Morals as Luxury Goods
and Political Polarization[*]

Benjamin Enke   Mattias Polborn   Alex A. Wu

April 21, 2022

Abstract

This paper develops a theory of political behavior in which moral values are a luxury good: the relative weight that voters place on moral rather than material considerations increases in income. This idea both generates new testable implications and ties together a broad set of empirical regularities about political polarization in the U.S. For instance, the model predicts (i) the emergence of economically left-wing elites; (ii) that more rich than poor people vote against their material interests; (iii) that within-party heterogeneity is larger among Democrats than Republicans; and (iv) realignment patterns that reflect an increased relevance of moral rather than class-based divisions: rich moral liberals who switch to the Democrats, and poor moral conservatives who swing Republican. Assuming that parties set policies by aggregating their supporters' preferences, the model also predicts increasing polarization on social issues over time. We relate these predictions to known stylized facts, and test our new predictions empirically.

"Erst kommt das Fressen, dann kommt die Moral." – "A hungry man has no conscience."

Bertolt Brecht, The Threepenny Opera

1   Introduction

A substantial and growing body of work documents that, in addition to the economic dimension traditionally emphasized by political economists, a social or cultural dimension is relevant for understanding political behavior. At the party level, candidate surveys and text analyses suggest that the ideological divide in Congress is increasingly on social and moral rather than economic issues (e.g., Moskowitz, Rogowski and Snyder, 2018; Gentzkow, Shapiro and Taddy, 2019; Enke, 2020).[1] Similarly, at the voter level, much work argues for an important role of morality and identity for political behavior (e.g., Ansolabehere, Rodden and Snyder, 2006; Shayo, 2009; Haidt, 2012; Besley and Persson, 2019; Mason, 2018; Bursztyn, Egorov and Fiorin, 2020; Enke, Rodríguez-Padilla and Zimmermann, 2019; Enke, 2020; Bonomi, Gennaioli and Tabellini, 2021), to the degree that various researchers emphasize an increasing "cultural divide" between Democrats and Republicans (Bertrand and Kamenica, 2018; Desmet and Wacziarg, 2021). Indeed, widely emphasized realignment patterns according to which the educational elite has swung Democrat, while the white, rural working-class swung Republican, are typically attributed to an increased relevance of moral rather than class-based divisions (Frank, 2007; Gethin, Martínez-Toledano and Piketty, 2021; Krasa and Polborn, 2014a).

While this burgeoning empirical literature shows that morals (broadly understood) generally matter for political behavior, it also raises many conceptual questions. How do morals affect the structure of political preferences? When and why do voters prioritize their moral over their material concerns? Which positions do parties take in response? What are the implications for political polarization, both in the electorate and between party positions? Why did the salience of moral divisions increase over time even though income inequality increased?

This paper formally studies the simple idea that *morals are luxury goods*: the relative weight that people place on moral versus material considerations increases in their absolute income. We show that this simple and evidence-backed assumption both generates various new testable implications and ties together a broad set of empirical regularities about the nature of political polarization in the United States.

The idea that morals constitute luxury goods is not new but has appeared in different terminology across the social sciences, such as in Maslow's (1943) "hierarchy of needs," the influential "postmaterialism" literature initiated by Inglehart (1997, 2020) or the argument

---

[1]Using candidate surveys, Moskowitz et al. (2018) find "increased polarization over social issues, [and] . . . little evidence of growing partisan differences over economic issues." In Congressional speeches, the strongest increase in polarization occurs on topics such as immigration, minorities, religion, environment and basic moral values, while many economic topics such as budget, business, economy or trade have seen little increase in polarization (Gentzkow et al., 2019; Enke, 2020).

that modernization increases demand for democracy (Lipset, 1959). A large body of empirical evidence supports the general idea of morals as luxury goods: starting with Inglehart's work, many researchers have used the World Values Survey to document that richer people consistently report placing greater importance on values and rights compared to economic growth and material security. These trends materialize in two ways: first, in any given survey year, rich people report being less materialist than the poor. Second, as average incomes increased over time, the U.S. population as a whole became less materially-oriented.

Building on this body of work, we posit that people also care more about expressing their morals at the voting booth when they worry less about putting food on the table.[2] Indeed, the recent "Hidden tribes" report on U.S. political polarization highlights that those voter groups that place a heavy emphasis on moral issues ("devoted conservatives" and "progressive activists") are the *richest* voter groups (Hawkins, Yudkin, Juan-Torres and Dixon, 2019).

In our model, voters' primitive types consist of their income and their moral values. Moral values determine the voter's ideal social policy. A voter's ideal economic policy (such as the level of redistributive taxation) is jointly determined by their relative income position in society and their moral values. Here, consistent with the extant evidence, we assume that moral liberals perceive a moral obligation to help the poor through government redistribution, while moral conservatism implies lower desired governmental redistribution.

Our main assumption is that while all voters (including the poor) potentially place non-zero utility weight on their morals, this weight strictly increases in absolute income. To see the most basic implication this assumption, consider the canonical comparative static effect of an increase in a voter's relative income on his ideal economic policy. In traditional models such as Meltzer and Richard (1981), richer voters have unambiguously more conservative ideal economic policies. In our model, however, income-dependent utility weights generate an additional indirect effect of higher income. For moral conservatives, whose morally-optimal economic policy is more conservative than their materially-optimal policy, a higher weight on moral values moves the ideal economic policy in a conservative direction. For moral liberals, however, this indirect effect points in the opposite direction as they can now "afford" to prioritize their liberal values over their material concerns. This produces a version of the "gauche caviar:" rich voters who support left-wing economic policies. As a result, our model predicts that the desired economic policies of rich and poor moral liberals are more similar to each other than those of rich and poor moral conservatives. An analogous logic applies to voting: as voters become richer, their moral values become more important to them, which implies that income is more strongly predictive of voting among moral conservatives than among moral liberals.

We empirically study these new predictions in two datasets, the American National Election Studies (ANES) and a nationally representative Moral Foundations Questionnaire (MFQ) dataset. In line with the predictions, we find that the effect of an income increase on

---

[2]As discussed in Section 3, decreasing marginal utility as such could also imply that the relative weight placed on moral considerations increases in income, but this is not generically true in a voting context.

a voter's desired economic policy is almost three times as large among moral conservatives than among moral liberals. Similarly, the positive effect of income on voting Republican is substantially more pronounced among moral conservatives than among moral liberals.

An alternative way of describing these results is that the effect of moral views on voting is larger among the rich. This has direct implications for who is more likely to "vote against their economic interests." Consider two "off-diagonal" groups of voters that have attracted considerable attention in the recent popular debate: voters who are either rich-and-morally-liberal or poor-and-morally-conservative, such that they align with each party on only one dimension. An influential body of narratives argues that the Republican party has convinced poor moral conservatives to accept low taxes on the rich, in exchange for the conservative social policies that they supposedly desire (Frank, 2007; Hacker and Pierson, 2020). Yet, while much of the public discussion tends to focus on the "puzzle" that poor moral conservatives vote Republican, it could be considered equally puzzling that rich moral liberals (such as the educational elite) often vote Democrat. Indeed, our model predicts that rich moral liberals are *more* likely to vote against their material interests than poor moral conservatives. The simple intuition is that both groups need to trade off their economic and moral objectives, and rich moral liberals place a higher weight on moral issues than poor moral conservatives. We again test our prediction empirically. In both the ANES and the MFQ, voters who are rich and morally liberal are more than 20 percentage points more likely to vote Democratic than poor and morally conservative people are to vote Republican.

The fundamental asymmetry in terms of whether the rich and poor vote based on material or moral considerations also has implications for understanding within-party supporter heterogeneity. In our model, people vote Democratic either because they are poor (in this case, more-or-less independent of their moral values, because the weight on these is relatively small), or because they are rich and have very liberal values. For Republican voters, on the other hand, the mirror image of this pattern is less likely to happen: the very poor will less often vote Republican even if they have conservative values because – being poor – they place a lower weight on their values. As a result, our model predicts that the set of Republican voters is less heterogeneous than the set of Democratic ones.

This prediction jives with popular media accounts that highlight the internal tensions between various factions within the Democratic party. At the same time, we are not aware of formal models that have made such a prediction. We again take this prediction to the data in the ANES and MFQ. Consistent with the model, Democratic voters exhibit a significantly greater standard deviation in both income and moral/ social views than Republican ones.

