cultural policy questions elicit approval of (i) affirmative action programs; (ii) tolerant policy toward refugees; (iii) making abortion legal (stronger approval is more progressive). Again, averaging these beliefs for a respondent yields a proxy for her primitive cultural belief c.

Group membership We obtain a proxy for membership in economic and cultural groups by asking respondents to self-categorize in the progressive or

conservative camp of different conflicts. For economics, after reporting her economic situation, the respondent is asked to choose, from the following list, which group "they and people like them feel most threatened by":

1. Billionaires who evade taxes
2. Big business
3. Wall Street lobbyists
4. Communists
5. Trade Union activists
6. Lazy people who could work but prefer to claim welfare benefits

The first three enemies are conservative, so their choice signals membership in a progressive economic group, and vice-versa for the last three enemies.

For cultural conflict, after reporting her ethnicity and religion, the respondent chooses her enemy from the following list:

1. Evangelicals
2. Supporters of the National Rifle Association (NRA)
3. Supporters of White Pride
4. Supporters of Black Lives Matter
5. The LGBTQ (Lesbian, Gay, Bisexual, Transgender, Queer) community
6. Cancel culture activists

The first three enemies are conservative, so their choice signals membership in a progressive cultural group, and vice-versa for the last three enemies.

Enemy choice measures two aspects. First, which particular conflict the respondent sees as more important in a domain (taxation or lobbying?, religion or race?). Second, whether the respondent is progressive or conservative in this conflict. We could have asked respondents to classify themselves as broadly progressive or conservative in economics and in culture, but we preferred enemy choice for three reasons. First, it lists concrete groups many respondents feel to be in conflict with. Second, it aligns with social identity theory (e.g., Tajfel et al., 1979; Bonomi et al., 2021), where salient conflict plays a key role. People have multiple latent identities and are more likely to identify with the ingroup

currently standing in salient conflict with its outgroup.[2] Third, and as we discuss next, making people think about a concrete outgroup and its features enhances the salience of conflict in the economic or cultural domain.

Treated variation in conflict salience We randomly assign participants to one of three groups. The Control group reports beliefs before the enemy selection. Beliefs in Control are thus unaffected by enemies.

In the "economic treatment" (ET henceforth), participants face the economic enemy question *before* reporting beliefs (and, like in Control, the cultural enemy question after reporting beliefs). To measure the groups we prime and foster engagement with them, we ask participants to describe the ingroup of people who feel threatened by the same enemy, and why they feel threatened. Before answering on each issue i, respondents also report whether they think that their ingroup and outgroup disagree on the issue. This measure of perceived disagreement on i which is also indicative of the groups we prime and – as we will see – relevant to our theory. The "cultural treatment" (CT) analogously sequences the cultural enemy choice, ingroup and outgroup descriptions, expected disagreement, and the belief elicitation.

We harmonize the Control survey with four initial questions where subjects state which animal they feel threatened by, explain why, describe which animals they like, and state how close they feel to people who like the same animals. These questions also involve "threat" and "closeness to others", but do not increase the salience of social conflict.

By eliciting enemies and group descriptions before beliefs, ET and CT increase the salience of economic or cultural conflict relative to Control. We mainly view such conflict as occurring between broad groups (E vs -E or C vs -C) relative to Control. It is possible, though, that ET and CT may prime a divide on the specific enemy issue (e.g. taxes or religion). This narrower conflict is also related to primitive beliefs, so it should have similar implications in our model. We later show that our results are robust to this case.

Crucially, our treatments do not provide any news: they neither mention

[2]The third survey wave slightly changed enemies in CT ("Religious extremists like Evangelicals", "Racists like White supremacists and White Pride supporters", "Illegal immigrants and those who defend them", and "Woke people and cancel culture extremists"; and combined them in ET ("Billionaires who evade taxes", "Big business and Wall Street lobbyists", "Communists and Trade Union activists", and "Lazy people who could work but prefer to claim welfare benefits" (the former two are progressive, the latter conservative cultural ingroups). Our results' cross-wave stability shows that specific wording does not matter.

Outcome:	OLS regressions					
	E	C	E,C	E,-C	-E,C	-E,-C
	(1)	(2)	(3)	(4)	(5)	(6)
ET	-0.0175*	-0.00185	-0.00725	-0.0102	0.00540	0.0121
	(0.010)	(0.009)	(0.009)	(0.008)	(0.007)	(0.009)
CT	-0.00191	0.0142	0.0173*	-0.0193**	-0.00317	0.00508
	(0.009)	(0.009)	(0.009)	(0.008)	(0.007)	(0.008)
Constant	0.599***	0.541***	0.426***	0.173***	0.115***	0.286***
	(0.007)	(0.006)	(0.006)	(0.006)	(0.005)	(0.006)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.20	0.28	0.28	0.03	0.02	0.23
N	12954	12954	12954	12954	12954	12954

Table 1: In Column (1) [(2)] we regress respondents' economic [cultural] identity membership (=1 if progressive, =0 if conservative) on an economic treatment dummy and a cultural treatment dummy and control variables (gender, age, race/ethnicity, educational attainment, annual income, labor market status, region of residence, and party affiliation). In Columns (3)-(4), we conduct the same regressions for each one of the social identity types (consisting to economic and cultural identity) separately. Robust standard errors in brackets. *** (**,*) indicate statistical significance at 0.01 (0.05, 0.1) level.

facts nor the pros and cons of different policies; they only harness subjects' reflections, altering the conflict context. Of course, ET and CT do not occur in a vacuum: respondents may have pre-existing "top of mind" conflicts, which our model takes into account. Our treatments are designed to minimize experimenter demand. The enemy questions are qualitatively similar to other policy questions we ask in the survey, making it unlikely that participants in ET or CT realize to be treated. Even in this unlikely case, it is not obvious that participants would infer the model implied treatment effects we test for. Our theory makes a rich set of non-obvious predictions that entail also domains and policy issues that are distant from the primed enemy and that work differently for mixed and pure types.

To assess treatment validity we check if the distribution of people in different groups is stable across Control, ET and CT. A non-constant distribution may suggest that participants have inferred at least the group-priming intent of our experiment and change their reported membership accordingly. Table 1 regresses in Columns 1 and 2 the share of economic progressives E and cultural progressives C on treatment dummies and also controls to ensure comparable

Baseline Disagreement in the Control				
$\sigma_{\tilde{e}}$	$\sigma_{\tilde{c}}$	σ_f	σ_n	corr($\tilde{e}$,$\tilde{c}$)
1.99	2.43	2.03	2.30	0.703

Table 2: Standard deviation of average economic beliefs, $\sigma_{\tilde{e}}$, cultural beliefs, $\sigma_{\tilde{c}}$, factual beliefs, σ_f, and normative beliefs, σ_n, and the correlation coefficient corr($\tilde{e}, \tilde{c}$) between average economic and cultural beliefs.

group compositions.[3] Columns 3 to 6 considers the share of respondents in each of the four granular social types. Estimated coefficients on the ET and CT dummies are tiny and almost always insignificant, consistent with random assignment.

Political affiliation and preregistration The end of the survey measures participants' political affiliations ("do you consider yourself a Republican, a Democrat or an independent?"). This question allows us to distinguish partisanship from membership in social groups.

We preregistered our main analyzes on AsPredicted.org in trial 133192 for Waves 1-2 and in trial 191959 for Wave 3. Both preregistrations feature the same tests on how ET and CT causally affect beliefs. We therefore pool all waves to leverage the sample size of 12,960 unique observations, although as shown in Table 4 our results also hold for each wave separately. Wave 3 included a subsample (N=1,366) where conflict over trade policy was made salient and related belief questions were elicited. The results are broadly consistent with those on ET and CT. Given the smaller sample size and the focus on a narrower conflict, we discuss them in Appendix G.

2.3 Latent Conflict and Salient Groups

We first document belief polarization and latent conflicts in our Control sample. We proxy for a respondent's economic and cultural outlooks by averaging her five economic beliefs into $\tilde{e}$ and her cultural beliefs into $\tilde{c}$. We also construct separate averages for the voter's factual beliefs and policy preferences. Each index ranges from 0 to 10, with higher values implying more progressivity.

[3]Controls include participants' political party affiliation, gender, age, US region of residency, education, labor market status, income, ethnicity and race, religion, and whether they were born in the US or not.

Polarization in the Population Table 2 proxies polarization by: a) the standard deviation of beliefs (Columns (1)-(4)), and b) the correlation of economic and cultural beliefs, which captures divided belief systems (Column (5)).

Polarization is significant, in economics and culture, over both facts and policy. It is larger in $\tilde{c}$ than in $\tilde{e}$, consistent with the importance of cultural conflict in the US. Economic and cultural beliefs are strongly positively correlated, creating two clusters. Conservatives who believe that immigrants commit more crimes or that there is little discrimination also tend to believe that income inequality has not increased or that effort is more important than luck for becoming rich, and vice-versa for progressives.[4]

Polarization between Groups To detect latent disagreement, based on enemy choice we allocate each subject to one social type in: (E, C), $(-E, -C)$, $(E, -C)$, $(-E, C)$. We then aggregate types into economic groups E and $-E$ and cultural groups, C vs $-C$. We also construct narrow groups by aggregating all people who chose the same specific enemy.

Panel A of Figure 1 reports the average $\tilde{c}$ and $\tilde{e}$ (x and y axis, respectively) of economic (E vs $-E$) and cultural (C vs $-C$) groups. Panel B reports the beliefs of narrow groups. We also report the beliefs of partisans, R and D, for comparison. Broad groups are strongly polarized in their core dimension: the average member of E (C) is more progressive across economics (culture) compared to that of $-E$ ($-C$).

Polarization is however strong also in the non-core dimension: members of E (C) are also more progressive on a range of cultural (economic) issues than their $-E$ ($-C$) outgroup. Enemy choice is informative about groups that exhibit broad disagreement, again pointing to the positive correlation in e and c, $\rho > 0$. Critically, correlation is far from perfect, creating the phenomenon of belief realignment studied in Section 5.

Disagreement in Panel A is not mechanically related to parties. Members of party D are more likely to be progressive than members of party R but the overlap is far from perfect (see Table H.9 in Appendix). First, about one

[4]Positive correlation of $\tilde{e}$ and $\tilde{c}$ could be due to education or values (Bonomi et al., 2021). In the data correlation is positive but far from full, consistent with the intuition that economics and culture tie to different domains. A person may think that the rich are undeserving or too powerful, and hence be progressive on redistribution, and yet cherish religious values, and hence be conservative in culture.

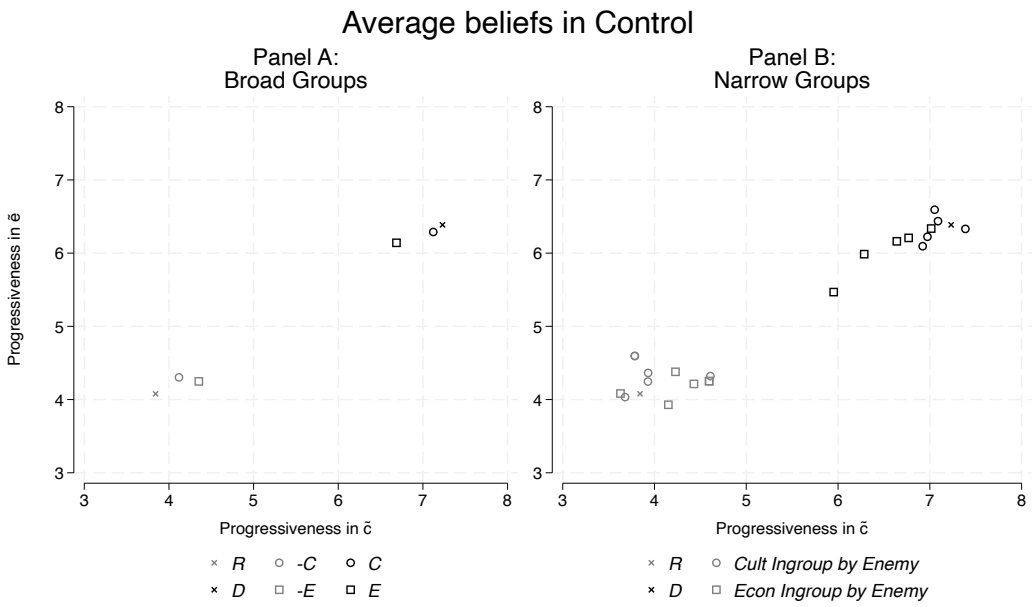


Figure 1: Panel A reports average economic and cultural beliefs by broad groups E, $-E$, C, $-C$, and by party groups, D and R. Panel B reports average beliefs of narrow groups, each consisting of people who chose the same enemy.

third of voters are Independent, and group-based polarization among them is comparable to that in Panel A.[5] Second, about one third of respondents are mixed types $(E, -C)$ and $(-E, C)$, who are not a-priori attributable to parties. Despite this, mixed types exhibit systematic disagreement: $(E, -C)$ is more conservative than $(-E, C)$ in both culture (a lot, about 1.16 points) and economics (a little, about 0.15 points). The fact that mixed types disagree based on culture signals that cultural conflict is more salient at baseline. Overall, then, group polarization goes well beyond parties.

Panel B shows that sharp conflict also arises for narrow groups: they lay along a progressive-conservative divide also correlated across domains. Interestingly, there is variation across narrow groups on the same progressive or conservative side. This variation allows us to test new model predictions.

Based on the observed latent conflicts among economic or cultural groups, Section 3 derives predictions for treatment effects. Before doing so, we study how treatment affects respondents' representations of ingroups and outgroups.

[5]Average disagreement is 1.8 among Independents in E and $-E$ and 2 for C and $-C$, compared to 2.1 and 2.5 in the full sample. Remarkably, the average progressive (resp. conservative) position is close to that of the average Democrat (resp. Republican) even though parties comprise only 70% of people, who are more extreme than average.

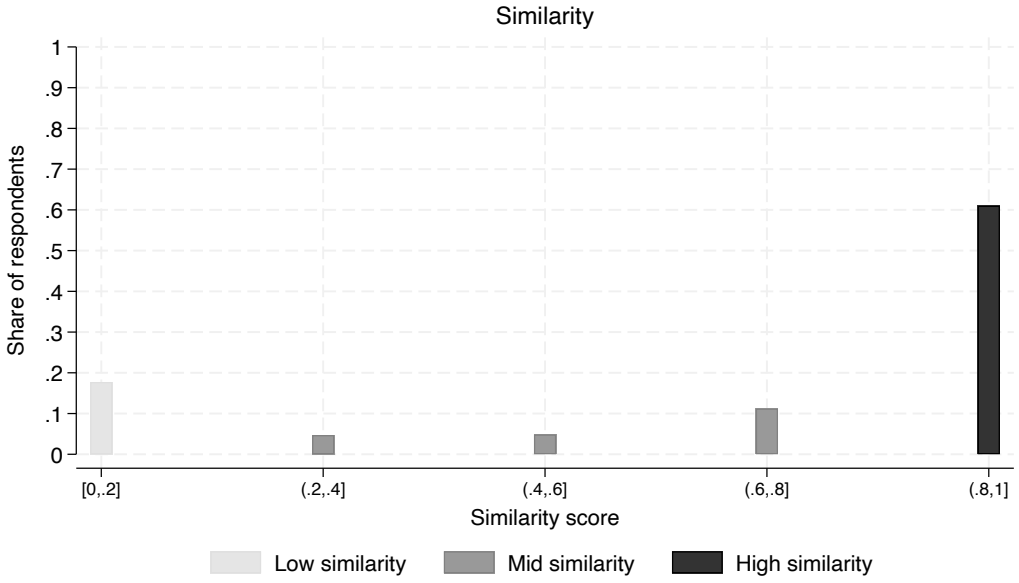


Figure 2: Histogram of subjects' maximum similarity scores across broad and narrow rubrics.

The Groups Primed by ET and CT Enemy choice seeks to increase the salience of conflict between economic or cultural groups. We check if this is the case by studying subjects' group descriptions and expected disagreement.

We instructed an RA to classify described ingroups and outgroup threats into: i) broad economic and cultural groups $E, -E$ and $C, -C$; ii) narrow enemy-based groups; iii) "no group" (meaning that ET or CT did not prime an economic or cultural group).[6] Based on the classification, we used TF-IDF to select a rubric of distinctive words for each broad and narrow group so as to minimize confounds from words common to all texts. We then tasked GPT-4o mini to compute a similarity score in [0,1] between a subject's ingroup description and a: i) broad or, ii) narrow group rubric. The subject's similarity score is the maximum of these two. See Appendix C for details.

No rubric mentions political parties, so similarity does not capture parties or their leaders. In fact, party-related words do not even appear in word clouds of subjects' texts, which instead focus on economic issues in ET and on cultural ones in CT, see Figure H.10 in the Appendix.

The similarity scores of subjects' text to the rubrics of different groups confirm that our treatments largely worked as intended. First, the ingroup texts of 77% of subjects in ET are *more* similar to the broad economic group

[6]We scaled the method by training GPT-4o mini on the RA's classification.

consistent with their enemy choice (E or $-E$) than to the other broad economic group or to any broad cultural groups, C or $-C$. In CT, the equivalent number is 81%. Similar results obtain for narrow groups.[7] In sum, for most people ET (resp. CT) primes the broad or narrow economic (resp. cultural) groups opposed along the economic or cultural conflict they make salient.

Second, few people fail to relate to an economic or a cultural group: the similarity of a subject's text to the assigned group is high, see Figure 2. Only 18% of texts are less than 0.2 similar to broad and narrow groups, while 62% of texts are more than 0.8 similar to a broad or a narrow group. Again, ET or CT prime conflict between economic or cultural groups.

Expected ingroup-outgroup disagreement is another theory relevant proxy for the success of ET and CT in priming conflicting groups. In ET (CT) the median respondent expects disagreement in 7 (8) out of 10 questions. The top quartile expects even more disagreement (9 in ET, 10 in CT), the bottom quartile perceives little disagreement (5 in ET, 6 in CT). Section 5.2 uses this variation to study heterogeneous effects of ET and CT.