All of the cross-sectional predictions summarized above are quite particular to our model. Indeed, a two-dimensional setup with utility weights that are constant in income does not generate these predictions. We then move on to exploring changes over time. Here, we focus on secular trends and the effects of long-run economic growth. For the sake of analytical simplicity, we first consider a setup in which all voters' incomes increase by the same multiplicative factor. Thus, all voters' relative economic positions remain constant, but the

weight that everyone in society places on moral values increases.

This setup generates a simple insight about the time-varying nature of ideological constraint (intra-correlations among people's economic and social policy views). In our model, because moral values affect both voters' ideal social policies and their ideal economic policies, an increased weight on moral values implies that voters' ideal economic and social positions become more correlated over time, which is a stylized fact (Kozlowski and Murphy, 2019; Wu, 2020). At the same time, our account adds a new prediction, namely that the increase in ideological constraint over time should be driven by that segment of the population whose incomes grew most (the rich). In ANES data, we find strong support for this prediction: the increase in ideological constraint over the last 30 years is more than twice as large among the top 1/3 of the income distribution than among the bottom 2/3.

The implications of our model for changes in voting behavior over time are somewhat involved because two effects occur at the same time: (i) For fixed party positions, changes in income induce voters to place a higher weight on moral values, which affects their vote choice. (ii) Party positions may change – possibly in response to changes in voter demand for different positions –, which itself leads to changes in voting behavior.

While we ultimately analyze these mechanisms jointly through simulations in a general equilibrium framework, it is instructive to additionally analyze them separately in partial equilibrium. Regarding (i), rising absolute incomes mean that voters place a higher weight on moral issues in the later period. As a result, "off-diagonal" voters become more likely to vote in line with their moral rather than their material concerns, such that poor moral conservatives (such as the white rural working class) swing towards the Republican party, while rich moral liberals (the canonical "educational elite") swing towards the Democrats.

Regarding (ii), we analytically show that exogenous social party polarization likewise generates the realignment discussed in the previous paragraph: when the parties diverge on social issues, voters' moral values become more decisive for their vote choice because their utility difference from the two parties' social policies increases, even for fixed utility weights. As a result, if social party positions polarize over time, poor moral conservatives (liberals) again become more likely to vote Republican (Democratic).

In Section 6, we put these pieces together in a framework in which party positions are endogenous and determined in equilibrium, along with agents' vote choices. This requires us to explicitly model how party positions are determined. We do so by assuming that party positions aggregate the preferences of party supporters. We discuss the various potential interpretations of this assumption in detail in Section 6.[3]

Because this two-dimensional model cannot be solved in closed form, we proceed nu-

---

[3]For example, our setup could be interpreted along the lines of Polborn and Snyder (2017) who model Congressional elections, in which voters in different districts choose between local candidates. Once elected, these local candidates all have an incentive to influence their party's position in the direction preferred by their local constituents (e.g., to ward of primary challenges). Then, the reduced-form assumption that parties aggregate the preferences of their supporters captures the idea that parties are not internally cohesive entities, but that they are influenced by the demands of elected representatives who face heterogeneous reelection incentives themselves.

merically. The objective of this analysis is to work out the robust comparative statics implications of our model, rather than to offer a calibrated model of the U.S. economy. A robust result of our simulations is that, as incomes increase over time, parties polarize on social issues. Intuitively, as voters get richer, they care more about moral issues, inducing some rich morally liberal voters to switch from Republicans to Democrats and some poor moral conservatives to swing Republican. Because parties maximize the average utility of their voters, this induces the Democratic party to move to the left and the Republicans to the right on social issues. This, in turn, induces further realignment among voters with relatively liberal or conservative values, which triggers new changes in party positions, until this process equilibrates at a higher level of social polarization. Thus, our "morals as luxury goods" approach offers a new lens through which the increased relevance of moral rather than class-based dimensions of political conflict and corresponding realignment patterns can be understood.

By a similar logic, our simulations robustly generate the well-known empirical regularity that the distributions of moral values of Republican and Democratic voters become less overlapping and more extreme over time, which is commonly referred to as "moral polarization in the electorate". Yet, in our model, this is not because the underlying distribution of morals in the population has changed, but purely a result of voter sorting. This result sheds light on a large body of empirical work in political science that has documented that increasing social and moral polarization in the general public is largely driven by voter sorting (e.g., Fiorina, Abrams and Pope, 2006; Levendusky, 2009).

In our baseline analysis, we assume that all voters' incomes increase by the same multiplicative factor over time. This is analytically convenient but counterfactual. Extending our approach to increasing income inequality, we use a combination of analytical derivations and model simulations to shed some light on the oft-discussed puzzle that rural working-class voters swung Republican over time *even though* their relative incomes decreased over time. In our model of morals as luxury goods, this pattern can easily occur even if income increases are concentrated at the top. The logic is that when the rich get disproportionately richer, they place a higher weight on moral considerations, which induces some rich moral liberals to swing Democratic. This, in turn, induces the parties to polarize on social issues because their voter bases have now both become more extreme. Faced with such socially increasingly polarized parties, a poor morally conservative voter may well become more likely to vote Republican, even when his materially-preferred economic policy has moved to the left as a result of increased income inequality. In turn, when poor moral conservatives swing Republican, this further pushes the Republican party position on social issues to the right and the Democratic one further left. Thus, in our model, poor moral conservatives are both "pushed" to vote Republican by elite-driven social party polarization, but also "push" the parties to more extreme positions themselves.

We view this set of results as consistent with a large body of qualitative sociological work that argues that poor, rural whites feel that the Democratic party has "abandoned"

them in favor of "extremist" social policy positions that cater to urban moral liberals (e.g., Cramer, 2016; Sherman, 2009). Indeed, in line with the results of the "Hidden tribes" report (Hawkins et al., 2019), our model and empirical results suggest that political polarization is partly the result of a culture war between two subsets of the elite, who push the parties to become morally more extreme.

To sum up, a simple and evidence-backed assumption – that the relative utility weights of material and moral considerations depend on income – sheds light on three of the most widely-discussed empirical regularities regarding U.S. political polarization: increasing party polarization on social policy issues, realignment of rich moral liberals and poor moral conservatives, and increasing moral polarization of the electorate. Moreover, our approach generates various new testable predictions that find support in the data, including (i) steeper voting-income and economic conservatism-income gradients among moral conservatives than among moral liberals; (ii) that more rich moral liberals than poor moral conservatives vote against their material interests; (iii) greater within-party heterogeneity on the left than on the right and (iv) increased ideological constraint, especially among the rich.

The paper proceeds as follows. Section 2 briefly describes the data we use to test our theoretical predictions. Section 3 exposits the model. Sections 4 and 5 analyze the model in the cross-section and over time. Section 6 defines endogenous party positions and simulates the general equilibrium model. In Section 7, we discuss how our model setup and results relate to the literature, and Section 8 concludes. All proofs of propositions are in Appendix A.

2   Data

Because various predictions of our model will be new, we present theoretical propositions interwoven with empirical analyses that are based on two complementary datasets: the American National Election Studies (ANES) and the Moral Foundations Questionnaire (MFQ). The ANES has the advantages that it allows for (i) time series analyses and (ii) the construction of separate economic and social conservatism indices that we will think of as proxies for voters' ideal economic and social policy positions. The downside of the ANES is that it does not contain many direct questions on basic moral values. The MFQ dataset assembled by Enke (2020), on the other hand, only consists of one cross-section and does not contain rich information on economic and social policy views, but it allows for the construction of a moral values index. We briefly describe the key variables here; details are in Appendix E.

American National Election Studies.   Since 1988, the ANES has consistently included questions on economic issues such as social security and federal spending on public schools, and on social issues such as gay rights and the importance of family values. We adopt the methodology of Ansolabehere et al. (2006) in aggregating survey questions to construct indices of economic and social conservatism. We classify issues as either economic or moral, and then compute weighted averages, where the weight placed on each question is

determined by its loading in a principal component analysis.

The income variable in the ANES consists of five buckets that partition the population into income percentiles. In our analyses below, we will either work with this variable directly, or impute (log) average incomes in the respective income percentile and survey year, using data from the Tax Policy Center. See Appendix E for details.

Moral Foundations Questionnaire. The MFQ dataset assembled by Enke (2020) consists of a nationally representative sample of 4,011 respondents. The data include information on voting in the 2016 Presidential election, household income, and a summary statistic of moral values. The MFQ is an influential psychological questionnaire. Its main idea is that individuals exhibit strong heterogeneity in how much emphasis they place on communal relative to universalist values (Haidt, 2012). Here, universalist concepts such as individual rights, justice, and impartial fairness apply irrespective of the context or identity of the people involved. Communal concepts, on ther other hand, are tied to certain relationships or groups, and emphasize the moral relevance of community, loyalty, respect, and tradition. Based on the MFQ, Enke (2020) develops a "moral conservatism" summary statistic of the relative importance of communal versus universalist moral values.

3   Model

3.1   Preferences

To preview, our model of preferences rests on two main assumptions. First, an agent's moral values affect both his economic and his social policy views. Second, the weight that an agent places on moral relative to material considerations increases in income. That is, as far as voting decisions are concerned, we will think of morality as a luxury good. It is convenient to write all relationships that depend on income as functions of the natural logarithm of income, denoted y. We assume $y \geq 0$ for all voters.

We assume that the deterministic part[4] of an agent's overall utility is given by a weighted average of two components, which we call material (M) and philosophical utility (P):

$$U = [1 - w(y)]M + w(y)P. \tag{1}$$

Here, the weight function ranges from 0 to 1, and is increasing and concave. For example, in the general equilibrium section and for some illustrations, we use $w(y) = \frac{c+\rho(y-k)}{1+\rho(y-k)}$, where $c \in (0,1)$ is a lower bound on the weight placed on philosophical utility, and $k \in [0, y)$ and $\rho > 0$ are parameters.[5] We can approach a standard framework (i.e., one where the weight

---

[4]When voting, individuals also receive idiosyncratic utility shocks, for example from the quality of candidates. We will introduce these (ex-ante stochastic) components further below.

[5]Including k in addition to ρ is a useful generalization because it allows to scale the level and the responsiveness of the weights separately. With $k = 0$, a high degree of responsiveness ρ implies that the weights are always close to one. By including both k and ρ, it is possible to have a strong

placed on philosophical utility is constant in income)[6] by letting $\rho \to 0$, so that the utility function collapses into one in which voters place constant weights c and $1-c$ on philosophical and material utility, respectively.

We do not assume that the moral utility weight is zero (or negligible) for relatively poor voters. It may well be that the poorest voter in the U.S. places a non-trivial weight on moral issues – we only assume that the weight increases in income.

Discussion. The idea that morals are luxury goods is not new but follows various prior literatures across the social sciences. Our contribution is to incorporate this idea into a model of political behavior. We here discuss interpetations of this assumption and associated evidence.

A first interpretation is that people's true utility weights actually change as they become richer. For instance, in standard household theory, some physical goods are inferior and some are normal. In much the same way, it would be very surprising if the rate at which voters trade off economic and moral policy positions is unaffected by their wealth. As Brecht famously writes in the Threepenny Opera, "Erst kommt das Fressen, dann kommt die Moral."[7] Relatedly, in his essay on morality and markets, Shleifer (2004) concludes that "as societies grow richer, their willingness to pay for ethical behavior ...increases as well." Outside of economics, this argument is typically referred to as "Maslow's hierarchy of needs," according to which values and self-actualization occupy a higher step in the hierarchy of human needs than material concerns and only become relevant to people when they have achieved material security (Maslow, 1943).

While diminishing marginal utility is a plausible reason for why higher incomes lead to a higher effective weight on moral issues, this is not mathematically true for all concave utility functions. The reason is that there are two countervailing effects. First, as people get richer, their marginal utility of money decreases, which makes morality more important at the margin. Second, however, as people get richer, economic policy also becomes more important to them in absolute terms as their tax base increases.[8] Nevertheless, it is clearly conceivable that the effective relative weight that a voter puts on moral issues changes with income due to diminishing marginal utility.

A second interpretation for why moral issues may become more important to people as they get richer is that voters have identity concerns, and the degree to which they derive utility from signalling their values or identity (either to others or to themselves) may increase

---

responsiveness without letting the weights tend to one.

[6] For models of electoral competition in multidimensional settings where voters have constant weights on different issues, see, e.g., Schofield (2007); Bade (2011); Krasa and Polborn (2014b,c); Dragu and Fan (2016); Nunnari and Zápal (2017); Xefteris (2017).

[7] Almost verbatim translated, "The first priority is food, morals come second." Or: "A hungry man has no conscience."

[8] For example, in a Meltzer-Richards model of redistributive taxation in which the utility from consumption is logarithmic, and the utility function is augmented by an additive term that captures the voter's moral concerns, the aforementioned effects cancel each other exactly out.

in income. For example, people may categorize themselves into a "morally liberal" or "morally conservative" social identity, and that social identity may become more important to them as they get richer.

Regardless of which interpretation is adopted, the broad idea that the importance placed on non-material considerations increases in income is supported by a wealth of evidence. In particular, in a highly influential line of work, the sociologist Ronald Inglehart and his collaborators devised the now-famous World Values Survey to document that, as people and societies get richer, they place increasing weight on non-material versus material considerations (Inglehart, 1997; Inglehart and Baker, 2000; Inglehart, 2020). The available evidence strongly suggests that this "modernization hypothesis" or "postmaterialism hypothesis" applies both over time and in the cross-section: (i) people in the U.S. today place much more emphasis on non-material considerations than they used to 40 years ago and (ii) in any given survey year, the rich place a higher emphasis on non-material issues than the poor. Broadly related to this body of evidence is also work in economics that shows that people place a higher weight on job meaning compared to wages if they grew up during economic booms rather than recessions (Cotofan, Cassar, Dur and Meier, 2020).[9]

There is also evidence to suggest that the idea of morals as luxury goods extends to the political domain (Gelman, Shor, Bafumi and Park, 2007). Indeed, the recent "Hidden tribes" report (Hawkins et al., 2019) highlights that those voter groups that are especially occupied with moral issues such as fairness or abortion ("devoted conservatives" and "progressive activists") are considerably richer than other voter groups.

Finally, note that our morals-as-luxury-goods approach differs from an approach in which income affects *which* values people hold. A common intuition is that higher income may make people socially more liberal. However, the conventional U.S. datasets that we also use here speak against such a mechanism. In the ANES, the correlation between income and social conservatism is only $r = -0.10$, and in the MFQ it is even positive ($r = 0.06$). This provides some justification for our approach of neglecting a potential dependence of μ on y.

Material utility. Material utility is a function of individual income and the implemented economic policy x_e. Rather than setting up an explicit model of material utility, we employ a reduced form one, which is compatible with several interpretations:

$$M(x_e, y, \bar{y}) = y - \frac{1}{2}\left[x_e - \underbrace{(r(y - \bar{y}))}_{\substack{\text{Materially-optimal} \\ \text{economic policy}}}\right]^2. \tag{2}$$

Here, $y \geq 0$ denotes individual (log) income, and $\bar{y}$ average (log) income in society. We assume that $r(\cdot)$ is an increasing and differentiable function of the individual's relative

---

[9]Broadly related is also a body of work that argues that higher income (or education) may lead to higher demand for democracy, which could be viewed as placing higher weight on moral issues (e.g., Lipset, 1959; Glaeser, Ponzetto and Shleifer, 2007).

economic position in society.[10]

Many natural economic models have the property that materially-optimal policies are monotone in relative income. For example, in a model of redistribution, richer agents prefer lower taxes. Similarly, if the level of taxation determines the amount of public goods that can be provided to citizens, the ideal level of taxation will also depend on relative income.

Note that (2) implies that, if all citizens' incomes increase by the same factor (i.e., log incomes increase by the same constant for all voters), the respective materially-optimal policy is unchanged for all citizens. Also, the equivalent variation of the utility loss from any non-optimal economic policy, relative to the individual's optimal policy, is a constant fraction of income that is unchanged if all citizens' incomes increase by the same factor. Thus, economic interests as captured by (2) scale proportionately.

Philosophical utility. Philosophical (or moral) utility P depends on the deviation of economic and social policy from one's ideal moral positions. Similarly to Bonomi et al. (2021), we assume that an agent's moral values matter not just for how he thinks about social policy but also for his assessment of economic policy. For example, the agent may have a bliss point that determines which tax rate or level of redistribution he finds morally fair. There is a fundamental parameter μ that determines those social and economic policies that the individual considers ideal from a purely moral perspective. Philosophical utility is given by

$$P(x_e, x_s, \mu) = -(1/2) \left[x_e - \underbrace{h(\mu)}_{\substack{\text{Morally-optimal} \\ \text{economic policy}}} \right]^2 - (1/2)(x_s - \mu)^2, \qquad (3)$$

where $h(\cdot)$ is an increasing differentiable function capturing how the morally-optimal economic policy depends on μ; x_s is social policy, i.e., on issues such as immigration, environmental protection or abortion. Our formulation implies that an agent who believes the morally appropriate economic policy is more conservative also believes that the morally appropriate social policy is more conservative. This assumption is directly supported by data. Enke et al. (2019) show that a person's degree of moral universalism is strongly correlated with their views on a number of different economic and social policy issues. According to this interpretation, low μ corresponds to high universalism, which translates into both left-wing economic views (support for welfare and universal health care) and left-wing social views (pro-immigration and affirmative action).[11]

---

[10]Since y and $\bar{y}$ are logarithms, the fact that $r(\cdot)$ depends on their difference is equivalent to assuming that a voter's optimal economic policy depends on the ratio of his and average income.