Similarity and Expected Disagreement are noisy proxies of the same phenomenon, identification with broad or narrow groups, so they should be positively correlated. This is indeed the case. The correlation coefficient of Similarity and Expected Disagreement is positive with a value of 0.1742 and statistically significantly different than zero (p-value$<$0.0001).

Overall, then, for most participants ET and CT prime conflict among economic or cultural groups that disagree on many issues. We now present our model and derive its implications for treatment effects.

3 The Model

In BGT, a person identifies with her salient ingroup, causing her beliefs to move toward its ingroup stereotype. For simplicity, we focus on identity with broad groups, also consistent with their importance in Section 2.

[7]Descriptions are nearly four times more similar to the broad group consistent with the chosen enemy than to the opposite group in the same (economic or cultural) domain, and nearly six times more similar than to the groups of the non-primed domain (0.59 vs. 0.15 vs. 0.10), and four times more similar to the narrow group consistent with the chosen enemy than to the other narrow groups in the same conflict domain and ten times more similar than to the narrow groups in the non-primed domain (0.50 vs 0.11 vs 0.05).

Identity and beliefs Following Bordalo et al. (2016), the ingroup stereotype moves a person's belief on issue i by a factor proportional to the average ingroup-outgroup disagreement on i itself. Formally, let $\hat{E}$ be the partition between E and $-E$ and $\hat{C}$ that between C and $-C$. Let $d_i(\hat{X})$ be disagreement on issue i between the average progressive and conservative in the partition, $X = E, C$. By definition, $d_i(\hat{X}) > 0$ for all i because primitive beliefs are positively correlated within and across economics and culture (and uncorrelated with issue specific factors v_i). A person's identity with her ingroup in partition $\hat{X}$, has a belief on issue i in domain $b = e, c$ equal to:

$$b_i(\hat{X}) \quad = \quad b_i + x \cdot d_i(\hat{X}). \tag{3}$$

b_i is the person's undistorted belief, equal to Equation (1) for economic issues ($b = e$), to Equation (2) for cultural ones ($b = c$). Term $x = e, c$ is her primitive belief along partition $\hat{X}$. A person identified with a progressive ingroup, $x > 0$, becomes more progressive on all issues i within and across domains. This effect increases with the extent of ingroup-outgroup disagreement $d_i(\hat{X})$ in the issue. This is the critical "Kernel of Truth" logic of stereotypes, which yields predictions that we later test.[8]

In our setup, when economic conflict is salient group disagreement is equal to $d_i(\hat{E}) = d > 0$ in all economic issues i (where d is the difference in primitive economic beliefs between E and $-E$) and equal to $d_i(\hat{E}) = \rho \cdot d > 0$ in all cultural issues i. Spillovers from economics to culture are due to $\rho > 0$. With salient cultural conflict group disagreement is $d_i(\hat{C}) = \rho \cdot d > 0$ in economics and $d_i(\hat{C}) = d > 0$ in culture. In this formulation group disagreement is constant across issues i within a domain. We later relax this restriction.

In Equation (3) identity increases the importance of primitive beliefs compared to issue-specific evidence or tastes v_i. When economic conflict is salient, a progressive person retrieves the stereotypical distrust of the market by her ingroup and becomes more favorable to redistribution, even if she had been exposed to contrary facts, and vice-versa for a conservative person. In this way, salient conflict produces extremism by itself, via the salient group stereotype in Equation (3), even absent any news.[9]

[8] The distortion increases in a person's primitive beliefs e, c, which is plausible and simplifies the algebra. In BGT, this is obtained if the parameter θ capturing the strength of stereotypes increases in primitive beliefs.

[9] One foundation here is selective attention as in Bordalo et al. (2025): context prompts

A person can identify with her economic group, E or $-E$, with her cultural group C or $-C$, or be unidentified. The latter state can reflect identification with a non-political group, e.g. a sport team, or focus on a specific policy issue, e.g. gun control, neglecting groups. The unidentified person fails to perceive ingroup-outgroup disagreement, $d_i = 0$, so her beliefs are undistorted. We use subscript $\hat{E}$ for economic identity, $\hat{C}$ for cultural identity, and $\hat{\emptyset}$ for no identity.

Identity Determination The salience of conflict between groups X and $-X$, is additive in a common component $\sigma_{\hat{X}}$ and a person-specific iid extreme value shock $\epsilon_{\hat{X}}$ with mean zero and unit variance. The salience of "unidentified" is set to $\sigma_{\hat{\emptyset}} = 0$. Compared to BGT, we allow for shock $\epsilon_{\hat{X}}$ and neglect contrast-driven salience. The latter can be important in the world (implying that voters who care more about an issue are more likely to see it as salient) but it is not part of our treatments, so we abstract from it. Based on our assumptions, the probability that economic or cultural identity is most salient and hence chosen is iid across people and equal to:

$$p_{\hat{X}} = \frac{e^{\sigma_{\hat{X}}}}{1 + e^{\sigma_{\hat{E}}} + e^{\sigma_{\hat{C}}}}, \quad \text{for } \hat{X} = \hat{E}, \hat{C}, \tag{4}$$

which is in turn equal to the share of people identified along $\hat{E}$ or $\hat{C}$. The remaining people are unidentified.

Section 2.3 showed that ET and CT prime most people to strongly connect to economic or cultural ingroups, and few people not to connect to them. Based on the prevailing response, we thus view ET as increasing the salience $\sigma_{\hat{E}}$ of E and $-E$, and CT as increasing the salience $\sigma_{\hat{C}}$ of C and $-C$ (we study the unidentified people later). Thus, ET and CT change identities as follows:

$$\frac{\partial p_{\hat{X}}}{\partial \sigma_{\hat{X}}} = p_{\hat{X}} \cdot (1 - p_{\hat{X}}), \tag{5}$$

$$\frac{\partial p_{\hat{Y}}}{\partial \sigma_{\hat{X}}} = -p_{\hat{X}} \cdot p_{\hat{Y}}, \tag{6}$$

for $\hat{Y} \neq \hat{X}$. Priming cultural enemies, higher $\sigma_{\hat{C}}$, increases cultural identity $p_{\hat{C}}$ in Equation (5), but it also reduces the economic and no-identity shares, $p_{\hat{E}}$ and ($p_{\emptyset}$ in Equation 6). Interference of cultural conflict with economic conflict

categorization of policy choices as "cultural-conflict problems", causing focus on culture, or as "economic-conflict problems", causing focus on economics, not on them together.

(and vice-versa) amplifies belief instability, with important implications.

The Effect of ET and CT on Beliefs A fraction $p_{\hat{E}}$ of people with primitive beliefs (e, c) is economically identified, a fraction $p_{\hat{C}}$ is culturally identified, the rest is unidentified. Let $\tilde{e}$ and $\tilde{c}$ be average economic and cultural beliefs, respectively, for person (e, c). Using Equation (3) and averaging the issue-specific factor υ_i to zero one finds that:

$$\tilde{e} = e \cdot (1 + p_{\hat{E}} \cdot d) + p_{\hat{C}} \cdot c \cdot \rho \cdot d, \tag{7}$$

$$\tilde{c} = c \cdot (1 + p_{\hat{C}} \cdot d) + p_{\hat{E}} \cdot e \cdot \rho \cdot d, \tag{8}$$

Beliefs reflect primitives (e, c) and latent group conflict $d > 0$, both within the identity domain (captured by $p_{\hat{E}}$ for $\tilde{e}$ and $p_{\hat{C}}$ for $\tilde{c}$), and across domains based on correlation $\rho > 0$ (captured by $p_{\hat{C}}$ for $\tilde{e}$ and $p_{\hat{E}}$ for $\tilde{c}$).

By Equations (5) and (6), increasing the salience of economic conflict $\sigma_{\hat{E}}$ in ET changes beliefs relative to Control as follows:

$$\frac{\partial \tilde{e}}{\partial \sigma_{\hat{E}}} = p_{\hat{E}} \cdot [e \cdot (1 - p_{\hat{E}}) - c \cdot p_{\hat{C}} \cdot \rho] \cdot d, \tag{9}$$

$$\frac{\partial \tilde{c}}{\partial \sigma_{\hat{E}}} = p_{\hat{E}} \cdot [e \cdot (1 - p_{\hat{E}}) \cdot \rho - c \cdot p_{\hat{C}}] \cdot d. \tag{10}$$

These equations show the key forces in our model. First, as conflict between two groups becomes more salient, their latent disagreement is polarized in beliefs. For ET, this is captured by the first term in square brackets, which causes economic progressives ($e > 0$) to become even more progressive, in economics $\tilde{e}$ and culture $\tilde{c}$, and the opposite for conservatives ($e < 0$).

Second, as conflict between two groups becomes more salient, the alternative latent disagreement de-polarizes in peoples' beliefs. For ET, this is captured by the second term in square brackets, which causes cultural progressives ($c > 0$) to become less progressive in economics $\tilde{e}$ and culture $\tilde{c}$ and the opposite for conservatives ($c < 0$).

The overall effect of ET on beliefs $\tilde{e}$ and $\tilde{c}$ combines these two forces, which analogously shape beliefs in CT. We now study two key consequences: polarization of latent disagreement among salient groups in Section 4 and realignment away from no longer salient groups in Section 5.

4 Polarization and Salient Conflict

Section 4.1 shows that ET and CT increase population-wide polarization. Section 4.2 studies how ET and CT polarize latent groups. Section 4.3 shows that, consistent with kernel of truth, belief polarize more in issues i where groups disagree more in Control.

4.1 Population Wide Polarization

In our model, ET (CT) should cause members of opposing economic (cultural) groups to become more extreme in economics and culture, which tends to boost overall polarization. The same treatment, however, dampens the non-primed cultural (economic) identity, reducing extremism. Appendix A finds conditions under which the former effect dominates.

Prediction 1. *If the correlation between e and c is high enough:*

$$\rho^2 > \max[\frac{p_{\hat{C}}}{1-p_{\hat{E}}}, \frac{p_{\hat{E}}}{1-p_{\hat{C}}}], \tag{11}$$

then ET *and* CT *should boost the variance of* $\tilde{e}$ *and* $\tilde{c}$.

Large ρ implies that the spillover of ET on culture and of CT on economics is strong enough to offset the de-identification of people along the non-primed conflict. Condition (11) is easier to meet if more people are unidentified at baseline ($p_{\hat{E}}$ and $p_{\hat{C}}$ are low): de-identification is weaker in this case.

Test of Prediction 1 Table 3 shows the standard deviation of beliefs across treatments, for aggregate indices $(\tilde{e}, \tilde{c})$, factual $(\tilde{e}_f, \tilde{c}_f)$ and normative $(\tilde{e}_n, \tilde{c}_n)$ beliefs. As expected, ET and CT increase disagreement in all indices. The effects are statistically significant and large, with magnitudes ranging from 2 to 10% of baseline disagreement.

Three messages follow. First, salient conflict causally boosts polarization via a context effect: our treatments do not convey information and the specific enemy conflicts are barely related to many beliefs we elicit. Second, there are strong spillovers: ET polarizes beliefs on a range of economic issues, cultural issues, and across economics and culture, and the same for CT. In our model, this arises because groups exhibit latent disagreement in both economics *and*

Standard Deviation in Beliefs				
	$\sigma_{\tilde{e}}$	$\sigma_{\tilde{c}}$	σ_f	σ_n
Control	1.99	2.43	2.03	2.30
ET	2.12***	2.54***	2.13***	2.41***
CT	2.18***	2.49**	2.15***	2.42***

Table 3: Standard deviations of average beliefs in economics, $\sigma_{\tilde{e}}$, culture, $\sigma_{\tilde{c}}$, facts, σ_f, and policies, σ_n, by treatment. *** (**) indicate statistically significant differences between ET or CT and Control at 0.01 (0.05) level.

culture, as documented in Figure 1. Third, we see large treatment effects on factual beliefs. When prompted to think about, say, a conflict with economic lobbies, a person believes that income inequality has increased more and that the share of crimes committed by immigrants is lower, even if no facts are provided. The reason is that salient conflict ignites primitive beliefs, dampening the role of prior information. This can explain the finding in Kahan (2015) that scientifically savvy people have more accurate beliefs than non savvy ones in many issues but not in politically charged ones. Group conflict can cause over-weighting of primitive beliefs at the expense of information. Such distortions are not due to fixed heuristics: their intensity and direction changes endogenously with the salience of specific conflicts.

One alternative explanation for the increase in polarization is that our treatments make social disorder salient, not group membership, or make people angry, affecting beliefs from mechanisms other than identity. Jost et al. (2003) show that people perceiving strong conflict are more conservative, Algan et al. (2025) that anger also prompts conservatism. If some respondent become more conservative, they could increase population-wide disagreement. Salient conflict may also cause distraction and noisy answers, increasing disagreement. Compared to these mechanisms, our model prediction – which we now test – is that salient conflict should polarize beliefs along latent group disagreement.

4.2 Polarization of Latent Group Disagreement

In our model, overall disagreement is caused by the polarization of latent conflicts, which boosts disagreement among members of primed groups compared to the baseline. We prove the following result.

Prediction 2 *If some people are unidentified at baseline,* $p_{\hat{C}} + p_{\hat{E}} < 1$*, ET*

polarizes the average belief of economic groups in each issue i in the direction of the group's disagreement in Control; CT does the same for cultural groups:

$$\frac{\partial \tilde{b}_{i,-E}}{\partial \sigma_{\hat{E}}} < 0 < \frac{\partial \tilde{b}_{i,E}}{\partial \sigma_{\hat{E}}} \ , \ \ \frac{\partial \tilde{b}_{i,-C}}{\partial \sigma_{\hat{C}}} < 0 < \frac{\partial \tilde{b}_{i,C}}{\partial \sigma_{\hat{C}}} \ , \ \ b = e, c$$

Test of Prediction 2 Prediction 2 should hold because it rests on a weaker condition than (11). To test it, we compare the beliefs of people who signaled, through enemy choice, membership in the same group in ET or CT or Control. This is a proper test if enemy choice is not affected by ET or CT themselves, which Table 1 shows to be indeed the case.

We test Prediction 2 by averaging and standardizing a subject's belief over the 10 questions, giving a general progressivity index $\tilde{b}$, and by estimating, separately for each treatment XT and Control, X=E,C, the regression:

$$\tilde{b} = \alpha_0 + \alpha_1 \cdot X + \alpha_2 \cdot \mathrm{XT} + \alpha_3 \cdot X \cdot \mathrm{XT} + \epsilon \ , \ \ X = E, C$$

where X is a dummy for the group and XT for the treatment. Coefficient α_1 captures latent disagreement between the primed progressive and conservative groups in Control. α_2 captures the belief change of the primed *conservative* group between XT and Control. Coefficient α_3 captures the effect of XT on disagreement between primed progressives and conservatives. By Prediction 2 the interaction of dummies should then have the same positive sign of the group dummy, $\alpha_1, \alpha_3 > 0$: By making group conflict salient, XT polarizes their disagreement compared to Control. The model also predicts that $\alpha_2 < 0$: treatment XT should increase the conservatism of $-X$.

We report regression results in Table 4. In Column (1), the sample includes ET and Control, in Column (2) it includes CT and Control. In Columns (3) to (6), we pool the observations from Columns (1) - (2) to show robustness wave-by-wave and to adding demographic and political controls interacted with the dummy for treatment XT.[10]

Prediction 2 is confirmed in all columns. The coefficients on E and C are positive: these groups are baseline more progressive than $-E$ and $-C$. Consistent with our model, then, also the interaction coefficients are positive and statistically significant: both ET and CT polarize the latent conflict between primed progressives and conservatives. The ET or CT dummy coefficients are

[10]We list the controls in Footnote 3.

Outcome	OLS regressions General Progressivity (Beliefs, $\tilde{e}$, $\tilde{c}$, averaged and standardized)					
Treatment	ET (X = E)	CT (X = C)	ET and CT (X = E,C)			
Waves	All	All	1	2	3	All
	(1)	(2)	(3)	(4)	(5)	(6)
X	0.997***	1.177***	1.121***	1.045***	1.094***	0.680***
	(0.025)	(0.023)	(0.038)	(0.036)	(0.034)	(0.020)
XT	-0.128***	-0.0144	-0.0701*	-0.0751**	-0.0664*	
	(0.026)	(0.025)	(0.037)	(0.035)	(0.035)	
X XT	0.199***	0.117***	0.161***	0.183***	0.132***	0.138***
	(0.036)	(0.034)	(0.049)	(0.048)	(0.046)	(0.027)
Controls	No	No	No	No	No	Yes
Adj. R^2	0.298	0.381	0.348	0.325	0.340	0.495
Clusters	8738	8708	4286	4094	4580	12954

Table 4: In Columns (1) and (2) we regress general progressivity (i.e. averaged and standardized beliefs) on a progressive-group-membership dummy ($X = E$ in ET, $X = C$ in CT), a treatment dummy (ET, CT), and the interaction of the two dummies. Columns (3) to (6) pool these regressions (with standard errors clustered at the individual level) to present the diff-in-diffs wave by wave (Columns (3) to (5)) and when adding demographic and political controls and their interactions with the treatment dummies (Column (6)). Robust standard errors in brackets. *** (**,*) indicate statistical significance at 0.01 (0.05, 0.1) level.

negative, showing that disagreement increases conservatives and progressive change their beliefs in opposite directions.[11]

Polarization is large. In Column (1), ET increases group disagreement by 20% (= .199/.997) relative to baseline disagreement in Control. In Column (2), CT does so by 10% (= .117/1.177). The increase in population-wide disagreement is thus tied to latent group conflict. If we break down the regression into economic facts and policies and cultural facts and policies, we see positive group dummies and interactions (see Table H.11 in the Appendix). Treatment effects are highly significant, with the exception of cultural policy preferences in CT, and range from 8 to 35% of disagreement between progressives and conservatives in Control.

Important, both ET and CT increase belief coherence: the covariance of subjects' beliefs increases within as well as across economics and culture, in

[11]This effect is significant in ET, not in CT, where polarization mostly ties to progressives getting more extreme. We offer an explanation for this fact in Section 5.2.

both facts and policies (see Appendix Table H.10). As in our model, then, salient conflict separates voters into highly polarized belief systems.

One may think that ET and CT cause polarization by priming parties. Our treatments do not mention parties or their jargon, and text shows that described ingroups are not tied to parties. We can directly assess party-identity using measured political affiliation. If ET and CT cue membership in D or R, including these dummies in our regressions should annihilate the explanatory power of social groups. Column (6), however, shows that our results are very robust to controlling for a range of demographics, including partisanship and the interactions of these controls with the treatment dummy. The polarizing effects of ET and CT are very robust, they are not due to parties.[12]

4.3 Kernel of Truth Polarization

In our model extremism increases in actual disagreement d_i between the primed ingroup and outgroup on issue i, see Equation (3). This property is due to the "kernel of truth" logic: stereotypes exaggerate real group differences (Bordalo et al., 2016). Thus, ET and CT should cause stronger polarization on issues where groups disagree more. If disagreement between E and $-E$ (C and $-C$) is not constant at d but varies across issues i we obtain the result below.[13]

Prediction 3 *Let $\tilde{b}_{iP}^{K}$ be the belief over issue i by the average member of primed group $P = X, -X$ in condition $K = \text{XT}$, Control. There is a $\mu^{XT} > 0$ with* X=E,C *such that*:

$$\tilde{b}_{i,X}^{\text{XT}} = \tilde{b}_{iX}^{\text{Control}} + \mu^{\text{XT}} \cdot (\tilde{b}_{i,X}^{\text{Control}} - \tilde{b}_{i-X}^{\text{Control}}) \quad (12)$$

$$\tilde{b}_{i,-X}^{\text{XT}} = \tilde{b}_{i,-X}^{\text{Control}} + \mu^{\text{XT}} \cdot (\tilde{b}_{i,-X}^{\text{Control}} - \tilde{b}_{iX}^{\text{Control}}) \quad (13)$$

Salient economic (cultural) conflict should cause each economic (cultural) group to become especially extreme in issues where the disagreement with the outgroup is larger, compared to issues where it is lower. This occurs precisely

[12]Table H.12 also finds very similar results when restricting separately to Independents, Democrats, Republicans, where ET and CT cannot prime party affiliation.

[13]To do so, we generalize Equations (1) and (2) with issue-specific loadings, so that $b_i = \pi_i \cdot b + \upsilon_i$ for progressives ($b > 0$) and $b_i = \kappa_i \cdot b + \upsilon_i$ for conservatives ($b < 0$), where $\pi_i, \kappa_i > 0$ are uncorrelated with υ_i. Variation of loadings across groups creates independent issue variation in disagreement across i. We could generalize the model further by allowing loadings to differ across economic and cultural groups.

because polarization builds on a latent conflict that varies across issues.[14]

This prediction is testable: in the Control sample, groups disagree more on some issues than in others. Disagreement between E and $-E$ ranges from 1.25 to 2.41 points on economics and from 1.95 to 2.68 points in culture. Disagreement between C and $-C$ ranges from 1.22 to 2.60 points in economics and from 2.46 to 3.44 points in culture. We then estimate the following regression for a group's average belief $\tilde{b}^{\text{XT}}_{G,i}$ in issue i as measured in treatment XT:

$$\tilde{b}^{\text{XT}}_{G,i} = \theta_0 + \theta_1 \cdot \tilde{b}^{\text{Control}}_{G,i} + \theta_2 \cdot (\tilde{b}^{\text{Control}}_{G,i} - \tilde{b}^{\text{Control}}_{-G,i}) + \epsilon_i \ , \tag{14}$$

where $G, -G = X, -X$ and $-G \neq G$. Prediction 3 implies $\theta_2 > 0$, namely holding the group's average belief in Control in issue i constant, the effect of ET or CT on the salient group's average belief is stronger in issues where disagreement with the outgroup is larger. Compared to the diff-in-diff of Table 4 this regression still detects whether ET and CT amplify baseline disagreement, which occurs when $\theta_1 + 2 \cdot \theta_2 > 1$, but it additionally tests whether such amplification is stronger in issues where baseline disagreement is larger. This test is easily implemented also at the enemy-level (with the ingroup comprising people who chose the same enemy and the outgroup comprising people who chose a different enemy) offering a robustness of our results to the possibility that ET and CT may prime narrow rather than broad groups.

Figure 3 reports in the left panels the estimated coefficients θ_2 of regression Equation (14) for broad groups in ET (in blue) and CT (in pink).[15] There are 20 data points, corresponding to progressive and conservative beliefs in each of the 10 questions. The right panels report the same regression estimated for narrow groups in ET and CT (same color coding). Now we have 100 data points, because there are 10 enemy groups for each of the 10 questions.[16]

We see a sharp and highly statistically significant gradient: Under both ET and CT, group beliefs load approximately by 1 on ingroup beliefs in Control. Crucially, and consistent with our model, there is a positive and significant loading on group disagreement: ET and CT cause a larger increase in ex-

[14] In Prediction 3 we set μ^{XT} constant, despite the fact that it should be larger on economic issues than cultural issues for ET (and vice-versa for CT), because we have only five questions in each domain so we cannot reliably test this property.

[15] See Columns (1)-(2) and (4)-(5) in Table H.13 for the full regression results.

[16] More precisely, there are 6 enemy groups in ET or CT in waves 1 and 2 and there are 4 enemy groups per treatment in wave 3.

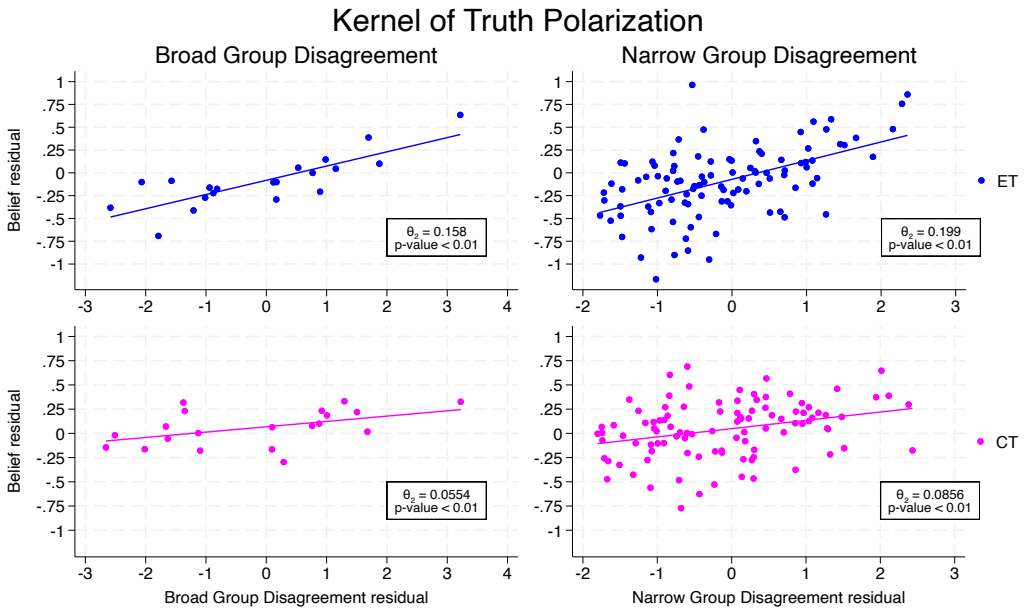


Figure 3: We plot group-issue level belief residuals against group disagreement residuals. The slope corresponds to the estimated θ_2 coefficient of regression Equation (14), for broad groups (left panels) and narrow groups (right panels).

tremism on issues in which the opposing social groups disagree more sharply in Control. A one point higher disagreement is associated with higher polarization between .16 and .20 points in ET and between .06 and .09 points in CT. This effect goes in opposite direction for the beliefs of progressives and conservatives, doubling the effect on polarization.[17] This result confirms our mechanism: salient conflict causes people to slant their belief to the salient ingroup stereotypes, causing a larger increase in extremism on issues where latent group disagreement is stronger.

5 Belief Realignment

The second implication of our model is that changing the salience of different conflicts should realign beliefs towards those of the salient ingroup, and away from those of ingroups which are no longer salient. We now show that this force can shape: i) which groups are more polarized - whether economic or cultural groups (Section 5.1); and ii) the overall extent of polarization (Section 5.2).

[17]The estimates confirm the conflict-amplification: both ET and CT entail $\theta_1 + 2 \cdot \theta_2 > 1$. A placebo regression based on our Control in the Appendix, see Columns (3) and (6) in Table H.13, gives the expected coefficient of 1 for the ingroup mean and 0 for the disagreement term.

5.1 Economic versus Cultural Conflict

Increasing the salience of economic compared to cultural conflict (i.e pointing to an economic rather than to a cultural enemy) changes which groups are more polarized. Upon a switch from CT to ET, disagreement between economic and cultural groups changes as follows:

$$\Delta^{\text{CT}\to\text{ET}}(\tilde{b}_E - \tilde{b}_{-E}) \propto (p^{\text{ET}}_{\hat{E}} - p^{\text{CT}}_{\hat{E}}) + \rho \cdot (p^{\text{ET}}_{\hat{C}} - p^{\text{CT}}_{\hat{C}}), \quad (15)$$

$$\Delta^{\text{CT}\to\text{ET}}(\tilde{b}_C - \tilde{b}_{-C}) \propto \rho \cdot (p^{\text{ET}}_{\hat{E}} - p^{\text{CT}}_{\hat{E}}) + (p^{\text{ET}}_{\hat{C}} - p^{\text{CT}}_{\hat{C}}), \quad (16)$$

where $\tilde{b}_G$ is the general progressivity of group G (i. e beliefs averaged over all issues and all members of group G), while $p^{\text{XT}}_{\hat{E}}$ and $p^{\text{XT}}_{\hat{C}}$ are the shares of respondent who are economically and culturally identified in treatment XT.

By Equation (15), a switch from CT to ET *directly* increases disagreement between opposite economic groups by fostering economic identity, $(p^{\text{ET}}_{\hat{E}} - p^{\text{CT}}_{\hat{E}}) > 0$, but *indirectly* lowers disagreement – also among economic groups – by reducing cultural identity, $\rho \cdot (p^{\text{ET}}_{\hat{C}} - p^{\text{CT}}_{\hat{C}}) < 0$. The indirect effect depends on ρ, which shapes the extent to which cultural groups are correlated with economic ones. For $\rho < 1$, then, the direct effect is ceteris stronger than the indirect one: the switch from CT to ET tends to foster disagreement between E and $-E$ because it causes many unidentified people to become economically identified. The same effects are at play in Equation (16) for disagreement among cultural groups. Critically, now the direct effect is negative and the indirect one is positive, so $\rho < 1$ implies that a switch from CT to ET ceteris paribus dampens conflict between C and $-C$.

This logic implies that changes in the salience of specific conflicts can polarize some groups and de-polarize others. Compared to the polarization Predictions 1, 2 and 3, which rest on ρ being large, de-polarization is only significant if ρ is moderate. In the data, $\rho < 1$, yielding the following realignment.

Prediction 4 *Moving from* CT *to* ET *reduces disagreement between cultural groups compared to disagreement between economic groups,* $\Delta^{CT\to ET}[(\tilde{b}_C - \tilde{b}_{-C}) - (\tilde{b}_E - \tilde{b}_{-E})] < 0$. *The switch from* CT *to* ET *dampens disagreement between opposite mixed types* $(E, -C)$ *and* $(-E, C)$*, and ambiguously affects disagreement between opposite pure types* (E, C) *and* $(-E, -C)$.

Imperfect correlation between economic and cultural beliefs implies that

	$(\tilde{b}_E - \tilde{b}_{-E})$ (1)	$(\tilde{b}_C - \tilde{b}_{-C})$ (2)	$(\tilde{b}_E - \tilde{b}_{-E}) - (\tilde{b}_C - \tilde{b}_{-C})$ (3)
CT	1.172^{***}	1.293^{***}	-0.122^{***}
ET	1.197^{***}	1.214^{***}	-0.018
ET−CT	0.025	-0.079^{**}	0.104^{***}

Table 5: We report average differences in beliefs between $X = E, C$ and $-X$ in CT and ET in the first two rows and columns. In the first two rows of the third column, we report the difference in differences across economic and cultural groups separately for each treatment. In row three, we report differences between treatments. *** (**) indicate statistically significant difference from zero at 0.01 (0.05) level.

changes in salience can affect the dimension, economic or cultural, defining which groups that are more polarized in their overall views, $\tilde{b}$. Realignment mostly occurs via mixed types. By increasing economic identity and reducing cultural identity, a switch from CT to ET boosts the general conservatism of type $(-E, C)$ and the general progressivity of type $(E, -C)$. This change reduces disagreement between them because in CT $(-E, C)$ is more overall progressive than $(E, -C)$. Critically, the stronger conservatism of type $(-E, C)$ increases the conservatism of broad groups $-E$ and C, while the stronger progressivity of type $(E, -C)$ increases the progressivity of broad groups E and $-C$. Thus, the belief change of mixed types causes economic groups to polarize, cultural ones to de-polarize.

By contrast, pure types do not necessarily favor realignment, in fact they may obstruct it. When moving from CT to ET higher economic identity boosts the general progressivity of type (E, C) but lower cultural identity dampens it, and vice versa for type $(-E, -C)$. Disagreement among pure types and between the broad groups they belong to is ambiguously affected.

Test of Prediction 4 We first test whether a switch from CT to ET reduces disagreement between cultural groups compared to disagreement between economic groups. Table 5 reports: i) disagreement between economic and cultural groups in CT and ET (first and second rows), ii) the effect of the switch from CT to ET on such disagreement (third row), and iii) the effect of the switch on disagreement between economic vs cultural groups (the southeast cell). The model predicts that the diff in diff in iii) should be positive, and under stricter conditions it causes absolute disagreement between C and $-C$ to drop.

When moving from CT to ET, disagreement among economic groups in-

Outcome	OLS regressions General Progressivity	
Treatments	ET and CT	
Types	Pure	Mixed
	(1)	(2)
C	1.677***	0.225***
	(0.027)	(0.045)
ET	-0.0714**	0.0356
	(0.029)	(0.042)
C ET	0.00524	-0.199***
	(0.038)	(0.063)
Adj. R^2	0.563	0.0103
Obs.	6149	2325

Table 6: We regress averaged and standardized beliefs $\tilde{b}$ on a dummy for being a cultural progressive C, an ET dummy, and the interaction of the two. In Column (1) the regression includes pure types and in Column (2) only mixed types, respectively. Robust standard errors in brackets. *** (**,*) indicate statistical significance at 0.01 (0.05, 0.1) level.

creases, but not significantly (Column (1)), while disagreement between cultural groups decreases significantly (Column (2)). Consistent with the model, the positive diff-in-diff coefficient shows that cultural groups depolarize compared to economic ones. The effect is about 10% of average group disagreement. A switch from CT to ET causes a realignment, making culture relatively less important than economics as the dimension defining which groups are polarized between progressives vs conservatives.

Table 6 tests for the role of mixed and pure types in Prediction 4. We regress a subject's general progressivity $\tilde{b}$ on a dummy for C, one for ET, and an interaction of the two, in the sample of CT and ET respondents (we thus omit the Control sample). Column (1) considers only pure types, Column (2) only mixed ones. For pure types the estimated coefficient on the dummy variable C measures the difference in general progressivity between (E, C) and $(-E, -C)$ in the CT sample. For mixed types it measures the difference in general progressivity between $(-E, C)$ and $(E, -C)$, again in CT. The estimated coefficient on the interaction measures the diff-in-diff variable of interest: the difference in general progressivity between the same types across ET and CT.

Consistent with Prediction 4, de-polarization of cultural conflict is *entirely* due to mixed types: the negative interactive coefficient in Column (2) says

that, compared to CT, ET makes $(-E, C)$ generally more conservative than $(E, -C)$ by about 0.2, almost fully reversing the 0.22 higher progressivity of the former type in CT. Nothing significant instead happens to disagreement among pure types in Column (1), again consistent with Prediction 4.[18]

We can perform a kernel of truth test for mixed and pure types, which also allows us to test whether disagreement among mixed types drops *more* on issues in which their baseline disagreement in CT is larger. For mixed types, following Equations (12) and (13) we estimate the following regressions:

$$\tilde{b}^{\text{ET}}_{(E,-C),i} = \theta_0 + \theta_1 \cdot \tilde{b}^{\text{CT}}_{(E,-C),i} + \theta_2 \cdot (\tilde{b}^{\text{CT}}_{E,i} - \tilde{b}^{\text{CT}}_{-E,i}) + \epsilon_i \quad (17)$$

$$\tilde{b}^{\text{ET}}_{(-E,C),i} = \theta_0 + \theta_1 \cdot \tilde{b}^{\text{CT}}_{(-E,C),i} + \theta_2 \cdot (\tilde{b}^{\text{CT}}_{-E,i} - \tilde{b}^{\text{CT}}_{E,i}) + \epsilon_i, \quad (18)$$

which predict a type's beliefs in ET conditional on its beliefs in CT and on the disagreement between its economic ingroup and outgroup in the CT baseline. We estimate analogous regressions for pure types.

Using the estimated regressions we predict disagreement in ET over issue i between mixed types $(E, -C)$ and $(-E, C)$, and between pure types (E, C) and $(-E, -C)$. The left hand panel of Figure 4 reports predicted (y axis) against CT-baseline (x axis) disagreement for mixed types (black "+" symbols). The right panel reports the same for pure types. The 45 degree line identifies issues in which disagreement does not change when moving from CT to ET. For completeness, the Figure reports also similarly constructed predicted and baseline disagreement for mixed and pure narrow types (i.e. based on the specific enemy choice) using gray "x" symbols (see Appendix D for details on how we construct this).