[11]In the data, we cannot separately observe the voter's morally-optimal economic policy, only their overall desired economic policy. However, we can still back out that a voter's morally-ideal economic and their morally-ideal social policy are correlated. The reason is as follows. In Enke et al. (2019), moral preferences are correlated in more or less the same way with economic and social policy views. Given that income and moral preferences are essentially uncorrelated, the only mechanism through which (in our model) moral preferences μ can be predictive of economic policy views and

Utility function. Substituting (2) and (3) into (1) yields

$$U = [1 - w(y)]\left\{ y - \frac{1}{2}\left[x_e - r\left(y - \bar{y}\right)\right]^2 \right\} - \frac{w(y)}{2}\left[(x_e - h(\mu))^2 + (x_s - \mu)^2\right] \qquad (4)$$

The key characteristic of this utility function is that the utility derived from social policy is a strongly superior good (luxury good). To see this, consider the equivalent variation of an individual's utility loss due to social policy differing from the individual's ideal position μ. Suppose the individual's income increases. For fixed weight w, the equivalent variation would increase proportionately with income. However, since w is increasing in y, the equivalent variation increases more than proportionately with income.

3.2 Ideal Policies

We now turn to characterizing a voter's ideal economic and social policies that maximize (4). Clearly, $x_s^* = \mu$ is optimal in terms of social policy. Maximizing with respect to x_e gives

$$x_e^* = [1 - w(y)]r\left(y - \bar{y}\right) + w(y)h(\mu). \qquad (5)$$

Thus, x_e^* is a weighted average of the materially-optimal economic policy and the morally-optimal economic policy, where the weight on the morally optimal policy increases in income. It is easy to see that the ideal economic policy increases in moral conservatism μ, but the relationship between income and the ideal economic policy is more involved.

3.3 Voting

Let x_e^i and x_s^i be the economic and social positions of party i. In Sections 4 and 5, party positions are exogenous, and we assume that, on both dimensions, Republican positions are more conservative than Democratic ones (i.e., $x_e^D < x_e^R$ and $x_s^D < x_s^R$). In Section 6, we endogenize the party positions.

In addition to their policy utility, voters also receive an idiosyncratic utility shock ε, which could reflect local candidates' competence or the voter's evaluation of differences between the parties beyond the economic and moral policy dimensions that are at the center of our model. The utility shock enters additively to utility from party positions, and is measured in a way that positive realizations indicate a non-policy preference for the Democratic party. That is, voter j supports the Republican party if and only if

$$U_j^R - U_j^D > \varepsilon_j,$$

where the left-hand side is voter j's net policy preference for the Republican positions. Let the density of the net utility shock ε be denoted by ψ, and assume that it is always positive,

---

social policy views in the same way, is if the morally-ideal economic policy is correlated with the morally-ideal social policy. This amounts to saying that they both depend on μ.

with associated cumulative distribution Ψ. Then, the probability that a voter votes for the Republican candidate is given by $\Psi(U^R - U^D)$, where, of course, the policy utility difference depends on the voter's type (y, μ).

It is useful and without loss of generality to normalize policies in a way that they are symmetric around zero (i.e., such that $x_e^D = -x_e^R$ and $x_s^D = -x_s^R$). Furthermore, let $\Delta x_e \equiv x_e^R - x_e^D$ and $\Delta x_s \equiv x_s^R - x_s^D$ denote the degree of economic and more polarization between Republicans and Democrats. Using this, we can write the utility difference ΔU as

$$\Delta U \equiv U^R - U^D = [1 - w(y)]\Delta x_e r\,(y - \bar{y}) + w(y)\left[\Delta x_e h(\mu) + \Delta x_s \mu\right]. \tag{6}$$

4   Policy Views and Voting in the Cross-Section

Our model analysis proceeds in three steps. In Section 4, we present predictions about voters' ideal policies and voting behavior that follow purely from the preferences described in the previous section, holding party positions fixed. We test the model predictions using cross-sectional data. Second, in Section 5, we present results on changes in policy views and voting behavior over time that result from assumed, exogenous changes in economic conditions and party positions. Then, in a third step, Section 6 analyzes endogenous party positions, how they change over time, and corresponding implications for voting behavior.

4.1   Economic Policy Views in the Cross-Section

To begin, it is instructive to consider how support for a higher (more conservative) economic policy changes as a function of income:

$$\frac{\partial x_e^*}{\partial y} = \underbrace{[1 - w(y)]r'\,(y - \bar{y})}_{\text{direct income effect, } +} + \underbrace{w'(y)\left[h(\mu) - r\,(y - \bar{y})\right]}_{\text{indirect moral effect, ?}} \tag{7}$$

The first term (direct effect) is positive, as in standard political economy models: from the perspective of material incentives, higher income pushes the voter to prefer a more conservative economic policy, as in Meltzer and Richard (1981). However, the sign of the second term (indirect effect) depends on whether an individual's morally-optimal economic policy, $h(\mu)$, is to the left or to the right of his materially-optimal economic policy, $r\,(y - \bar{y})$. For voters who are sufficiently morally conservative (high μ), both the direct and the indirect income effect are positive. For morally liberal people, on the other hand, the two effects point in opposite directions because the higher weight on moral values leads them to favor left-wing economic policies. Thus, the desired economic policies of rich and poor moral liberals are more similar to each other than those of rich and poor moral conservatives. This gives rise to the concept of the "gauche caviar": people who are sufficiently rich that their moral values make them economically left-wing.

Proposition 1. *The relationship between economic policy views and income is stronger among social conservatives. Equivalently, the relationship between economic and social policy views is stronger among the rich. Formally,*

$$\frac{\partial^2 x_e^*}{\partial y \partial x_s^*} = h'(\mu)w'(y) > 0.$$

The proof is obvious from (7) by noting that $x_s^* = \mu$.

Empirical evidence. To our knowledge, this is a new (and distinctive) prediction of our model that crucially depends on the idea that the moral weight increases in income. For example, canonical models such as Meltzer and Richard (1981) do not make such a prediction. Similarly, two-dimensional models that feature both income and moral values do not make such a prediction as long as the utility weights do not depend on income. We therefore proceed by investigating the empirical patterns in the ANES.[12] We use the economic conservatism index as proxy for x_e^* and the social conservatism index as proxy for $x_s^* = \mu$.[13] Figure 1 summarizes the results. We plot standardized beta coefficients that report by which fraction of a standard deviation economic conservatism changes in response to a one standard deviation increase in log income. We show the results of two regression specifications, one without controls and one that conditions on state and year FE, age, age squared, and gender. Throughout the paper, we reference regression evidence in figure notes and relegate the tables to appendices. Consistent with Proposition 1, we find that the coefficient in a regression of economic conservatism on income is more than twice as large among above- than among below-median social conservatism respondents.

4.2 Voting in the Cross-Section

Recall that the probability of voting R is given by $\Psi(\Delta U) = \Psi(U^R - U^D)$. Differentiating (6) with respect to μ shows that

$$\frac{dPr(R|y,\mu)}{d\mu} = \psi(\Delta U)\frac{d\Delta U}{d\mu} = \psi(\Delta U)w(y)\left(h'(\mu)\Delta x_e + \Delta x_s\right) > 0, \tag{8}$$

which implies that morally-conservative types are more likely to vote Republican.

In contrast, the effect of an increase in income y on the probability of voting R is more complicated:

---

[12]The MFQ dataset does not contain an economic conservatism index.