On the right panel, points lie on the 45 degree line: kernel of truth confirms that the switch from CT to ET does not change disagreement among pure types. On the left panel, points instead lie above the 45 degree line: the switch from CT to ET increases the progressivity of economically progressive mixed types compared to economically conservative ones, reducing their disagreement compared to CT. Consistent with kernel of truth, this moderation is stronger

[18] If we decompose the effect into transitions from CT to Control and from Control to ET we see that both matter to reduce cultural identity. The switch from CT to Control depolarizes cultural groups, being the opposite of the effect in Column (3) of Table 4. The switch from Control to ET is associated with unchanged disagreement among cultural groups, $\Delta^{\text{ET,Control}}(\tilde{b}_C - \tilde{b}_{-C}) = 0$, which in Equation (16) signals that this transition reduces cultural identity, $\Delta^{\text{ET,Control}} p_{\hat{C}} < 0$, offsetting the first positive term.

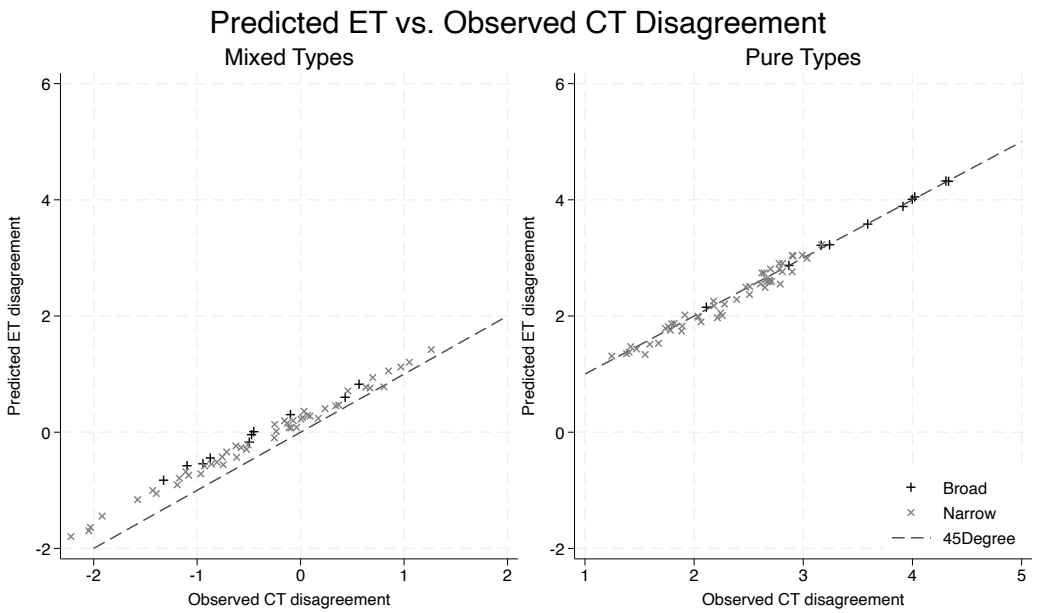


Figure 4: We predict average beliefs for each issue i in ET based on the corresponding beliefs in CT and disagreement in CT, separately for pure and mixed types, and for broad and narrow groups. Based on predicted beliefs, we compute predicted disagreement between economic progressives and conservatives, separately for pure and mixed types, and broad and narrow groups. We plot predicted disagreement in ET against disagreement in CT.

in issues i where baseline disagreement is larger (i.e. more negative).[19]

Overall, then, an increase in the salience of a specific conflict is not merely a "polarization booster": it affects the dimension of polarization, de-polarizing no longer salient cleavages. This realignment is especially due to types with "incongruent" memberships and can have important real world implications. As described in Gethin et al. (2022), in the US and in several Western democracies, culturally conservative lower class voters - intuitively linked to $(E, -C)$ - now support right wing parties, while highly educated and progressive affluent voters - intuitively linked to $(-E, C)$ - have turned to the left. Frank (2004) describes the rightward realignment of religious working class voters as a consequence of political conflict on abortion. Gennaioli and Tabellini (2025) present a model where growing salience of cultural conflict causes voters in $(E, -C)$ to switch from a progressive economic identity to a conservative cul-

[19]If we regress predicted disagreement in ET on actual disagreement in CT for broad mixed types we find a positive and statistically significant intercept (of value 0.31 with p-value <0.001) and a positive coefficient (of value 0.85) that is statistically significantly below 1 (p-value <0.01). For pure types, the intercept barely differs from zero (with a value of 0.06 and corresponding p-value of 0.096) and the coefficient (of value 0.99) on the actual disagreement in CT is not significantly different from 1 (p-value=0.14)

tural one, causing them to become politically "right-wing", and the same shock causes voters in $(-E, C)$ to become "left-wing". By making economic conflict salient, our experiment induces a small realignment in the opposite direction. This result is not explained by changing party identity. Mixed types do not have coherent party views. They switch from a generally conservative to a generally progressive position when different social conflicts are salient.

5.2 Expected Disagreement and Polarization

In our model a person slants her beliefs toward her ingroup stereotype, which in turn increases in ingroup-output disagreement. The same person can then be very polarized when identified with an ingroup featuring strong disagreement and little polarized when disagreement is weak. As a result, making a policy conflict salient does not necessarily boost extremism if such conflict is not associated with groups that sharply disagree across the board. For instance, making an external threat salient could re-frame conflict away from domestic groups, de-polarizing the latter. Compared to Section 5.1, this kind of belief realignment changes not just the locus but also the overall level of polarization.

We can test for this effect using measured expected disagreement and similarity. Both proxies capture identity with economic or cultural groups that disagree on many issues, acting as strong polarizing forces. Our model implies that respondents expecting high ingroup-outgroup disagreement should exhibit strong polarization under ET and CT compared to Control. For subjects expecting little disagreement, instead, ET and CT should not amplify disagreement relative to Control, or even dampen disagreement.[20] Similarly, respondents whose descriptions are similar to conflicting economic or cultural groups should exhibit strong polarization, while this should not occur (or the reverse should occur) for low similarity respondents. We indeed saw that similarity of text to economic and cultural groups and expected disagreement are positively correlated in the data.

We split subjects in ET and CT into those expecting ingroup-outgroup disagreement on at least three questions, and those expecting it on less than three questions (3 is median expected disagreement in Control). We call the former

[20] As we further discuss below, for people for whom ET and CT prime group with little disagreement our treatments effectively increase the salience of no identity $\sigma_\emptyset$, switching off conflicting economic and cultural identities, $(p_{\hat{E}}^{\text{XT}} - p_{\hat{E}}^{\text{Control}}) < 0$ and $(p_{\hat{C}}^{\text{XT}} - p_{\hat{C}}^{\text{Control}}) < 0$, in Equations (15) and (16).

Outcome:	OLS regressions General Progressivity			
Treatments:	ET and CT (X = E,C)			
Sample:	LED	HED	Low	High
	(1)	(2)	(3)	(4)
X	0.925***	1.100***	1.085***	1.085***
	(0.019)	(0.021)	(0.021)	(0.021)
XT	0.229***	-0.119***	0.185***	-0.209***
	(0.042)	(0.021)	(0.032)	(0.022)
X XT	-0.360***	0.229***	-0.484***	0.456***
	(0.061)	(0.028)	(0.049)	(0.030)
Adj. R^2	0.252	0.357	0.287	0.390
Clusters	5436	12010	5985	9696

Table 7: The dependent variable is average and standardized beliefs $\tilde{b}$. In Columns (1) and (2) we respectively compare treated LED participants and treated HED participants with their corresponding synthetic Controls. In Columns (3) and (4), we respectively compare treated Low Similarity participants and High Similarity participants with the Control. The regressions pool ET, CT and Control, as in Columns (3)-(6) in Table 4. Robust standard errors in brackets. *** (**,*) indicate statistical significance at 0.01 (0.05, 0.1) level.

group High Expected Disagreement (HED), the latter group Low Expected Disagreement (LED). When comparing each group to Control, one concern is that HED people may have more extreme primitive beliefs than LED ones. We thus create synthetic Control groups, comparing HED to Control subjects located in the same percentiles of HED subjects in ET or CT, and analogously for LED.[21] In this way, we compare HED respondents to a more extreme Control than LED ones.

Table 7 estimates our diff in diff regression for LED in Column (1) and HED in Column (2). Consistent with our model, Column (2) shows a much stronger treatment effects for HED than in our baseline: pooling ET and CT, polarization increases by 21% among HED compared to 14% in the baseline. Column (1) shows de-polarization for LED, who exhibit *moderated* beliefs relative to comparable people in Control by about 39%, as captured by the negative interactive coefficient.[22]

[21]This is appropriate in our model, where observed beliefs are more extreme than primitive beliefs, but preserve the ranking between different types (e, c). Details on how we construct the synthetic Control are reported in Appendix E.1.

[22]In Appendix E.2, we study individual-level variation in expected disagreement. In both ET and CT, treated progressives (conservatives) are more progressive (conservative) on

Outcome	OLS Regression Average Beliefs: $\tilde{e}$ and $\tilde{c}$							
Treatments	ET and CT							
Ingroup	Broad				Narrow			
Sample	LED	HED	Low	High	LED	HED	Low	High
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Ingroup-Outgroup Beliefs	-0.116**	0.129***	-0.105***	0.208***	-0.113**	0.189***	-0.117***	0.279***
	(0.049)	(0.023)	(0.033)	(0.024)	(0.046)	(0.024)	(0.039)	(0.025)
Adjusted R^2	0.77	0.99	0.93	0.99	0.64	0.97	0.82	0.97
N	40	40	40	40	200	200	200	200

Table 8: We report the kernel of truth regression Equation (14) separately estimated for broad groups, Columns (1)-(4), or narrow groups, Columns (5)-(8), for ET participants who are LED, Columns (1) and (5), or HED, Columns (2) and (6), and for ET participants who have Low Similarity, Columns (3) and (7), or High Similarity, Columns (4) and (8). Robust standard errors in brackets. *** (**) indicate statistical significance at 0.01 (0.05) level.

One possible explanation for these results is that LED subjects are less attentive and hence are more likely to respond at random or in the middle of the response scale. This seems unlikely. First, all subjects had to pass an attention check in order to remain in the survey (following our preregistration). Second, respondents' survey time, a standard proxy for survey quality, is not correlated with LED/HED status (see Table H.14 in the Appendix). In our model the behavior of LED arises because the treatment primes them to think about groups that do not disagree over a broad set of issues. In doing so, it switches off their more broadly polarized identities in the baseline.

We can test this mechanism also using using our similarity proxy, which directly measures the extent to which subjects' descriptions resemble highly conflicting economic or cultural groups. We separate subjects into "High" and "Low" similarity, and expect stronger treatment effects for the former. Table 7 confirms this prediction showing that the interactive coefficient for High similarity people is large, corresponding to a 41% amplification of conflict compared to Control, and High respondents, while it is negative for Low respondent, corresponding to a 44% moderation.

Table 8 repeats these tests for our kernel of truth Equations 12 and 13. Table 8 separates respondents into HED and LED in Columns (1), (2), (5), (6) and into Low vs High in Columns (3), (4), (7), (8). For both HED and High respondents, beliefs become more extreme the larger baseline disagreement

issues in which they expect disagreement than when they do not expect disagreement—consistent with our theory. Importantly, this pattern is absent in the Control, as our theory predicts, since expected disagreement is measured there for placebo groups not based on existing political conflict with associated stereotypes.

between economic or cultural groups. The reverse is true for LED and Low similarity respondents, confirming their treatment induced moderation.

Tables 7 and 8 pool ET and CT. If we split them, we see similar polarization and de-polarization effects within each treatment. Interestingly, we see that moderation among LED and Low is stronger in CT than in ET. This result is also inconsistent with LED subjects being merely inattentive. Our model can explain this result if cultural identity is more salient at baseline, so that its de-polarization in CT is stronger. Gennaioli and Tabellini (2025) find evidence for higher salience of cultural identity in a 2022 survey in the US.[23,24]

6 Concluding Remarks

A chief role of democratic institutions is to resolve group conflicts. Naturally, then, politics is often framed in terms of "us versus them": workers versus capitalists, religious versus secular, nationalists versus cosmopolitans, etc. This framing matters. As in the Biblical proverb "Hate stirs up conflict, but love covers all offenses", an antagonistic frame itself fuels conflict. But what does "us versus them" do? How? And to which people?

Identity sheds light on these questions. On the "What" question, it says that conflict also changes the beliefs of people over "how the world works or should work", it does not just trigger animosity or retribution among groups. On the "How" question, beliefs polarize along the salient ingroup stereotype, boosting latent group disagreement, with surprising spillovers across issues. On the "Who" question, it says that conflict is more polarizing for people who perceive stronger group disagreement and can realign some people across the

[23]If ET and CT increase the salience of the "unidentified" option , the identity choice Equation (4), and the belief Equations (7) and (8) imply that a LED type (e, c) reacts as:

$$\frac{\partial \tilde{e}}{\partial \sigma_{\emptyset}} = -p_{\emptyset} \cdot (p_{\hat{E}} \cdot e + p_{\hat{C}} \cdot c \cdot \rho) \cdot d \tag{19}$$

$$\frac{\partial \tilde{c}}{\partial \sigma_{\emptyset}} = -p_{\emptyset} \cdot (p_{\hat{E}} \cdot e \cdot \rho + p_{\hat{C}} \cdot c) \cdot d \tag{20}$$

Thus, there is *moderation* of LED compared to Control and such effect is stronger for CT than for ET provided that culture is more salient than economics at baseline, $p_{\hat{C}} > p_{\hat{E}}$.

[24]CT may exhibit a stronger de-polarization effect also because cultural enemies are more heterogeneous than economic ones. For instance, a culturally conservative person may feel no affinity with the "White Pride" enemy reported in the list, reducing her cultural identity and creating low expected disagreement and a moderating force. This may be particularly relevant for culturally conservatives who are non-White, for which we provide supportive evidence in Appendix F.

progressive-conservative divide based on the groups they belong to.

Identity can help explain the recent growth of populism and cultural polarization in the US and other countries. Consistent with our approach, cultural issues have become more salient over time, as measured by political advertising (Gennaioli and Tabellini, 2025) or cable news (Noy and Rao, 2025). Our mechanism then sheds light on three important aspects of cultural conflict and polarization stressed in previous work. A first aspect is the documented role of globalization (Autor et al., 2020), immigration (Abramowitz, 2018; Alesina et al., 2023), or of the 2008 Obama election (Sides et al., 2019) in fostering cultural conflict. As shown by Bonomi et al. (2021) and Gennaioli and Tabellini (2025), this can happen because these shocks increase the salience of conflict between educated/culturally progressive people and less educated/culturally conservative ones. Critically, our experiment shows that a salience switch from economics to culture exerts two effects that are consistent with the evidence. First, it changes the nature of conflicting groups, from economic to cultural, without changing population-wide disagreement much, consistent with Demset et al. (2025). Indeed, in our experiment polarization over economic and cultural issues reaches similar levels when either economic or cultural conflict are salient, due to spillovers. The second key consequence of growing cultural salience is realignment, which sharply changes the beliefs of mixed voters. Realignment explains why we see: i) redistributive conflict to have increased between religious vs. secular voters and to have decreased between lower vs upper class voters (Bonomi et al., 2021) and ii) sharp electoral realignments, with lower class conservative voters turning towards the Republican party (Gennaioli and Tabellini, 2025).

Another strand of work connects growing polarization to traditional or social media, which have allegedly distorted beliefs through targeted messages or echo chambers (Levy and Razin, 2019). Identity can explain why these channels can fuel persistent distortions. Targeting of propaganda to core supporters (Allcott et al., 2025) makes it optimal for politicians to diverge (Glaeser et al., 2005). Identity then explains why politicians want to diverge by making certain conflicts or stereotypes salient (Gennaioli and Tabellini, 2025) rather than by providing policy information.[25] Indeed, politicians often fuel extremism (Zhang et al., 2025; Ottinger and Posch, 2025) and their Twitter messages

[25]Antagonizing content is also more easily disseminated by social media (Van Bavel et al., 2024; Brady and Van Bavel, 2025), making it easier to raise conflict salience.

become more polarized before elections (Boeri et al., 2025). Identity in fact offers a theory of propaganda in which politicians compete to make their favored social conflict more salient than others. Once a social cleavage is "top of mind" and polarizes conflict, its salience becomes self-sustaining and persistently affects beliefs.[26]

A third body of work stresses affective polarization: growing distaste and even hatred of party outgroups (Gentzkow, 2016). Identity can explain why members of opposing salient groups can exaggerate their actual disagreement (Bordalo et al., 2016). As shown theoretically by Bonomi et al. (2021) when a conflict is salient, group stereotypes increase disagreement in policy views, but such effect in turn amplifies *perceived* disagreement in the same views above reality. This multiplier effect then naturally boosts affective polarization. We do not test this channel here, but Bordalo et al. (2020) show – consistent with it – that after the end of the cold war disagreement between Republicans and Democrats on domestic issues increased compared to foreign policy, and stereotypes of party outgroups moved in the same direction.

Given the social costs of polarization, which policies may help reduce it? Standard news-based account stress the role of unbiased information provision. The implications of our theory are different: when beliefs stereotype along a group, they neglect issue specific information. Thus, promoting accurate information or balanced exposure to news may have little effect or even backfire because it does not tackle the key cause of information neglect: salient conflict (Colonnelli et al., 2024). Interventions priming non-conflictual memberships, such as increasing the salience of an external threat, a common destiny, or similarity among people, could perhaps reduce conflict and polarization more effectively (Alan et al., 2021; Bursztyn and Yang, 2022; Andries et al., 2024).

A final set of implications concern social and group behavior. The assimilation of beliefs to salient groups can lead to choices that exhibit peer effects and emulation of others, and identity implies that conflict with outgroups can enhance such phenomena. It can also support group coordination and cooperation, softening the well known free riding problem in collective action (Olson, 1971; Ostrom, 1990). Specifically, by moving beliefs toward group stereotypes, identity can cause people to over-weight group welfare relative

[26]In this approach, political entrants may be able to dispel mainstream parties by proposing new frames (e.g. the people vs the elite) and by framing themselves as typical ingroup members (Dal Bó et al., 2017).

to social welfare, and to neglect private costs (Passarelli and Tabellini, 2017). This mechanism can however be only at play for members of salient groups. Differences in group salience, then, could explain why some groups are more effective at achieving their goals than others, and why external threats may harness cooperation.

References

Abramowitz, Alan I, *The great alignment: Race, party transformation, and the rise of Donald Trump*, Yale University Press, 2018.

Achen, Christopher H and Larry M Bartels, "Democracy for realists: Why elections do not produce responsive government," 2017.

Alan, Sule, Ceren Baysan, Mert Gumren, and Elif Kubilay, "Building social cohesion in ethnically mixed schools: An intervention on perspective taking," *The Quarterly Journal of Economics*, 2021, *136* (4), 2147–2194.

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

Algan, Yann, Eva Davoine, Thomas Renault, and Stefanie Stancheva, "Emotions and Policy Views," Technical Report, Working Paper 2025.

Allcott, Hunt, Matthew Gentzkow, Ro'ee Levy, Adriana Crespo-Tenorio, Natasha Dumas, and et al., "The Effects of Political Advertising on Facebook and Instagram before the 2020 US Election," *Nature and Human behavior*, 2025, *forthcoming.*

Andries, Marianne, Leonardo Bursztyn, Thomas Chaney, Milena Djourelova, and Alex Imas, "In their shoes: Empathy through information," Technical Report, National Bureau of Economic Research 2024.

Autor, David, David Dorn, Gordon Hanson, and Kaveh Majlesi, "Importing political polarization? The electoral consequences of rising trade exposure," *American Economic Review*, 2020, *110* (10), 3139–3183.

Baker, Scott R, Nicholas Bloom, and Steven J Davis, "Measuring economic policy uncertainty," *The quarterly journal of economics*, 2016, *131* (4), 1593–1636.

Bavel, Jay J Van, Claire E Robertson, Kareena Del Rosario, Jesper Rasmussen, and Steve Rathje, "Social media and morality," *Annual review of psychology*, 2024, *75* (1), 311–340.

Bénabou, Roland and Jean Tirole, "Mindful economics: The production, consumption, and value of beliefs," *Journal of Economic Perspectives*, 2016, *30* (3), 141–164.

Bó, Ernesto Dal, Frederico Finan, Olle Folke, Torsten Persson, and Johanna Rickne, "Who becomes a politician?," *The Quarterly Journal of Economics*, 2017, *132* (4), 1877–1914.

Boeri, Tito, Nina Nikiforova, and Guido Tabellini, "Do Elections Moderate or Polarize Political Rhetoric?," Technical Report, Bocconi University, Working Paper 2025.

Bonomi, Giampaolo, Nicola Gennaioli, and Guido Tabellini, "Identity, beliefs, and political conflict," *The Quarterly Journal of Economics*, 2021, *136* (4), 2371–2411.

Bordalo, Pedro, Katherine Coffman, Nicola Gennaioli, and Andrei Shleifer, "Stereotypes," *The Quarterly Journal of Economics*, 2016, *131* (4), 1753–1794.

_ , Marco Tabellini, and David Y Yang, "Issue salience and political stereotypes," Technical Report, National Bureau of Economic Research 2020.

_ , Nicola Gennaioli, Giacomo Lanzani, and Andrei Shleifer, "A Cognitive Theory of Reasoning and Choice," Technical Report, Working Paper 2025.

Brady, William J and Jay J Van Bavel, "Social identity shapes antecedents and functional outcomes of moral emotion expression.," *Journal of Experimental Psychology: General*, 2025.

Bursztyn, Leonardo and David Y Yang, "Misperceptions about others," *Annual review of economics*, 2022, *14* (1), 425–452.

Colonnelli, Emanuele, Niels Joachim Gormsen, and Tim McQuade, "Selfish corporations," *Review of Economic Studies*, 2024, *91* (3), 1498–1536.

Demset, Klaus, Ignacio Ortuno-Ortin, and Romain Wacziarg, "Latent polarization," Technical Report, National Bureau of Economic Research 2025.

Druckman, James N, Erik Peterson, and Rune Slothuus, "How elite partisan polarization affects public opinion formation," *American political science review*, 2013, *107* (1), 57–79.

Enke, Benjamin, "What you see is all there is," *The Quarterly Journal of Economics*, 2020, *135* (3), 1363–1398.

_ **and Florian Zimmermann**, "Correlation neglect in belief formation," *The review of economic studies*, 2019, *86* (1), 313–332.

Frank, Thomas, *What's the Matter with Kansas? How Conservatives Won the Heart of America*, Macmillan, 2004.

Gennaioli, Nicola and Guido Tabellini, "Presidential Address: Identity Politics," *Econometrica*, 2025, *93* (6), 1937–1967.

Gentzkow, Matthew, "Polarization in 2016," *Toulouse Network for Information Technology Whitepaper*, 2016, *1.*

Gethin, Amory, Clara Martínez-Toledano, and Thomas Piketty, "Brahmin left versus merchant right: Changing political cleavages in 21 western democracies, 1948–2020," *The Quarterly Journal of Economics*, 2022, *137* (1), 1–48.

Glaeser, Edward L, Giacomo Ponzetto, and Shapiro. Jesse, "The political economy of hatred," *The Quarterly Journal of Economics*, 2005, *120* (1), 45–86.

Graham, Matthew H and Milan W Svolik, "Democracy in America? Partisanship, polarization, and the robustness of support for democracy in the United States," *American Political Science Review*, 2020, *114* (2), 392–409.

Grossman, Gene M and Elhanan Helpman, "Identity politics and trade policy," *The Review of Economic Studies*, 2021, *88* (3), 1101–1126.

Iyengar, Shanto and Sean J Westwood, "Fear and loathing across party lines: New evidence on group polarization," *American journal of political science*, 2015, *59* (3), 690–707.

Jost, John T, Jack Glaser, Arie W Kruglanski, and Frank J Sulloway, "Political Conservatism as Motivated Social Cognition," *Psychological Bulletin*, 2003, *129* (3), 339–375.

Kahan, Dan M, "Climate-science communication and the measurement problem," *Political Psychology*, 2015, *36*, 1–43.

Klar, Samara, "Partisanship in a social setting," *American journal of political science*, 2014, *58* (3), 687–704.

Klein, Ezra, *Why we're polarized*, Simon and Schuster, 2020.

Kunda, Ziva, "The case for motivated reasoning.," *Psychological bulletin*, 1990, *108* (3), 480.

Levitsky, Steven and Daniel Ziblatt, *How democracies die*, Crown, 2018.

Levy, Gilat and Ronny Razin, "Echo chambers and their effects on economic and political outcomes," *Annual Review of Economics*, 2019, *11* (1), 303–328.

Levy, Ro'ee, "Social media, news consumption, and polarization: Evidence from a field experiment," *American economic review*, 2021, *111* (3), 831–870.

Noy, Shakked and Aakaash Rao, "The business of the culture war," *Available at SSRN 5694622*, 2025.

Olson, Mancur, *The Logic of Collective Action: Public Goods and the Theory of Groups, with a new preface and appendix*, Vol. 124, harvard university press, 1971.

Ortoleva, Pietro and Erik Snowberg, "Overconfidence in political behavior," *American Economic Review*, 2015, *105* (2), 504–535.

Ostrom, Elinor, *Governing the commons: The evolution of institutions for collective action*, Cambridge university press, 1990.

Ottinger, Sebastian and Max Posch, "The Political Economy of Propaganda: Evidence from US Newspapers," 2025.

Passarelli, Francesco and Guido Tabellini, "Emotions and Political Unrest," *Journal of Political Economy*, 2017, *125* (3), 903–946.

Shayo, Moses, "A model of social identity with an application to political economy: Nation, class, and redistribution," *American Political science review*, 2009, *103* (2), 147–174.

_ , "Social Identity and Economic Policy," *Annual Review of Economics*, 2020, *12* (1), 355–89.

Sides, John, Michael Tesler, and Lynn Vavreck, *Identity crisis: The 2016 presidential campaign and the battle for the meaning of America*, Princeton University Press, 2019.

Tajfel, Henri and John C. Turner, "An Integrative Theory of Intergroup Conflict," in William G. Austin and Stephen Worchel, eds., *The Social Psychology of Intergroup Relations*, Monterey, CA: Brooks/Cole, 1979, pp. 33–47.

_ **, John C Turner, William G Austin, and Stephen Worchel**, "An integrative theory of intergroup conflict," *Organizational identity: A reader*, 1979, *56* (65), 9780203505984–16.

Thaler, Michael, "The fake news effect: Experimentally identifying motivated reasoning using trust in news," *American Economic Journal: Microeconomics*, 2024, *16* (2), 1–38.

Tversky, Amos and Daniel Kahneman, "Judgment under Uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking under uncertainty.," *Science*, 1974, *185* (4157), 1124–1131.

Zaller, John, *The nature and origins of mass opinion*, Cambridge university press, 1992.

Zhang, Lehan, Gabriele Gratton, Pauline Grosjean, and Hasin Yousaf, "Adding Fuel to the (Gun) Fire: How Politicians Polarize the Public Debate," *Center for Law & Economics Working Paper Series*, 2025.

A Proofs

Proof of Prediction 1. Let $x = c, e$ denote person's (e, c) primitive beliefs in one domain and by $y = e, c, y \neq x$ as her primitive beliefs in the other. Let $\hat{X}$ and $\hat{Y}$ accordingly denote different dimensions of identity. Given Equations (5) and (6) we have:

$$\frac{\partial p_{\widehat{X}}}{\partial \sigma_{\widehat{X}}} = p_{\widehat{X}} \cdot \left(1 - p_{\widehat{X}}\right); \frac{\partial p_{\widehat{Y}}}{\partial \sigma_{\widehat{X}}} = -p_{\widehat{Y}} \cdot p_{\widehat{X}}.$$

The population variance of beliefs is the average of squared beliefs (average beliefs are in fact zero in each domain). For a type (x, y) the average squared belief $\widetilde{x}$ is equal to:

$$\mathbb{E}\left(\widetilde{x}^2\right) = p_{\emptyset} \cdot x^2 + p_{\widehat{X}} \cdot x^2 \cdot (1+d)^2 + p_{\widehat{Y}} \cdot (x + y \cdot \rho \cdot d)^2,$$

The population variance of beliefs is then the average of the above expression across types (x, y):

$$V\left(\widetilde{x}\right) = 1 + p_{\widehat{X}} \cdot (1+d)^2 + p_{\widehat{Y}} \cdot (1 + \rho^2 \cdot d^2 + 2\rho^2 d).$$

The effect of higher salience of conflict $\widehat{X}$ or $\widehat{Y}$ are respectively equal to:

$$\begin{aligned} \frac{\partial V\left(\widetilde{x}\right)}{\partial \sigma_{\widehat{X}}} &\propto (1 - p_{\widehat{X}}) - \rho^2 \cdot p_{\widehat{Y}} \\ \frac{\partial V\left(\widetilde{x}\right)}{\partial \sigma_{\widehat{Y}}} &\propto -p_{\widehat{X}} + \rho^2 \cdot (1 - p_{\widehat{Y}}) \end{aligned}$$

which are both positive provided:

$$\rho^2 > \frac{p_{\widehat{C}}}{1 - p_{\widehat{E}}}, \frac{p_{\widehat{E}}}{1 - p_{\widehat{C}}},$$

Which is the condition of Prediction 1, ensuring that beliefs in each domain (factual and normative) have higher variance when one conflict is salient.

Proof of Prediction 2

The generalized versions of Equations (15) and (16) for average beliefs in any issue economic or cultural issue $\widetilde{x}, \widetilde{y} \in \{\widetilde{c}, \widetilde{e}\}$ with $\widetilde{x} \neq \widetilde{y}$ under a treatment

making conflict $\widehat{x}$ salient are equal to:

$$\frac{\partial \widetilde{x}}{\partial \sigma_{\widehat{X}}} = p_{\widehat{X}} \cdot [x \cdot \left(1 - p_{\widehat{X}}\right) \cdot d - y \cdot p_{\widehat{Y}} \cdot \rho \cdot d], \tag{21}$$

$$\frac{\partial \widetilde{y}}{\partial \sigma_{\widehat{X}}} = p_{\widehat{X}} \cdot [x \cdot (1 - p_{\widehat{X}}) \cdot \rho \cdot d - y \cdot p_{\widehat{Y}} \cdot d]. \tag{22}$$

The average progressive type is has undistorted beliefs x_X and $\rho \cdot x_X$, the average conservative type has undistorted beliefs x_{-X} and $\rho \cdot x_{-X}$, where $x_X = -x_{-X} > 0$. The treatment effect on the average progressive is then positive (they become more progressive) provided:

$$(1 - p_{\widehat{X}}) - p_{\widehat{Y}} \cdot \rho^2 > 0, \tag{23}$$

$$(1 - p_{\widehat{X}}) - p_{\widehat{Y}} > 0. \tag{24}$$

which is fulfilled provided some people are not identified. The effect on the average conservative is the opposite due to the negative multiplicative term obtained by replacing x_X with x_{-X} in the above expressions. The analysis holds both for ET and CT, so Prediction 2 then follows.

Proof of Prediction 3. Let i be an economic or cultural issue. Disagreement in undistorted beliefs between the average member of a progressive and conservative group is equal to:

$$d_i\left(\widehat{E}\right) \equiv b_{i,E} - b_{i,-E} = \begin{cases} (\pi_i + \kappa_i) \cdot e_E & if \quad i \in \widetilde{E} \\ (\pi_i + \kappa_i) \cdot \rho \cdot e_E & if \quad i \in \widetilde{C} \end{cases}$$

$$d_i\left(\widehat{C}\right) \equiv b_{i,C} - b_{i,-C} = \begin{cases} (\pi_i + \kappa_i) \cdot \rho \cdot c_C & if \quad i \in \widetilde{E} \\ (\pi_i + \kappa_i) \cdot c_C & if \quad i \in \widetilde{C} \end{cases},$$

where $\widetilde{X}$ is the economic ($X = E$) or cultural ($X = C$) domain of the issue. Given the symmetric Guassian distribution is equal to:

$$d_i\left(\widehat{E}\right) \equiv b_{i,E} - b_{i,-E} = \begin{cases} (\pi_i + \kappa_i) \cdot d/2 & if \quad i \in \widetilde{E} \\ (\pi_i + \kappa_i) \cdot \rho \cdot d/2 & if \quad i \in \widetilde{C} \end{cases}$$

$$d_i\left(\widehat{C}\right) \equiv b_{i,C} - b_{i,-C} = \begin{cases} (\pi_i + \kappa_i) \cdot \rho \cdot d/2 & if \quad i \in \widetilde{E} \\ (\pi_i + \kappa_i) \cdot d/2 & if \quad i \in \widetilde{C} \end{cases}.$$

Thus, $d_i\left(\widehat{C}\right) = \rho \cdot d_i\left(\widehat{E}\right)$ for $i \in \widetilde{E}$ and $d_i\left(\widehat{E}\right) = \rho \cdot d_i\left(\widehat{C}\right)$ for $i \in \widetilde{C}$.

The observed beliefs on issue i of the average member of groups $X = E, C$ and $-X$ in condition K = ET, CT, Control are:

$$\begin{aligned}\widetilde{b}^{\mathrm{K}}_{i,X} &= b_{i,X} + p^{\mathrm{K}}_{\widehat{\mathrm{X}}} \cdot x_X \cdot d_i\left(\widehat{\mathrm{X}}\right) + p^{\mathrm{K}}_{\widehat{\mathrm{Y}}} \cdot \rho \cdot x_X \cdot d_i\left(\widehat{\mathrm{Y}}\right),\\ \widetilde{b}^{\mathrm{K}}_{i,-X} &= -b_{i,X} - p^{\mathrm{K}}_{\widehat{\mathrm{X}}} \cdot x_X \cdot d_i\left(\widehat{\mathrm{X}}\right) - p^{\mathrm{K}}_{\widehat{\mathrm{Y}}} \cdot \rho \cdot x_X \cdot d_i\left(\widehat{\mathrm{Y}}\right),\end{aligned}$$

where $\widehat{\mathrm{X}}$, X=E,C is the partition between X and $-X$ and $\widehat{\mathrm{Y}}$ is partition Y$\neq$X.