[13]While, in our model, a voter's x_e^* is a deterministic function of y and μ, we could generalize the model such that, in addition, the preferred policy also depends on an iid random shock. In the interest of simplicity, we refrain from doing so, but a previous version with such a setup is available from the authors upon request.
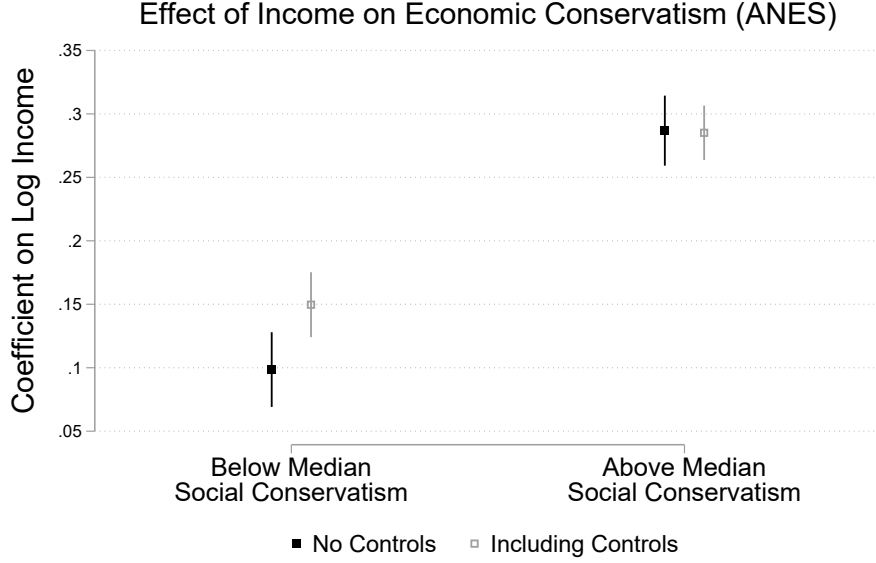
Effect of Income on Economic Conservatism (ANES)

Figure 1: The figure reports coefficients from OLS regressions of the economic conservatism index on log income, split by moral conservatism. 95% confidence intervals reported. In a Wald test of the equality of coefficients between groups, the p-value is $p < 0.0001$. Data from ANES, total $N = 17,576$. Income and the conservatism indices all have mean zero and standard deviation one. Controls include state and year FE, gender, age, and age squared. The corresponding regression results are reported in Table 2 in Appendix C.

$$
\frac{dPr(R|y,\mu)}{dy} = \psi(\Delta U) \left\{ \underbrace{[1 - w(y)]\Delta x_e r'(y - \bar{y})}_{\text{direct income effect, +}} + \underbrace{w'(y)[\Delta x_e(h(\mu) - r(y - \bar{y})) + \Delta x_s \mu]}_{\text{indirect moral effect, ?}} \right\}.
$$
(9)

As in the analysis of economic policy views, there is a direct income effect and an indirect moral effect. The sign of the direct income effect is positive because, from a purely material perspective, richer voters prefer the Republicans' more conservative economic policy (i.e., $r(\cdot)$ is increasing).

The sign of the indirect moral effect again depends on a voter's moral values. It is positive for morally conservative voters, which implies that the probability of voting Republican is unambiguously increasing in income for these voters. However, the indirect effect is negative for morally liberal voters. Intuitively, if an increase in income makes a morally liberal voter rich enough to "afford" voting based on moral values, they can become more likely to vote Democrat.

To make this statement formally in Proposition 2, it is useful to compute the cross-partial with respect to income and moral values; for example, (9) with respect to μ. There are two effects. The first one, corresponding to the intuition discussed above, has a definitive sign. The second one is related to the fact that an increase in μ changes ΔU and thus $\psi(\Delta U)$, the value of the density of ε at ΔU. If the distribution of ε is sufficiently close to a uniform

14

distribution, then the size of the second effect is small, and the first effect dominates.

Proposition 2. *We have $\frac{d^2 Pr(R)}{d\mu dy}$, as given by (13) in the Appendix, positive if $|\psi'(\Delta U(y, \mu))|$ is sufficiently small. In particular, this is the case if ψ is uniform.*

If $\psi' \approx 0$, so that we can neglect the effect of changes in the density, Proposition 2 states that voting is more responsive to variation in income among moral conservatives than among moral liberals. Equivalently, voting is more responsive to variation in moral values among the rich than among the poor.[14] As the proof in Appendix A highlights, this prediction is again driven by the idea of income-dependent utility weights $w(y)$. A two-dimensional setup with constant weights does not generate this prediction

Empirical evidence.   The idea that the voting-income gradient is steeper among social conservatives has loosely been discussed in the literature. For example, McCarty, Poole and Rosenthal (2016) and Gelman et al. (2007) both report that the income gradient is particularly steep among evangelicals and religious people more generally. Gelman et al. (2007) also report that the income gradient is steeper in rural and Southern states than in urban ones, where voters in rural and Southern states are plausibly more morally conservative.

To test our prediction more directly, we regress a binary voting indicator on household income, separately for above- and below-median social conservatism (ANES) and moral conservatism (MFQ). As shown in Figure 2, in both datasets, the effect of a one standard deviation increase in log income on the probability of voting Republican is about 8-10 percentage points larger among social/ moral conservatives.

4.3   Voting Against One's Material Interests

The insight that moral priorities are more relevant for the voting behavior of the rich affords interesting insights into the behavior of "off-diagonal" voters: those (y, μ) types who are either rich-and-morally-liberal or poor-and-morally-conservative. These voters align with the Democrats and the Republicans on one dimension each, such that they are potential swing voters in elections. Moreover, these groups have recently received a large amount of attention in the popular discussion. For instance, in the bestseller "What's the matter with Kansas?," Thomas Frank (2007) explores the phenomenon of relatively poor white working-class voters who have moved from historically being a core constituency of the Democratic party to nowadays often voting for Republican candidates. Frank argues that these voters are moved by "god, guns and gays" (i.e., their preference for the Republican position on social issues) to vote "against their interests," which Frank identifies exclusively with their economic interests. Interestingly, while Frank accuses the rich moral liberals in the Democratic party of being responsible for Democrats losing the heartland working class, he does not discuss that they themselves often vote against their (economic) interests, too. In fact, the existing

---

[14]We discuss the case where ψ is not approximately constant after the proof of Proposition 2 in Appendix A.
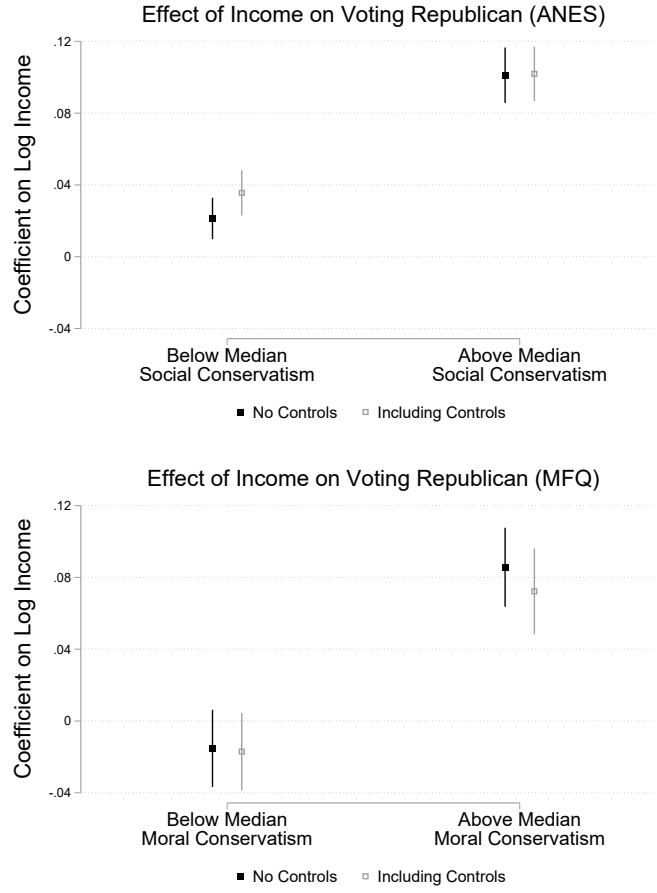
Figure 2: Log income coefficient in a voting regression, by level of moral / social conservatism. The top panel uses data from the ANES ($N = 15,505$) and reports the OLS coefficients of regressions of voting Republican on log income, conditioned on the respondent's social conservatism index. The bottom panel uses data from the MFQ ($N = 3,471$) and reports the OLS coefficients of regressions of voting Republican on log income, conditioned on the respondent's moral values index. In a Wald test of the equality of coefficients between groups, the p-value is $p < 0.0001$ in both datasets. The log income variables are standardized to have mean zero and standard deviation one. The corresponding regression results are reported in Tables 3 and 4 in Appendix C.

literature has rarely compared how much poor moral conservatives actually vote Republican *relative to* how much rich moral liberals vote Democratic.