So that observed group disagreement on issue i in condition K is equal to:

$$\begin{aligned}\widetilde{b}^{\mathrm{K}}_{i,X} - \widetilde{b}^{\mathrm{K}}_{i,-X} &= d_i\left(\widehat{\mathrm{X}}\right) + 2p^{\mathrm{K}}_{\widehat{\mathrm{X}}} \cdot x_X \cdot d_i\left(\widehat{\mathrm{X}}\right) + 2p^{\mathrm{K}}_{\widehat{\mathrm{Y}}} \cdot \rho \cdot x_X \cdot d_i\left(\widehat{\mathrm{Y}}\right)\\ &= d_i\left(\widehat{\mathrm{X}}\right) + p^{\mathrm{K}}_{\widehat{\mathrm{X}}} \cdot d \cdot d_i\left(\widehat{\mathrm{X}}\right) + p^{\mathrm{K}}_{\widehat{\mathrm{Y}}} \cdot \rho \cdot d \cdot d_i\left(\widehat{\mathrm{Y}}\right)\end{aligned}$$

where the second line uses the fact that $2x_X = d$. This is in turn equal to:

$$\widetilde{b}^{\mathrm{K}}_{i,X} - \widetilde{b}^{\mathrm{K}}_{i,-X} = \begin{cases} \left[1 + p^{\mathrm{K}}_{\widehat{\mathrm{X}}} \cdot d + p^{\mathrm{K}}_{\widehat{\mathrm{Y}}} \cdot d \cdot \rho^2\right] \cdot d_i\left(\widehat{\mathrm{X}}\right) & if \quad i \in \widetilde{X} \\ \left[1 + p^{\mathrm{K}}_{\widehat{\mathrm{X}}} \cdot d + p^{\mathrm{K}}_{\widehat{\mathrm{Y}}} \cdot d\right] \cdot d_i\left(\widehat{\mathrm{X}}\right) & if \quad i \in \widetilde{Y} \end{cases}$$

which, in turn, implies that:

$$d_i\left(\widehat{\mathrm{X}}\right) = \delta^{\mathrm{K}}_{\mathrm{D,X}} \cdot \left(\widetilde{b}^{\mathrm{K}}_{i,X} - \widetilde{b}^{\mathrm{K}}_{i,-X}\right), \qquad \mathrm{D} \;=\; E,\, C. \tag{25}$$

Observed disagreement on issue i in condition K between two groups is informative about the groups' disagreement on the same issue, with loading $\delta^{\mathrm{K}}_{\mathrm{D,X}}$ that depends on the condition itself and is higher for the domain D along which the group is defined. The belief of groups $X, -X$ in a target condition K=ET, CT, Control, expressed in terms of a baseline condition B=ET,CT, Control, B$\neq$K, are equal to:

$$\begin{aligned}\widetilde{b}^{\mathrm{K}}_{i,X} &= \widetilde{b}^{\mathrm{B}}_{i,X} + \left(p^{\mathrm{K}}_{\widehat{\mathrm{X}}} - p^{\mathrm{B}}_{\widehat{\mathrm{X}}}\right) \cdot (d/2) \cdot d_i\left(\widehat{\mathrm{X}}\right) + \left(p^{\mathrm{K}}_{\widehat{\mathrm{Y}}} - p^{\mathrm{B}}_{\widehat{\mathrm{Y}}}\right) \cdot \rho \cdot (d/2) \cdot d_i\left(\widehat{\mathrm{Y}}\right),\\ \widetilde{b}^{\mathrm{K}}_{i,-X} &= \widetilde{b}^{\mathrm{B}}_{i,X} - \left(p^{\mathrm{K}}_{\widehat{\mathrm{X}}} - p^{\mathrm{B}}_{\widehat{\mathrm{X}}}\right) \cdot (d/2) \cdot d_i\left(\widehat{\mathrm{X}}\right) - \left(p^{\mathrm{K}}_{\widehat{\mathrm{Y}}} - p^{\mathrm{B}}_{\widehat{\mathrm{Y}}}\right) \cdot \rho \cdot (d/2) \cdot d_i\left(\widehat{\mathrm{Y}}\right),\end{aligned}$$

Which for group X and $-X$ can in turn be written as:

$$\begin{aligned}\widetilde{b}^{\mathrm{K}}_{i,X} &= \widetilde{b}^{\mathrm{B}}_{i,X} + \mu^{\mathrm{K,B}}_{D,\mathrm{X}} \cdot \left(\widetilde{b}^{\mathrm{B}}_{i,X} - \widetilde{b}^{\mathrm{B}}_{i,-X}\right),\\ \widetilde{b}^{\mathrm{K}}_{i,-X} &= \widetilde{b}^{\mathrm{B}}_{i,-X} + \mu^{\mathrm{K,B}}_{D,\mathrm{X}} \cdot \left(\widetilde{b}^{\mathrm{B}}_{i,-X} - \widetilde{b}^{\mathrm{B}}_{i,X}\right),\end{aligned}$$

where the equation is obtained by writing $d_i(\widehat{\mathrm{Y}})$ in terms of $d_i\left(\widehat{\mathrm{X}}\right)$ so that:

$$\mu_{D,\mathrm{X}}^{\mathrm{K,B}} = (d/2)\begin{cases} \left[\left(p_{\widehat{\mathrm{X}}}^{\mathrm{K}} - p_{\widehat{\mathrm{X}}}^{\mathrm{B}}\right) + \left(p_{\widehat{\mathrm{Y}}}^{\mathrm{K}} - p_{\widehat{\mathrm{Y}}}^{\mathrm{B}}\right)\cdot\rho^2\right]\cdot\delta_{X,\mathrm{X}}^{\mathrm{B}} & if \quad i\in\widetilde{X} \\ \left[\left(p_{\widehat{\mathrm{X}}}^{\mathrm{K}} - p_{\widehat{\mathrm{X}}}^{\mathrm{B}}\right) + \left(p_{\widehat{\mathrm{Y}}}^{\mathrm{K}} - p_{\widehat{\mathrm{Y}}}^{\mathrm{B}}\right)\right]\cdot\delta_{Y,\mathrm{X}}^{\mathrm{B}} & if \quad i\in\widetilde{Y} \end{cases}.$$

Thus, a groups' beliefs in condition K increase in the same group's beliefs in the baseline B and in disagreemement with outgorups in the same baseline.

The coefficient $\mu_{D}^{\mathrm{K,B}}$ varies across domains and across the target and baseline conditions. In our specification, where K = XT, so groups X and $-X$ are primed, the coefficient $\mu_{D}^{\mathrm{K,B}}$ is larger in the domain defining the primed groups than on the other domain. The difference in the two coefficients is however small if ρ is large. The coefficient also varies across different treatments K depending on the magnitude of their effect on different identities. We estimate the equation under the approximation of a constant coefficient $\mu^{\mathrm{K,B}}$ that does not vary across domains. This will not affect the sign of the coefficient provided variation in disagreement across issues is sufficiently large compared to variation in coefficient magnitude, which is realistic.

Proof of Prediction 4. Beliefs are:

$$\tilde{e} = e\cdot(1+p_{\hat{E}}\cdot d) + p_{\hat{C}}\cdot c\cdot\rho\cdot d, \tag{26}$$

$$\tilde{c} = c\cdot(1+p_{\hat{C}}\cdot d) + p_{\hat{E}}\cdot e\cdot\rho\cdot d. \tag{27}$$

So, disagreement between the average member of a progressive and conservative economic group is:

$$\tilde{e}_E - \tilde{e}_{-E} = 2e_E\cdot(1+p_{\hat{E}}\cdot d) + 2\cdot p_{\hat{C}}\cdot e_E\cdot\rho^2\cdot d, \tag{28}$$

$$\tilde{c}_E - \tilde{c}_{-E} = 2e_E\cdot\rho\cdot(1+p_{\hat{C}}\cdot d) + 2\cdot p_{\hat{E}}\cdot e_E\cdot\rho\cdot d. \tag{29}$$

In turn, given that then, disagreement in general progressivity is equal to $\tilde{b}_E$ - $\tilde{b}_{-E} \propto 1 + p_{\hat{E}} + p_{\hat{C}}\cdot\rho$, and among cultural types it is respectively equal to $\tilde{b}_C$ - $\tilde{b}_{-C} \propto 1 + p_{\hat{C}} + p_{\hat{E}}\cdot\rho$. It is then clear that a change from CT to ET that increases economic identity and reduces cultural identity causes a change in disagreement among E and $-E$ larger than among C and $-C$ when $\rho < 1$.

Consider now pure and mixed types. Taking into account the beliefs equations and the fact that, by the properties of the standard bivariate Gaussian,

$e_{EC} = c_{EC}$ and $e_{E-C} = -c_{E-C} = c_{-EC} = -e_{-EC} < c_{EC}$ and so on, where b_{XY} is the undistorted belief of type (X, Y).[27]

Disagreement in general progressivity among pure types is then equal to:

$$(\tilde{b}_{E,C} - \tilde{b}_{-E,-C}) = (b_{E,C} - b_{-E,-C}) + 2e_{E,C} \cdot (p_{\hat{E}} + p_{\hat{C}}) \cdot (1 + \rho) \cdot d. \quad (30)$$

A switch from CT to ET has an ambiguous effect on $(p_{\hat{E}} + p_{\hat{C}})$, so it moves disagreement among pure types in an ambiguous way. Disagreement in general progressivity among mixed types is instead equal to:

$$(\tilde{b}_{E,-C} - \tilde{b}_{-E,C}) = (b_{E,-C} - b_{-E,C}) + 2e_{E,-C} \cdot (p_{\hat{E}} - p_{\hat{C}}) \cdot (1 + \rho) \cdot d. \quad (31)$$

A switch from CT to ET unambiguously increases $(p_{\hat{E}} - p_{\hat{C}})$, so it increases $(\tilde{b}_{E,-C} - \tilde{b}_{-E,C})$, proving Prediction 4.

[27] Specifically:

$$\begin{aligned} e_{EC} &= \frac{1+\rho}{\Phi_+\sqrt{2\pi}} = c_{EC} \\ e_{E-C} &= \frac{1-\rho}{\Phi_-\sqrt{2\pi}} = -c_{E-C} < c_{EC} \end{aligned}$$

where $\Phi_+ = \Pr(e > 0, c > 0)$ and $\Phi_- = \Pr(e > 0, c < 0)$ is the cumulative distribution of the standard bivariate.

B Survey experiment

B.1 Study Sample

	All waves	Waves 1	Wave 2	Wave 3	Census
Female	50%	51%	49%	50%	51%
Male	50%	49%	51%	50%	49%
18-24yo	11%	12%	12%	9%	12%
25-34yo	19%	18%	19%	18%	18%
35-44yo	16%	16%	16%	16%	16%
45-54yo	17%	17%	17%	18%	17%
55-64yo	17%	17%	16%	18%	17%
≥65yo	20%	20%	19%	21%	20%
White	80%	79%	81%	79%	72%
Black	10%	11%	8%	12%	12%
Other	10%	10%	10%	9%	15%
No high school	1%	1%	1%	2%	12%
High School	23%	19%	22%	26%	27%
Some College	26%	27%	25%	25%	20%
College	36%	38%	35%	34%	28%
Grad School	15%	17%	14%	14%	12%
<10k	6%	5%	6%	6%	6%
[10k, 15k)	4%	3%	5%	5%	6%
[15k, 25k)	8%	7%	9%	9%	9%
[25k, 35k)	10%	10%	10%	11%	9%
[35k, 50k)	14%	13%	14%	14%	12%
[50k, 75k)	21%	21%	21%	20%	17%
[75k, 100k)	14%	15%	13%	14%	13%
[100k, 150k)	14%	16%	14%	12%	15%
[150k, 200k)	5%	6%	5%	5%	7%
≥200k	4%	4%	5%	3%	6%
Northeast	17%	17%	17%	17%	17%
Midwest	20%	20%	22%	19%	21%
West	24%	25%	21%	26%	25%
South	39%	38%	40%	38%	38%

Table B.1: Average socio-economic and demographic information of respondents of the entire study sample by wave and pooled, relative to Census data.

B.2 Survey design: Experimental structure

	Survey Parts in Temporal Order				
	Part 1: Identity Membership (i)	**Part 2:** Add Identity Questions	**Part 3:** Outcomes and expected disagreement	**Part 4:** Identity Membership (ii)	**Part 5:** Add info
Control	Placebo	Placebo Ingroup & Outgroup	Political Beliefs	Cultural, Economic	Partisan Identity
ET	Economic	Economic Ingroup & Outgroup	Political Beliefs	Cultural	Partisan Identity
CT	Cultural	Cultural Ingroup & Outgroup	Political Beliefs	Economic	Partisan Identity

Table B.2: Survey design overview

B.3 Screenshots Parts 1-5

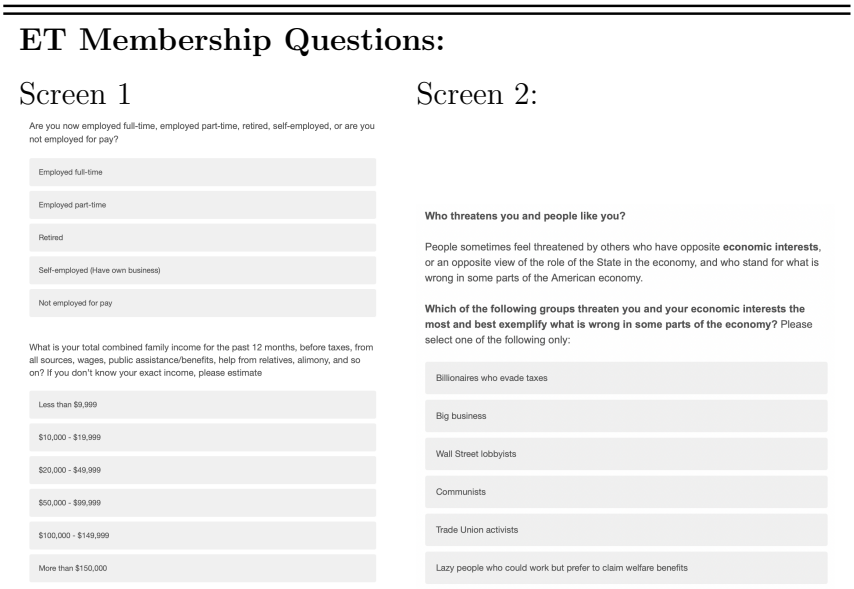


Figure B.1: ET participants face the ET Membership Questions in Part 1 of the survey. CT and Control participants face the ET Membership Questions in Part 4 of the survey.

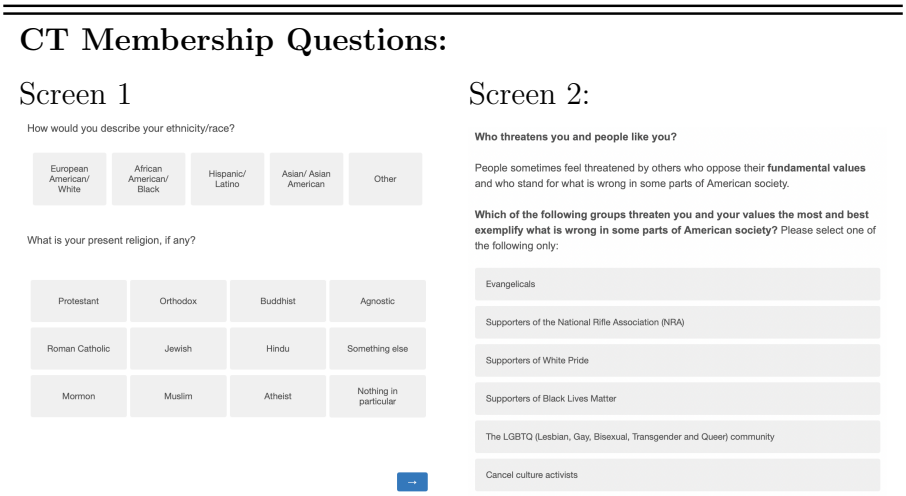
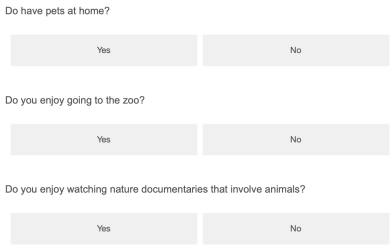


Figure B.2: CT participants face the CT Membership Questions in Part 1 of the survey. ET and Control participants face the CT Membership Questions in Part 4 of the survey.

Control Membership Questions:

Screen 1



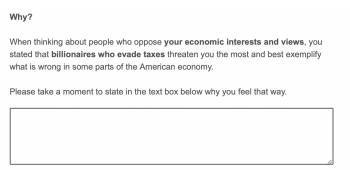
Screen 2:



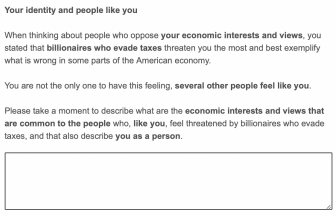
Figure B.3: Control participants face the Control Membership Questions in Part 1 of the survey.

ET Add Identity Questions

Screen 1:



Screen 2:



Screen 3:

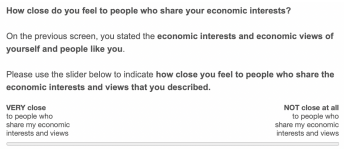
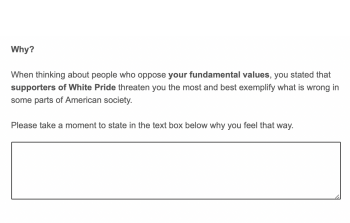


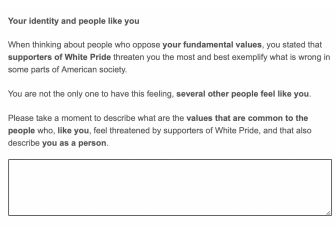
Figure B.4: ET participants face additional questions on their economic identity in Part 2 of the survey.

CT Add Identity Questions

Screen 1:



Screen 2:



Screen 3:

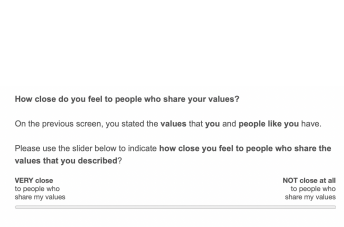
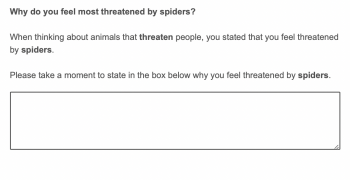


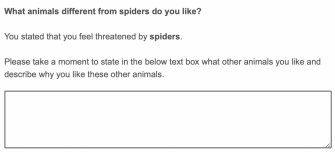
Figure B.5: CT participants face additional questions on their cultural identity in Part 2 of the survey.

Part 2, Control Add Identity Questions

Screen 1:



Screen 2:



Screen 3:

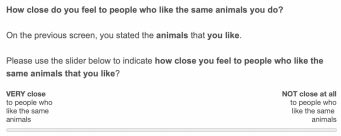


Figure B.6: Control participants face additional placebo questions in Part 2 of the survey.

Example Screen:
Outcome Question & Expected Disagreement in ET

Question: Some people feel that the government in Washington should see to it that every person has a job and a good standard of living. Others think the government should just let each person get ahead on their own. Where would you place yourself on this scale?

1. Task: Do you expect agreement or disagreement between **people like you** and **lazy people who could work but prefer to claim welfare benefits**?