Proposition 3 compares the voting behavior of a rich moral liberal (RML) and a poor moral conservative (PMC) who are, in a sense, mirror images of each other: they have moral ideal points that are symmetric around zero,[15] and their materially-optimal economic policies are also symmetric around zero. Thus, RML is better off economically with the Republican platform, but morally prefers the Democrats, and vice versa for PMC.

Proposition 3 shows that RML is actually more likely to vote against their material interests than PMC. The simple intuition is that, because of the income effect on the utility weights, it is more important for RML to morally align with the Democrats than it is for

---

[15]Remember that we have normalized party positions in such a way that the average of the Democratic and Republican position, on both economic and social issues, is zero.

PMC to morally align with the Republicans. Thus, the following proposition is again a reflection of the assumption of income-dependent utility weights.

Proposition 3. *Suppose that h is a linear function (i.e., $h(\mu) = \ell\mu$ for some $\ell > 0$), and that the function ψ is symmetric around 0 (i.e., no party has an expected valence advantage).*

Consider two voters, RML and PMC, such that $-\mu_{PMC} = \mu_{RML}$ and $r\left(y_{RML} - \bar{y}\right) = -r\left(y_{PMC} - \bar{y}\right)$. Then $Pr(D|RML) > Pr(R|PMC)$.

Since we cannot observe individual *probabilities* of voting Republican, an empirical test of Proposition 3 requires that we aggregate across several individuals and observe their vote realizations. If, in such an aggregation, every voter in the set of "rich social liberals" is perfectly matched with a mirror "poor social conservative" type as defined in Proposition 3, then the percentage of Democratic votes in the former set should be larger than the percentage of Republican votes in the latter set (assuming sufficiently many voters in each set). Moreover, by continuity, the result holds in such an aggregation even if the two empirical groups are slightly unbalanced (i.e., do not consist of voters that are *exact* mirror images of each other, such as in Proposition 3).

Empirical evidence. Given that this prediction is in stark contrast with many popular (qualitative) narratives, we test it using ANES and MFQ data. Figure 3 plots the fraction of rich social liberals and poor social conservatives who vote "against their material interests." Here, we define "rich" and "poor" as the top/bottom third of household income, and "socially conservative" and "socially liberal" as top/bottom third of the social / moral conservatism indices. Figure 3 shows that rich social liberals are about 35% more likely to vote Democrat than poor social conservatives are to vote Republican. These comparisons are statistically significant in both datasets.[16]

4.4 Within-Party Supporter Heterogeneity

The higher emphasis on moral issues among the rich also has direct implications for within-party constituent heterogeneity in income and moral values. Our main insight is that the set of Democratic voters is more heterogeneous than the set of Republican ones. To show this, we first introduce the concept of an iso-probability curve.

Definition 1. *Iso-probability curves are given by the set of (y, μ) for which the voting probability is constant, i.e., for which ΔU is constant.*

A special case of an iso-probability curve is the level set for which the voting probabilities are given by $1/2$. All voters who are located to the right of this curve are more likely to vote R rather than D, and vice versa.

---

[16]A potential concern is that these results are spurious and driven by an overall higher vote share for Democrats than Republicans in the ANES and MFQ datasets. To correct for such a potential confound, Appendix Figure 8 shows the demeaned vote shares for each group by subtracting the overall Republican and Democratic vote shares. Rich liberals are still significantly more likely to vote against their economic interests relative to poor conservatives.
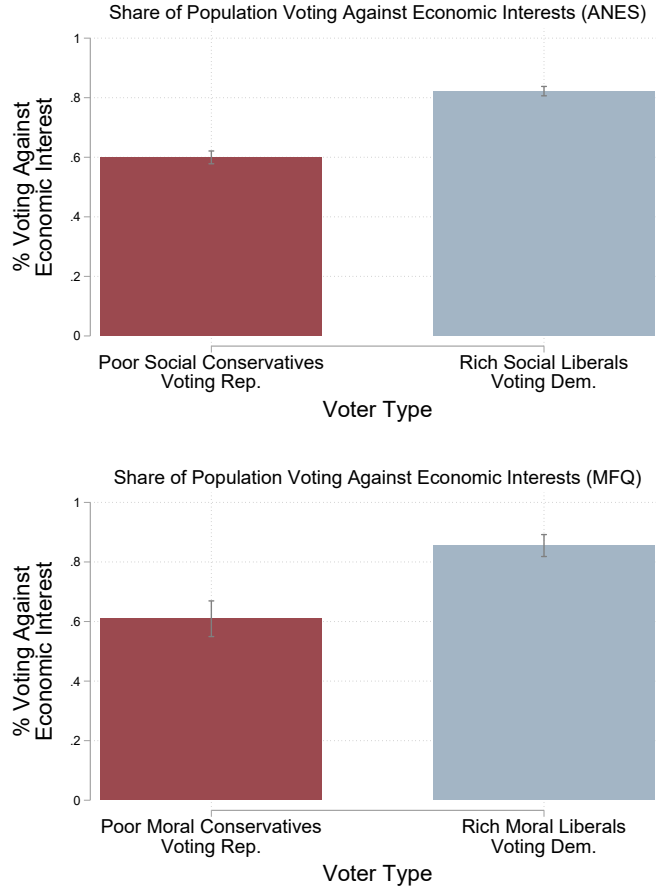
Figure 3: This figure reports the sample mean and 95% confidence interval of the share of poor social conservatives who vote Republican, and rich social liberals who vote Democrat. Here, we define "rich" and "poor" as the top/bottom third of household income, and "socially conservative / liberal" as top/bottom third of the social / moral conservatism indices. The top panel uses ANES data ($N = 4,274$) and the bottom panel MFQ data ($N = 608$). In t-tests of the equality of means across groups, the p-value is $p < 0.001$ in both datasets.

Proposition 4 below analyzes properties of the iso-probability curves. Iso-probability curves are usually downward-sloping. We show that this holds also in our model for voters who are not too liberal, and also applies for all voters if the weight function is of the logistic form (i.e., $w(y) = \frac{c+\rho(y-k)}{1+\rho(y-k)}$, with ρ sufficiently small).[17]

More importantly, if the weight function is logistic, then iso-probability curves are always strictly convex, as illustrated in Figure 4. Convexity establishes a sense in which the set of D voters is more heterogeneous than the set of R voters. Intuitively, the reason for this asymmetry is that rich moral liberals are more willing to vote against their material interests than poor moral conservatives. Thus, in our model, the D party is a "big tent" that attracts both the very poor and some of the very rich.

We note that this prediction on differential within-party heterogeneity is quite particular

---

[17]In principle, the indirect income effect that operates through $w(\cdot)$ could be so strong for moral liberals that it makes some iso-probability curves upward-sloping. As the indirect income effect is proportional to ρ, assuming that ρ is not too large effectively limits the size of this effect.
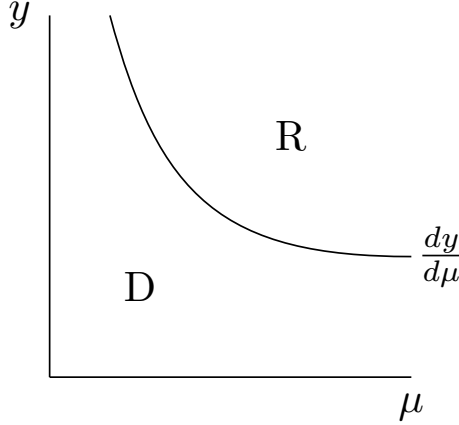
Figure 4: This figure depicts a voting iso-probability curve in (μ, y) space, where larger values of μ and y correspond to more conservative moral views and higher income respectively. Voters with preferences in the upper contour set of the curve vote Republican, and those with preferences in the lower contour set vote Democrat (in expectation).

to our model in that it crucially depends on income-dependent utility weights. If, in contrast, the utility weight of social issues is constant ($\rho = 0$), then the iso-probability curves are linear and downward-sloping. This clarifies that what generates our prediction about differential supporter heterogeneity is indeed morals as luxury goods, rather than a two-dimensional setup as such.

Proposition 4. *1. The slope of the isoprobability curve is given by*

$$
\begin{aligned}
\frac{dy}{d\mu} &= -\frac{w(y)\left[h'(\mu)\Delta x_e + \Delta x_s\right]}{[1-w(y)]\Delta x_e r'(y-\bar{y}) + w'(y)[\Delta x_e(h(\mu) - r(y-\bar{y})) + \Delta x_s \mu]} \\
&= -\frac{\frac{w(y)}{1-w(y)}\left[h'(\mu) + \frac{\Delta x_s}{\Delta x_e}\right]}{r'(y-\bar{y}) + \frac{d}{dy}\left(\frac{w(y)}{1-w(y)}\right)[1-w(y)][h(\mu) - r(y-\bar{y}) + \frac{\Delta x_s}{\Delta x_e}\mu]}
\end{aligned}
\tag{10}
$$

A sufficient conditions for (10) to be negative at (μ, y) for any weight function is that $h(\mu) - r(y - \bar{y}) + \frac{x_s^R - x_s^D}{x_e^R - x_e^D}\mu \geq 0$ (i.e., isoprobability curves are downward-sloping for all individuals who are not too liberal).

An alternative condition for (10) to be negative for all voters is that $w(y) = \frac{c + \rho(y-k)}{1 + \rho(y-k)}$, with ρ sufficiently small.

2. Suppose that $r(\cdot)$ and $h(\cdot)$ are linear functions. If the weight function satisfies $\frac{d}{dy}\left(\frac{w'(y)}{w(y)}\right) \leq 0$ and $w''(y)(1 - w(y)) + 2[w'(y)]^2 \geq 0$ for all y, then all downward-sloping isoprobability curves are convex.

In particular, a linear weight function, and $w(y) = \frac{c + \rho(y-k)}{1 + \rho(y-k)}$ with $\rho > 0$ satisfy these conditions.

Proposition 4 suggests that it is interesting to empirically compare intraparty standard deviations of income and moral values for Democrats and Republicans. For example, when

the type distributions of y and μ are uniform, then the proposition says that the standard deviation of moral values and income should be higher in the set of Democratic voters.[18]

Empirical evidence. To test the novel prediction that the set of Democratic supporters is more heterogeneous, we compute the standard deviation of income and social / moral conservatism within the sets of ANES and MFQ respondents that voted Republican and Democrat, respectively. Figure 5 summarizes the results. All four comparisons point in the hypothesized direction, as Democratic voters are more heterogeneous in terms of their values and income across datasets. Three of these comparisons are statistically highly significant. This evidence is also consistent with the data summarized in Rodden (2019), who argues that the Democratic party represents a more heterogeneous set of districts than the Republicans. Moreover, Bock (2020) finds that Democrats exhibit greater within-party variation in views on immigration than Republicans.

5 Policy Views and Voting over Time

5.1 Policy Views and Ideological Constraint Over Time

We now turn to time series analyses. Because party positions potentially (endogenously) change over time, we begin by analyzing our model's time series predictions for policy views. This has the analytical advantage that all results are a direct consequence of our posited utility function and do not require assumptions on party positions.

In studying changes in policy views over time, we focus on secular change and the consequences of long-term economic growth. To do so, it is convenient to analyze a framework in which voters' relative incomes remain constant. Thus, we assume that the income of all voters is scaled multiplicatively by $\beta > 1$ between an "early" period and a "late" period. Since the income variable in the model, y, is the logarithm of income, this means that y grows by $\ln(\beta)$.

We now investigate what our model implies for the relationship between ideal economic policies, x_e^*, and ideal social policy views, x_s^*, and how it changes with β (over time). The fact that people's social and economic policy views are typically correlated across seemingly-unrelated domains has been labeled "ideological constraint" (e.g. Ansolabehere, Rodden and Snyder, 2008). The following Proposition 5 shows that economic and social policy views become more tightly linked over time. Note that this is true in a framework where the correlation between the fundamental parameters – moral values and income – does not change over time.

---

[18]A uniform distribution of types is sufficient but not necessary. In our model simulations in Section 6, we assume normal distributions, and in these simulations the standard deviation of income and moral values is always higher in the set of Democratic voters when we restrict attention to simulations in which the overall vote share is between 45% and 55% (clearly, Democrats can be less heterogeneous in the counterfactual scenario that their vote share is very small).
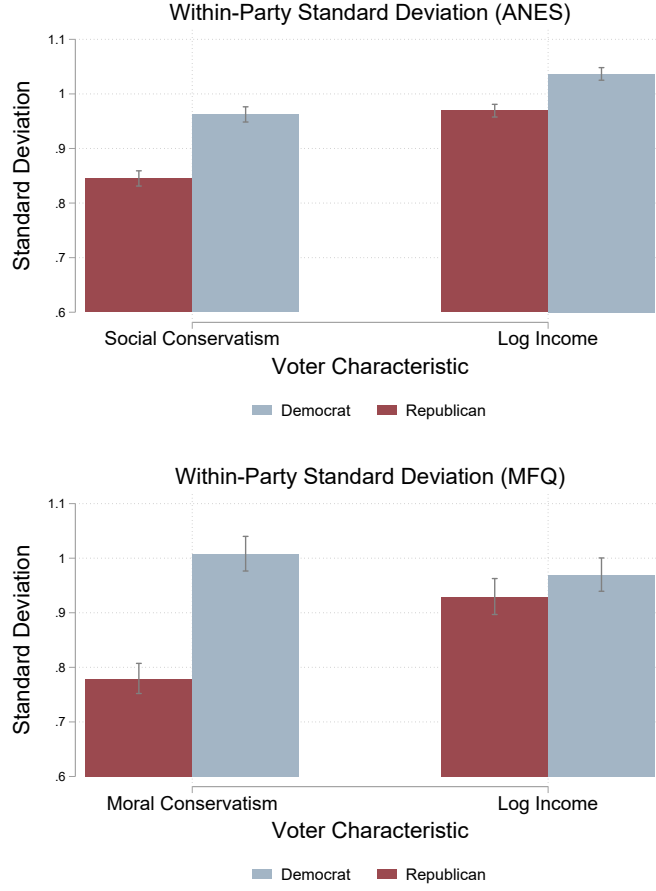
Figure 5: This figure reports the sample standard deviation and 95% confidence interval of social/moral conservatism and log income, split by whether a respondent voted Republican or Democrat. The panel on the top uses ANES data ($N = 13,175$) and the panel on the bottom uses MFQ data ($N = 3,471$). In F-tests of the equality of standard deviations across groups, the p-value is $p < 0.01$ for social conservatism and log income in the ANES, and for moral conservatism in the MFQ. For log income in the MFQ it is $p = 0.08$.

Proposition 5. *Assume that each voter's income increases by a factor of $\beta - 1$ (i.e., y increases to $y + \ln(\beta)$). Then, differentiating eq (5),*

$$\frac{\partial x_e^{*,late}}{\partial \mu} = w(y^{late})h'(\mu) > w(y^{early})h'(\mu) = \frac{\partial x_e^{*,early}}{\partial \mu}. \tag{11}$$

To develop some intuition for Proposition 5, consider a population in which moral values and income are independently distributed. Because voters' ideal economic positions consist of a convex combination of their materially-optimal economic policy and their morally-optimal economic policy, in our model ideal economic and social policies are always positively correlated, even in the early period. What happens as all voters' incomes increase by the same factor? Materially-optimal economic policies remain constant for all voters, as do morally-optimal policies. However, the weight on the latter increases, so that the correlation between social policy views (that are directly determined by moral values) and economic
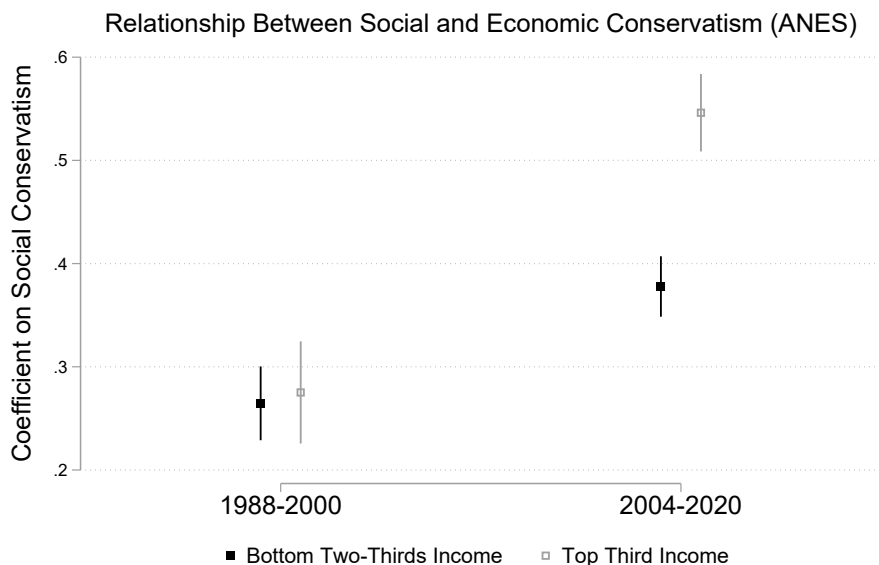
21

Figure 6: The figure reports coefficients from OLS regressions of the economic conservatism index on social conservatism (controlling for log income) split by whether the survey year is pre- or post-2000. 95% confidence intervals reported. In a Wald test of the equality of coefficients between groups, the p-value is $p = 0.6102$ for the 1988-2000 sample, and $p < 0.0001$ for the 2004-2020 sample. Data are from the ANES. Total $N = 17,576$. The conservatism indices both have mean zero and standard deviation one. The corresponding regression results are reported in Table 6 in Appendix C.

policy views increases.