Agreement between people like you and lazy people

Disagreement between people like you and lazy people

2. Task: What do you personally think? How would you personally respond to the question?

Take a score between 0 and 10 with the following meaning:
0: a lot of help should be provided,
10: no help at all should be provided.
What score between 0 and 10 comes closest to how you feel?

Government should HELP 0	1	2	3	4	5	6	7	8	9	Government should STAY AWAY 10

Figure B.7: Example screen featuring an outcome question and the corresponding expected disagreement question for one question of Part 3 for ET participants. For CT and Control participants, the screen would look the same but the expected disagreement question would refer to their chosen enemy from Part 1.

B.4 Wording of all outcome questions

	Questions: Economic and Cultural Beliefs
e^1:	Which has more to do with why a person is poor: lack of effort or bad luck?
e^2:	How has the share of total U.S. income that goes to the richest 1% of the U.S. population evolved over the past 30 years?
e^3:	The federal estate tax is a tax imposed on the transfer of wealth from a deceased person to his or her heirs. Some people feel that the estate tax should be increased to equalize opportunities among the next generation. Others think that estate taxes should be reduced to avoid double taxing. Where would you place yourself on this scale?
e^4:	Some people feel that the government in Washington should see to it that every person has a job and a good standard of living. Others think the government should just let each person get ahead on their own. Where would you place yourself on this scale?
e^5:	Some people think the government should provide fewer services, even in areas such as health and education, in order to reduce spending. Other people feel that it is important for the government to provide many more services even if it means an increase in spending. Where would you place yourself on this scale?
c^1:	Do you think that Black Americans are currently discriminated against in the hiring decisions of firms?
c^2:	Consider all crimes committed in the past 12 months. Relative to their share in population, how many of these crimes were committed by immigrants?
c^3:	Do you favor or oppose affirmative action programs that are designed to help blacks, women and other minorities to find better jobs and education?
c^4:	Consider the immigration policy of the U.S. What is your opinion about whether your country should provide food and shelter for political refugees?
c^5:	Some people argue that abortion should always be legal in order to protect mothers. Others say that abortion should always be illegal in order to protect unborn life. What is your view on this issue?

Table B.3: Questions used to elicit economic and cultural beliefs.

B.5 Additional information

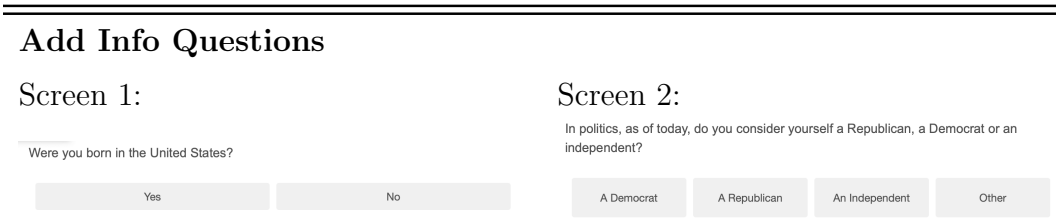


Figure B.8: All participants are asked whether they are US born and about their political affiliation in Part 5 of the survey..

C Text Analyses and Similarity

For each treated participant, we compute a similarity score that captures how similar their ingroup description is to their broad ingroup or narrow ingroup as implied by their enemy choice. Consider a participant in ET, for instance, who chooses as their enemy Trade Union activists. Their broad economic ingroup implied by this enemy choice is economic conservatives, and their narrow economic ingroup is those who chose Trade Union activists as their enemy. The similarity score of this participant captures how similar their ingroup description is either to the words that are commonly used by people in their broad ingroup to describe themselves (but not commonly used in their broad outgroup) or to the words that are commonly used by people in their narrow ingroup (but not commonly used in all other narrow groups). We call the former broad similarity score and the latter narrow similarity score. Participants' similarity score is then the maximum of their broad similarity score and their narrow similarity score. We describe how we compute these scores in the following.

C.1 Broad Similarity

Classification We created broad ingroup descriptions (BID) separately for E, -E, C and -C.

BID for E: Classify text into this category if participants write about their ingroup by referring to (economic) inequality in ways that justify redistribution of income/wealth between poorer and richer groups of people and provide services to help the economically less well-off, and by referring to inequality in ways that emphasize the struggles of the working or middle class, that everyone has to pay their fair share of taxes and privileges of elites, big corporations,

and lobbyists

BID for $-E$: Classify text into this category if participants write about their ingroup by referring to (economic) inequality in ways that justify why such inequality should not be changed (by the government) and that social welfare provision is wasteful, gets misused by recipients and leads to shirking/laziness and that undeserving people receive too much social security, and by referring to frustration with excessive government spending and bureaucratic inefficiency

BID for C: Classify text into this category if participants write about their ingroup by referring to universal/progressive cultural/moral aspects that relate to civil rights, equal rights and openness, tolerance, altruism towards minorities/women and disregard of conservatives.

BID for -C: Classify text into this category if participants write about their ingroup by referring to conservative cultural/moral aspects that relate to traditional/American values (for instance: free speech, ability to express beliefs without judgement/gun use), hierarchies, religion, and potentially disregard of minorities.

Each ingroup text in ET (CT) is then classified to be E, -E (C, -C) or neither. First, RA classifies for each treatment 400 random texts by hand. Second, RA hands 300 of these classifications over to LLM and LLM classifies all texts (including the 400 texts classified by RA). For ET ingroup texts, 26.9% were classified by the LLM to be E and 26.93% to be $-E$. For CT ingroup texts, 39.65% were classified to be C and 28.8% were classified to be $-C$. Based on the 400 texts classified by both RA and LLM, we find that the LLM accuracy is 92.25% for ET texts and 90.25% for CT texts.

Rubrics All ingroup texts classified to E were grouped into group E and all classified to $-E$ were grouped into group $-E$. The E-rubric was then created which consists of 20 words/bigrams that were common in the texts of the E group but not common in the text of the $-E$ group. Appendix C.3 describes the method used to generate rubrics and Appendix C.4 shows them. The same was done analogously for ingroup texts classified to be C and $-C$ in CT.

Broad Similarity Scores Each ingroup text in ET was compared to the E-rubric and to the $-E$-rubric according to the details stated in Appendix C.5. For subjects who we classified to be E according to their enemy choice

in the experiment, we then take the similarity score of the ingroup text with respect to the E-rubric to be their broad similarity score. For $-E$ subjects we take their similarity score to the $-E$-rubric to be their broad similarity score. The same is done in CT to arrive at subjects' broad similarity scores.

C.2 Narrow Similarity

Classification Each enemy text in ET (CT) is classified to belong to one of the economic (cultural) enemy choices. The RA did this first by hand (knowing only the relevant labels of the enemy options) for a random subset of texts and then the LLM completed the classification, like for the broad ingroup classification stated above.

Rubrics All ingroup texts classified to one of the narrow groups (e.g., billionaires) were grouped into one group. Narrow rubrics were then created like for broad rubrics by taking the 20 words/bigrams that were common in the respective narrow group but not common in the other narrow groups, see also Appendix C.3. The resulting Narrow rubrics for are shown in Appendix C.6

Narrow similarity scores Each ingroup text in ET was compared to each narrow rubric. For subjects who picked a particular enemy in the experiment, we then take the similarity score of the ingroup text with respect to that same rubric to be their narrow similarity score. The same is done in CT to arrive at subjects' narrow similarity scores.

C.3 TF-IDF

TF-IDF quantifies the importance of a term within a "document" by combining its frequency within a text with its rarity across multiple other documents. The formula is:

$$TF - IDF(t, d, D) = \frac{\text{No.of times a term t appears in doc d}}{\text{Tot.No.of terms in doc d}} \times \log\left(\frac{\text{Tot.No.of documents in corpus D}}{\text{Number of documents containing term t}}\right)$$

"Documents" are the ingroup text responses. We have 2 corpora, the collection of broud ingroup texts and the collection of narrow ingroup texts.

Separately for each corpus, we build a vocabulary of terms that occur at least 3 times across all text responses (filtering out too rare terms). Then we convert each survey text into a TF-IDF vector: for each term in each survey text, we get the TF-IDF frequency of that term (how frequent it is in the text multiplied by its inverse document frequency, or how rare it is in the entire collection of ingroup/outgroup texts). Then, for each group, we compute the centroid of the TF-IDF vectors: so, for each group (broad: progressive, conservative; narrow: big business etc.), we have a simple average of all the TF-IDF frequencies of the texts belonging to that group. This is our group prototype. Then, separately for broad and narrow groups, we take the difference of the centroids to get the relative importance of each term in the group compared to other groups. So for broad ingroups, we have progressive-centroid minus conservative-centroid, and the terms with highest differences are more representative of the progressive ingroup compared to the conservative ingroup. We take the top 20 words and bigrams. For the narrow ingroups, we do the difference of each narrow group with the centroid of all the other narrow ingroups averaged together. Once again we take the top 20 words and bigrams for the rubric.

C.4 Broad Rubrics

Rubric for broad ET ingroups E: rich, billionaires, fair, taxes, pay, poor, share, resources, fair share, pay fair, allocated, economic, fairly, economic resources, pay taxes, class, share taxes, wealthy, allocated fairly, resources allocated, rich poor, wealth, income, poor people, rich people.

Rubric for broad ET ingroups $-E$: work, hard, lazy, work hard, government, working, money, welfare, lazy people, people, bureaucrats, free, benefits, people work, giving, job, washington, support, do, earn, worked, giving money, communism, advantage, taking

C.5 LLM Prompt for Similarity Scores

Similarity was calculated by asking GPT-4o mini to produce a score between 0 and 1 for each profile (again made up of the concatenated ingroup and outgroup texts), comparing it to each broad and narrow ingroup rubric. The model was then fed the following prompt:

Each profile was evaluated 3 times against the rubrics, and the overall

```
    SYSTEM_PROMPT = """
You are a careful rater. Return ONLY strict JSON.

Task:
- Given a PROFILE and a set of class rubrics (content prototypes), rate the
↪  semantic similarity of the PROFILE to ALL classes.
- Similarity is topical/semantic overlap in themes, vocabulary, and content (not
↪  sentiment).

Output format (and nothing else):
{
  "ingroup":  {"1": <number in [0,1]>, "2": <number in [0,1]>},
  "outgroup": {"1": <number in [0,1]>, "2": <number in [0,1]>, ..., "10": <number
  ↪  in [0,1]>}
}

Scoring rules:
- 0.0 = no topical/semantic overlap; 1.0 = extremely strong overlap.
- Use continuous decimals (e.g., 0.27, 0.53, 0.88).
- Base your scores ONLY on the PROFILE text.
""".strip()
```

Figure C.9: Prompt to receive similarity scores.

similarity score of the profile to the rubric was calculated by taking the average of these 3 numbers. To be able to average and thus obtain a continuous measure of similarity, the temperature of the model was set to 0.2 (not fully deterministic).

C.6 Narrow Rubrics

Rubric for narrow CT ingroups for people threatened by "Evangelicals": evangelicals, religion, freedom, rights, tolerance, freedom religion, equal rights, gay, women, include, practice, choice, state, threatened evangelicals, body.

Rubric for narrow CT ingroups for people threatened by "NRA supporters": guns, gun, safety, violence, peace, safe, gun violence, nra, weapons, shootings, need, peace safety, assault, mass, right.

Rubric for narrow CT ingroups for people threatened by "White Pride supporters": equal, equal rights, white, rights, equality, pride, white pride, race, tolerance, color, races, peace, treated, peace freedom, skin.

Rubric for narrow CT ingroups for people threatened by "BLM supporters": black, matter, lives matter, blm, lives, black lives, police, respect, lawfulness, law, blacks, black people, loyalty, white, just black.

Rubric for narrow CT ingroups for people threatened by "LGBTQ community": god, woman, man, Igbta, children, man woman, family, marriage, female, bible, traditional, tradition, christian, values, community.

Rubric for narrow CT ingroups for people threatened by "Cancel-culture activists": cancel, culture, cancel culture, speech, history, respect, free, opinion, loyalty, things, freedom speech, free speech, opinions, say, mistakes.

Rubric for narrow CT ingroups for people threatened by "Religious extremists like Evangelicals": religious, religion, extremists, religious extremists, beliefs, abortion, personal, freedom, autonomy, believe, thinking, right, personal freedom, critical, bodily.

Rubric for narrow CT ingroups for people threatened by "Racists like White Supremacists": people, race, color, white, racists, skin, love, treated, racism, supremacists, diversity, white supremacists, person, equality, better.

Rubric for narrow CT ingroups for people threatened by "Illegal immigrants / their defenders": illegal, immigrants, hard, work, citizens, country, come, illegal immigrants, law, taxes, illegals, americans, pay, help, legally.

Rubric for narrow CT ingroups for people threatened by "Woke / cancel-culture extremists": woke, traditional, speech, family, god, believe, conservative, freedom, history, able, free, genders, free speech, woke people, right.

Rubric for narrow ET ingroups for people threatened by "Billionaires who evade taxes": rich, pay, taxes, billionaires, fair, yes, resources, share, fair share, poor, allocated, pay taxes, pay fair, economic resources, fairly.

Rubric for narrow ET ingroups for people threatened by "Big business": big, businesses, business, big businesses, big business, poor, resources, fairly, allocated, allocated fairly, profits, rich poor, profit, economic resources, prices.

Rubric for narrow ET ingroups for people threatened by "Wall Street lobbyists": lobbyists, lobbyist, wall street, wall, influence, street, rich, american, congress, laws, economic, lobbying, interests, street lobbyists, companies.

Rubric for narrow ET ingroups for people threatened by "Communists": communism, freedom, government, washington, communists, communist, dc, free market, market, believe, capitalism, free, want, bureaucrats, freedoms.

Rubric for narrow ET ingroups for people threatened by "Trade Union activists": unions, unsure, washington dc, owners, felt, dc, washington, speak, little, trade, money, people believe, causing, parts, drug.

Rubric for narrow ET ingroups for people threatened by "Lazy people who could work but prefer to claim welfare benefits": work, yes, lazy, people, money, bureaucrats, hard, welfare, giving, benefits, lazy people, work hard, giving money, government, people work.

Rubric for narrow ET ingroups for people threatened by "Billionaires who evade taxes": taxes, billionaires, pay, class, pay taxes, middle class, middle, aFord, tax, paying, fair, make, don know, wealthy, share.

Rubric for narrow ET ingroups for people threatened by "Big business and Wall Street lobbyists": big, businesses, business, big business, prices, lobbyists, interests, corporations, big businesses, wages, corporate, small, street, wall street, wall.

Rubric for narrow ET ingroups for people threatened by "Communists and Trade Union activists": unions, communism, communists, believe, communist, trade, freedom, government, economy, union, american, control, free, rights, america.

Rubric for narrow ET ingroups for people threatened by "Lazy people who could work but prefer to claim welfare benefits": work, hard, lazy, work hard, working, lazy people, welfare, family, hard working, sure, worked, people, help, people work, really.

D Narrow predicted disagreement

We computed predicted disagreement among all economically progressive narrow groups displayed in Figure 4. First, for all narrow economic groups in ET, separately for mixed and pure types, we predict their beliefs based on the same narrow group in CT as well as on disagreement between the narrow ingroup and all other groups in CT. Then, we compute predicted disagreement for each progressive narrow group by taking its predicted beliefs and subtracting from it the average predicted beliefs of all other narrow groups. This yields, for each belief in issue i, predicted disagreement of five progressive narrow groups (three from waves 1 and 2, two from wave 3).

E Expected disagreement

E.1 Synthetic Control

We apply the following procedure to compute the synthetic Control for treatment $X \in \{\text{ET}, \text{CT}\}$ on outcome $Y \in \{e, c\}$ and disagreement types $Z \in \{\text{HED}, \text{LED}\}$. We separately compute the quintiles of Z in treatment X and in the Control. Then we calculate the share of participants who are Z in the treatment for each quintile. For $Z = \text{HED}$ (LED) participants, shares in the top and bottom quintiles exceed (are below) 20% and in the center quintile are below (exceed) 20%. Then we compute weights for each Control participant

Outcome:	OLS regressions Progressivity by question							
Progressives:	X = E				X = C			
Treatment:	XT		Control		XT		Control	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
D	-1.314***	-1.231***	-0.688***	-0.776***	-1.760***	-1.694***	-0.654***	-0.755***
	(0.070)	(0.065)	(0.067)	(0.063)	(0.074)	(0.068)	(0.062)	(0.059)
X D	1.734***	1.640***	0.155*	0.295***	2.790***	2.520***	0.111	0.297***
	(0.093)	(0.086)	(0.084)	(0.078)	(0.106)	(0.096)	(0.082)	(0.077)
ID FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Issue FE	No	Yes	No	Yes	No	Yes	No	Yes
$\beta_1 + \beta_2 > 0$	Yes	Yes	No	No	Yes	Yes	No	No
$\beta_1 + \beta_2 < 0$	No	No	Yes	Yes	No	No	Yes	Yes
Adj. R^2	0.377	0.450	0.330	0.430	0.396	0.484	0.330	0.430
Clusters	4252	4252	4486	4486	4222	4222	4486	4486

Table E.4: Results of estimating regression Equation 32 for economic progressives vs conservatives in Columns (1)-(4), separately for ET (Columns (1)-(2)) and the Control (Columns (3)-(4)), and for cultural progressives vs conservatives in Columns (5)-(8), separately for CT (Columns (5)-(6)) and the Control (Columns (7)-(8)). We include participant fixed effects in all cases as well as question fixed effects in even rows. Robust standard errors in brackets. *** (**,*) indicate statistical significance at 0.01 (0.05, 0.1) level.

based on the share of disagreement types Z in the treatment in the corresponding quintile divided by the share Control participants in that quintile. In consequence, when we compare Control participants to Z = HED (LED) participants in the treatment, Control participants in the top and bottom quintiles are over-weighted (under-weighted) and in the center are under-weighted (over-weighted).

E.2 Individual-level variation in expected disagreement

An alternative approach to analyzing the effect of expected disagreement is to exploit individual-level variation in expected disagreement. Recall that participants state for each outcome question whether they expect disagreement or not between their economic ingroup and outgroup in ET, their cultural ingroup and outgroup in CT, and their placebo ingroup and outgroup in Control. We then estimate the following regression equation:

$$\tilde{b}_{ij} = \beta_1 \cdot D_{ij} + \beta_2 \cdot D_{ij} \cdot X_j + \text{ID}_j + \text{Issue}_i + \epsilon_{ij}, \tag{32}$$

where $\tilde{b}_{ij}$ is participant j's belief on issue i, D_{ij} is the expected disagreement dummy of participant j for issue i, X_i captures whether participant j identifies

with a progressive or conservative group (referring to economic groups in ET, $X = E$, to cultural groups in CT, $X = C$, and to economic or cultural groups in Control), ID_j to participant fixed effects, and Issue_i to issue fixed effects.[28]

Table E.4 reports the results of estimating regression Equation 32 separately all treatment conditions (whereby the uneven (even) Columns drop (include) issue fixed effects). For both ET and CT the results are consistent with our model. First, conservative participants in both ET and CT state significantly less progressive beliefs on issues in which they expect disagreement than on issues for which they do not expect disagreement (negatively estimated β_1, see Columns (1), (2), (5) and (6)). Second, we find the reverse pattern for progressive participants in both ET and CT: progressives state more progressive beliefs for issues in which they expect disagreement than for issues in which they do not expect disagreement ($\beta_1 + \beta_2$ is statistically significantly larger than zero, see again Columns (1), (2), (5) and (6)).

We do not find these patterns in the Control. Both conservatives and progressives in the Control are less progressive on questions in which they expect disagreement than on questions in which they expect no disagreement (both β_1 and $\beta_1+\beta_2$ are statistically significantly smaller than zero).[29,30] This is an important difference to the results in ET and CT and is consistent with our model. Expected disagreement is measured for placebo groups in the Control that are *not* based on existing political conflict with associated stereotypes.

F Non-Whites Culturally Conservatives

Footnote 24 in Section 5.2 suggests that CT may exhibit stronger de-polarization effects because cultural enemies are more heterogeneous than economic ones. In particular, non-White culturally conservative participants may feel little affinity with the "White Pride" enemy reported in the list, reducing their cultural identity and expected disagreement. This appendix provides supporting evidence.

[28] We do not include the progressive-membership dummy in isolation because of collinearity with the participant fixed effects.

[29] In the specifications with question fixed effects (Columns (4) and (8)), these differences in expected disagreement is significantly smaller for progressives than for conservatives (β_2 is statistically significantly larger than zero).

[30] Additional analyses (available upon request) show that in both ET and CT conservative (progressive) group members are statistically significantly less (more) progressive on questions with expected disagreement than conservative group members in the Control.

Treatment	OLS ET and CT		
Outcome	Expected Disagreement	Text Similarity	
	All (10)	Low	High
	(1)	(2)	(3)
Non-White	-0.0552***	0.0417***	-0.0585***
	(0.008)	(0.011)	(0.013)
Constant	0.703***	0.167***	0.629***
	(0.004)	(0.005)	(0.006)
Adjusted R^2	0.01	0.00	0.00
N	8474	8474	8474

Table F.5: In Column (1), we report the results of regressing the expected disagreement across the economic and cultural outcome questions of treated participants on a dummy that is equal to one if the participant self-described her race/ethnicity to be non-White or zero otherwise. In Columns (2) and (3), we report the results of regressing an indicator variable of whether a participants belongs to the Low Similarity (Column (2)) group or not, or whether a participants belongs to the High Similarity (Column (3)) group or not, on the non-White dummy. Robust standard errors in brackets. *** (**,*) indicate statistical significance at 0.01 (0.05, 0.1) level.

Table F.5 shows that non-White treated participants have significantly lower expected disagreement and lower similarity scores than White participants, and are significantly more likely to fall into the Low Similarity category. These patterns are consistent with CT failing to prime a strong cultural identity for non-White respondents.

Table F.6 reports our standard diff-in-diff regression separately for White and non-White respondents. For White respondents (Columns (1)-(4)), both ET and CT produce significant polarization, consistent with our baseline results. For non-White respondents (Columns (5)-(8)), ET effects are directionally similar to those for White respondents though smaller and insignificant for cultural beliefs, likely reflecting the substantially smaller sample size. Critically, CT produces a statistically significant negative interaction coefficient for cultural beliefs (Column (8)), indicating de-polarization — consistent with CT priming cultural groups that non-White conservatives do not identify with, thereby reducing rather than amplifying their cultural identity.

	OLS regressions							
Sample	White				Non-White			
Outcome	$\tilde{e}$		$\tilde{c}$		$\tilde{e}$		$\tilde{c}$	
Treatment	ET (X = E)	CT (X = C)	ET (X = E)	CT (X = C)	ET (X = E)	CT (X = C)	ET (X = E)	CT (X = C)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
X	0.944***	0.991***	0.978***	1.238***	0.676***	0.677***	0.650***	0.926***
	(0.028)	(0.028)	(0.028)	(0.025)	(0.057)	(0.060)	(0.061)	(0.058)
XT	-0.149***	-0.0843***	-0.0962***	0.0137	-0.140**	0.0733	-0.0690	0.164**
	(0.028)	(0.029)	(0.030)	(0.026)	(0.067)	(0.077)	(0.073)	(0.073)
X XT	0.294***	0.226***	0.118***	0.0892**	0.192**	-0.0148	0.0676	-0.179**
	(0.040)	(0.041)	(0.041)	(0.037)	(0.083)	(0.090)	(0.086)	(0.085)
Adj. R^2	0.297	0.295	0.260	0.412	0.161	0.122	0.135	0.201
Clusters	6986	6965	6986	6965	1752	1743	1752	1743

Table F.6: Regressing economic beliefs (Columns (1), (2), (5), and (6)) and cultural beliefs (Columns (3), (4), (7), and (8)) on a progressive-group-membership dummy ($X = E$ in ET, $X = C$ in CT), a treatment dummy (ET, CT), and the interaction of the two dummies, separately for White (Columns (1)-(4)) and non-White (Columns (5)-(8)) respondents. Robust standard errors in brackets. *** (**,*) indicate statistical significance at 0.01 (0.05, 0.1) level.

G Trade Conflict

In Wave 3 we included a treatment on trade conflict that combines economic features (lower consumer prices vs jobs) with cultural ones (globalization vs nationalism). We designed the trade treatment (TT) analogously to ET and CT, asking subjects to choose an enemy from a list of trade-related groups. The trade enemies are: "Big domestic businesses who make huge profits from import bans by charging higher prices to consumers", "Nationalists who believe in American superiority and oppose cooperation with foreign countries", "Cosmopolitan elites who blindly support globalization", and "Capitalists who outsource American jobs and make huge profits from exploiting workers in other countries". We define subjects who choose one of the first two to be trade progressives, and refer to them as T, and subjects who choose one of the latter two to be trade conservatives or $-T$. We also added an outcome question on trade that asks for respondents' factual assessment of whether international trade has "benefited the U.S. population economically in the last 10 years".

Table G.8 shows in Column (1) that trade conflict does not increase disagreement on pooled political beliefs (including on trade) relative to Control. Turning to respondents' expected disagreement and text data, we see that

Treatment	OLS TT vs ET&CT		
Outcome	Expected Disagreement	Text Similarity	
	All (11)	Low	High
	(1)	(2)	(3)
TT	-0.0846***	0.0393***	-0.0748***
	(0.011)	(0.013)	(0.016)
Constant	0.680***	0.166***	0.586***
	(0.006)	(0.007)	(0.009)
Adjusted R^2	0.01	0.00	0.00
N	4311	4311	4311

Table G.7: In Column (1), we report the results of regressing the expected disagreement across all the standard 10 outcome questions and the trade question of treated participants in Wave 3 on a treatment trade dummy. In Columns (2) and (3), we report the results of regressing an indicator variable of whether a participants belongs to the Low Similarity (Column (2)) group or not, or whether a participants belongs to the High Similarity (Column (3)) group or not, on a treatment trade dummy. Robust standard errors in brackets. *** (**,*) indicate statistical significance at 0.01 (0.05, 0.1) level.

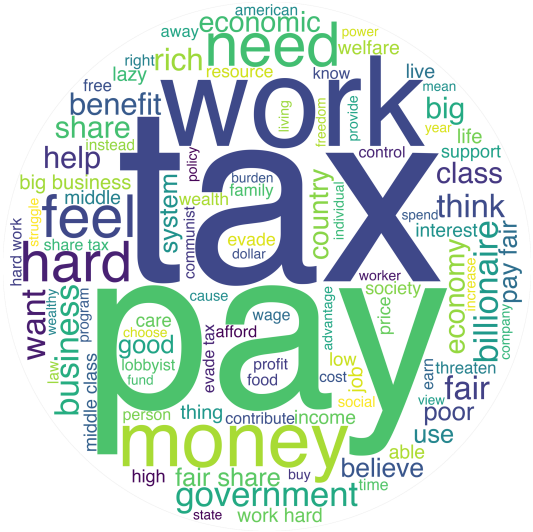
trade conflict led to significantly fewer identified respondents, see Table G.7. When we account for HED vs LED and High vs Low similarity, we find a large positive trade conflict effect on the trade outcome for respondents who expect disagreement on trade (HED) as well as who received a large similarity score in their ingroup description (High) and large backfiring effects on trade for respondents who expect no disagreement on trade (LED) as well as who received a low similarity score (Low), see Columns (3)-(10) in Table G.8.

Treatment	OLS regressions TT vs Control									
Sample	All		HED		LED		High		Low	
Oucome	all	t	all	t	all	t	all	t	all	t
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
T	1.004***	0.849***	0.988***	0.873***	1.086***	0.972***	1.004***	0.849***	1.004***	0.849***
	(0.095)	(0.132)	(0.095)	(0.154)	(0.098)	(0.147)	(0.095)	(0.132)	(0.095)	(0.132)
TT	0.0704	-0.324**	0.0389	-0.395**	0.204	0.337*	-0.126	-0.573***	0.390**	0.117
	(0.100)	(0.129)	(0.107)	(0.154)	(0.173)	(0.174)	(0.115)	(0.150)	(0.164)	(0.243)
T TT	-0.219	0.192	0.0323	0.843***	-0.765***	-0.800***	0.494***	0.812***	-1.208***	-0.570
	(0.144)	(0.205)	(0.155)	(0.259)	(0.242)	(0.274)	(0.182)	(0.260)	(0.233)	(0.358)
Constant	5.082***	5.339***	5.156***	4.709***	4.778***	5.573***	5.082***	5.339***	5.082***	5.339***
	(0.067)	(0.084)	(0.067)	(0.093)	(0.067)	(0.097)	(0.067)	(0.084)	(0.067)	(0.084)
Adjusted R^2	0.05	0.03	0.06	0.04	0.06	0.02	0.08	0.04	0.05	0.02
N	3001	3001	2667	2375	1969	2261	2335	2335	1951	1951

Table G.8: We report the results of the our difference in differences regression comparing the disagreement between T and $-T$ in Control to the disagreement between T and $-T$ in TT. Uneven Columns and even Columns focus on the outcomes of all pooled political beliefs (including on trade) and only trade beliefs, respectively. Columns (1) and (2) focus on the entire sample of TT and Control. Columns (3)-(6) focus on HED or LED treated participants relative to their corresponding synthetic Control. Columns (7)-(10) focus on High Similarity or Low Similarity treated participants relative to Control. Robust standard errors in brackets. *** (**,*) indicate statistical significance at 0.01 (0.05, 0.1) level.

H Additional Figures and Tables

Panel A:



Panel B:



Figure H.10: Panel A (B): Word cloud of subjects' ingroup descriptions and enemy explanations in ET (CT).

Party Identity Type	Social Identity Type (E,C) %	(-E,C) %	(E,-C) %	(-E,-C) %	Total %
Democrats	27.7	4.7	3.9	3.1	39.4
Republican	2.7	2.8	6.7	16.9	29.1
Independents	12.4	4.1	6.4	8.6	31.5
Total	42.7	11.6	17.0	28.7	100.0

Table H.9: Ingroup choices and political affiliation in Control

	Covariance		
	$(\tilde{e}, \tilde{c})$	$(\tilde{e}_f, \tilde{c}_f)$	$(\tilde{e}_p, \tilde{c}_p)$
Control	3.4022	2.5766	4.1701
ET	3.8402***	2.9900***	4.6243***
CT	4.0484***	3.2591***	4.7299***

Table H.10: Covariances in beliefs across the economic and cultural domain (both factual and policies in Column (1), only facts in Column (2), and policies in Column (3)) by treatment. *** indicate statistically significant differences between ET or CT and Control at 0.01 level.

	OLS regressions							
Treatments:	ET (X = E)				CT (X = C)			
Outcome:	$\tilde{e}_f$	$\tilde{e}_p$	$\tilde{c}_f$	$\tilde{c}_p$	$\tilde{e}_f$	$\tilde{e}_p$	$\tilde{c}_f$	$\tilde{c}_p$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
X	0.713***	0.842***	0.848***	0.887***	0.675***	0.926***	1.073***	1.153***
	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)	(0.026)	(0.025)	(0.024)
XT	-0.146***	-0.121***	-0.0700**	-0.0986***	-0.0607**	-0.0563*	0.0341	0.0250
	(0.028)	(0.029)	(0.029)	(0.030)	(0.028)	(0.030)	(0.027)	(0.027)
X XT	0.251***	0.232***	0.106***	0.101***	0.164***	0.163***	0.0802**	0.0120
	(0.038)	(0.038)	(0.039)	(0.039)	(0.039)	(0.038)	(0.035)	(0.035)
Adjusted R^2	0.18	0.23	0.20	0.22	0.14	0.25	0.32	0.34
N	8738	8738	8738	8738	8708	8708	8708	8708

Table H.11: Regressing beliefs about economic facts (Columns (1) and (5)), economic policies (Columns (2) and (6)), cultural facts (Columns (3) and (7)), and cultural policies (Columns (4) and (8)) on a progressive-group dummy ($X = E$ in ET, $X = C$ in CT), a treatment dummy (ET, CT), and the interaction of the two dummies. Robust standard errors in brackets. *** (**,*) indicate statistical significance at 0.01 (0.05, 0.1) level.

Outcome	OLS regressions General Progressivity					
Treatment	ET and CT (X = E,C)					
Sample	Independents		Democrats		Republican	
	(1)	(2)	(3)	(4)	(5)	(6)
X	0.909***	0.818***	0.712***	0.611***	0.530***	0.419***
	(0.036)	(0.035)	(0.040)	(0.038)	(0.037)	(0.034)
XT	-0.0509		-0.0481		-0.0827***	
	(0.036)		(0.049)		(0.026)	
X XT	0.162***	0.156***	0.142***	0.123**	0.138***	0.153***
	(0.047)	(0.046)	(0.053)	(0.052)	(0.051)	(0.050)
Controls	No	Yes	No	Yes	No	Yes
Adj. R^2	0.296	0.367	0.149	0.236	0.126	0.216
Clusters	3815	3815	4730	4730	3581	3581

Table H.12: In Column (1) we regress general progressivity (i.e. averaged and standardized beliefs) on a progressive-group dummy ($X = E$ in ET, $X = C$ in CT), a treatment dummy (ET, CT), and the interaction of the two dummies for self-reported independents, pooling all treatments. Column (2) adds to the regression in Column (1) demographic controls and their interactions with the treatment dummies. Columns (3) and (4) [(5) and (6)] repeat Columns Columns (1) and (2) but for Democrats [Republicans]. Robust standard errors in brackets. *** (**,*) indicate statistical significance at 0.01 (0.05, 0.1) level.

Outcome	OLS Average beliefs: $\tilde{e}$ and $\tilde{c}$					
Ingroup	Broad			Narrow		
Treatments	ET	CT	Control	ET	CT	Control
	(1)	(2)	(3)	(4)	(5)	(6)
Ingroup Beliefs	0.864***	0.966***	0.980***	0.857***	0.968***	0.980***
	(0.046)	(0.033)	(0.016)	(0.034)	(0.021)	(0.026)
Ingroup-Outgroup Beliefs	0.160***	0.0578***	0.0120	0.233***	0.0832***	0.00102
	(0.034)	(0.017)	(0.014)	(0.036)	(0.019)	(0.026)
Constant	0.670**	0.281	-0.0439	0.732***	0.266**	-0.0401
	(0.249)	(0.180)	(0.081)	(0.189)	(0.115)	(0.147)
Adjusted R^2	0.99	0.99	0.99	0.97	0.98	0.97
N	20	20	20	100	100	100

Table H.13: We report the results of estimating our kernel of truth regression Equation (14) separately for each treatment and for broad groups in Columns (1)-(3) or narrow groups in Columns (4)-(6). For Control, we conduct a placebo regression. A random subset of Control respondents is assigned to the placebo treatment group. The remaining respondents are the placebo Control group. We then regress average beliefs of issue i of the placebo treatment group, separately for E and $-E$ respondents, on the average beliefs of issue i of the placebo Control group, again separately for E and $-E$ respondents, and on the average disagreement on issue i between E and $-E$ respondents in the placebo Control. Robust standard errors in brackets. *** (**,*) indicate statistical significance at 0.01 (0.05, 0.1) level.

Outcome:	OLS regressions LED			
Treatments	ET and CT			
	(1)	(2)	(3)	(4)
Survey Time Overall	-0.000138	-0.000120		
	(0.000)	(0.000)		
Outcome Stage Time			-0.0000843	-0.0000700
			(0.000)	(0.000)
Constant	0.114***	0.296***	0.113***	0.296***
	(0.004)	(0.052)	(0.003)	(0.052)
Controls	No	Yes	No	Yes
Adjusted R^2	0.00	0.03	0.00	0.03
N	8470	8470	8474	8470

Table H.14: In Columns (1) and (2) [(3) and (4)] we regress treated respondents' LED status on their overall survey time [their time spent during the outcome stage]. Robust standard errors in brackets. *** (**,*) indicate statistical significance at 0.01 (0.05, 0.1) level.