It is also instructive to think about what would happen to ideological constraint of different income groups if their incomes grow differentially. In particular, over the last generation, the income growth among the rich has been much higher than among the poor. To illustrate the formal implications of this, suppose for simplicity that there are only two income groups ("poor" and "rich"), and that only the latter's income increases. Then, by eq. (11), ideological constraint only increases over time among the rich.

Empirical evidence. That economic and social policy views became more strongly correlated in the general public over time is a well-known stylized fact (see, e.g., Kozlowski and Murphy, 2019; Wu, 2020), and Appendix Table 5 documents this fact in the ANES data. More interestingly, Figure 6 tests the more novel model implication that the increase in ideological constraint is concentrated among the rich. We see that this is indeed the case. The increase in the relationship between social and economic conservatism is about twice as large among the top 1/3 of the income distribution than among the bottom 2/3.

5.2 Voting over Time: Realignment

As in the previous subsection, our interest lies in understanding the effects of long-run increases in income, but we now focus on implications for voting. In principle, an increase in aggregate income could have two effects that are relevant for voting behavior. First, for

fixed party positions, voters' preferences could change in a way that affects their voting behavior. Second, party positions could change in response to changes in voter preferences. In Section 6, we will simultaneously study these effects in a general equilibrium analysis through simulations.

However, as a prequel, it is instructive to first study each of these two cases separately analytically. To do this, we first ask how voting behavior changes in response to an increase in incomes, holding party positions fixed. Then, in a second step, we ask how exogenous party polarization on social issues affects voting behavior, holding incomes fixed. This latter thought experiment is relevant because in Section 6 we will see that increasing incomes indeed induce social party polarization.

The bottom line of this analysis is that these two effects both point in the same direction: a realignment of rich-and-morally-liberal voters (such as the "educational elite") from Republicans to Democrats, and a realignment of poor-and-morally-conservative voters (such as the rural working class) from Democrats to Republicans. These are arguably two of the most widely discussed voter realignments that have taken place over the last 50 years (Frank, 2007; Hacker and Pierson, 2020; Gethin et al., 2021).

5.2.1 Increasing Incomes and Realignment

The following Proposition shows that, over time, poor moral conservatives become more likely to vote Republican, and rich moral liberals become more likely to vote for Democrats.

Proposition 6. *Assume that each voter's income increases by a factor of $\beta - 1$ (i.e., a voter's y increases by $\ln(\beta)$), while party positions remain fixed. Then a voter's probability of voting Republican increases if and only if*

$$h(\mu) + \frac{(x_s^R - x_s^D)}{(x_e^R - x_e^D)}\mu - [r(y - \bar{y})] > 0. \tag{12}$$

To interpret this proposition, define a moral conservative (liberal) as a voter for whom $h(\mu) + \frac{(x_s^R - x_s^D)}{(x_e^R - x_e^D)}\mu > (<)0$. Likewise, define a voter as rich (poor) if $r(y - \bar{y}) > (<)0$.

For morally conservative and poor voters, the first term in (12) is positive, and the second (deducted) term is negative, so that the left-hand side of (12) is positive. Thus, poor moral conservatives become more likely to vote Republican. Analogously, for rich moral liberals, both effects go in the opposite direction, so they become more likely to vote for Democrats. These effects are intuitive: because voters weight their moral concerns more highly after everyone's income increases, a realignment based on moral values takes place.

It is instructive to compare these patterns for the "off-diagonal" voters with those for the two "core supporter" groups (i.e., poor moral liberals among the Democrats, and rich moral conservatives among Republicans). For both of these groups, the effect of a uniform income increase on voting is ambiguous, as the two partial effects go in opposite directions. This may shed light on why the popular discussion is oftentimes particularly focused on the two

"off diagonal" rather than the "core supporter" groups: for the latter, an increase in income does not imply systematic changes in voting behavior.

5.2.2 Party Polarization and Realignment

It is useful to build intuition for the model mechanics by considering how voting behavior would change *if* one knew that party positions changed in particular ways.[19] To analyze the effect of party polarization on the different types of voters, we consider a symmetric change in party positions that leaves the midpoint between the two parties' positions unchanged; that is, when Democrats move to the left (on economic and/ or social issues), then Republicans move to the right by the same amount.[20] As before, and without loss of generality, we normalize the mid-point of party positions to zero.

The first part of Proposition 7 deals with a balanced polarization on both economic and social issues and shows that this intensifies the probability that any given type votes for his policy-preferred party ("sorting"). In contrast, the second part deals with polarization on social issues, which leads to a systematic realignment of which voter types prefer which party, on average. Any mixed change in polarization can be decomposed into the two cases covered by the proposition.

Proposition 7. *1. If both economic and social party platforms polarize symmetrically over time, then voter sorting increases. Formally, if*

$$\frac{x_{s,1}^R - x_{s,1}^D}{x_{s,0}^R - x_{s,0}^D} = \frac{x_{e,1}^R - x_{e,1}^D}{x_{e,0}^R - x_{e,0}^D} = k > 1,$$

then $\Psi(U_1^R - U_1^D) \gtreqqless \Psi(U_0^R - U_0^D)$ if and only if $\Psi(U_0^R - U_0^D) \gtreqqless 1/2$.

2. If social party platforms polarize symmetrically over time and economic platforms are unchanged, a realignment based on moral values takes place, meaning that moral conservatives (liberals) become more (less) likely to vote R. Formally, if

$$\frac{x_{s,1}^R - x_{s,1}^D}{x_{s,0}^R - x_{s,0}^D} > 1 \quad \wedge \quad x_{e,1}^R = x_{e,0}^R \quad \wedge \quad x_{e,1}^D = x_{e,0}^D,$$

then $\Psi(U_1^R - U_1^D) \gtreqqless \Psi(U_0^R - U_0^D)$ if and only if $\mu \gtreqqless 0$.

Both claims follow from (6) in a straightforward way.[21] A proportional symmetric polarization in both policy dimensions will simply multiply a voter's ideological utility difference.

If this utility difference was positive (i.e., the likelihood that this voter type would vote for R was larger than $1/2$ in the early period), it will grow, so that the valence shock becomes less important. Thus, the probability of voting R will increase, and vice versa. This clarifies how party polarization induces more sorting of voters into parties based on their ideological utility. This result is helpful because it shows that the increasing amount of voter sorting that is often discussed in political science (Levendusky, 2009; Mason, 2018) can rationally occur even without any changes in voters' preferences, purely as a result of party polarization.

More important for our purposes, Part 2 of Proposition 7 implies a form of realignment based on moral values. When only social polarization increases, then the utility difference changes by $w(y)\mu\Delta x_s$, which is positive if and only if $\mu > 0$. As a result, moral values become "more important" for voting: any fixed voter type becomes more likely to vote R if and only if he is morally conservative. Note that this occurs without any voter becoming more extreme (in terms of their moral preferences), or putting more weight on social issues (in terms of the weight w). This effect is of particular relevance for those voters who are close to being indifferent, such as "off-diagonal" voters that are either poor-and-morally-conservative or rich-and-morally-liberal.

Part 2 of Proposition 7 arguably informs the discussion about the realignment of socially conservative working class voters towards the Republican party – the focus, for example, of "What's the matter with Kansas?". An oft-discussed puzzle is that many of these voters have swung Republican *even though* their relative economic position deteriorated over time, which should push them towards the Democrats. In our model, a decrease in relative income implies a more left-wing optimal economic policy. However, if the decrease in relative income coincides with increased social party polarization, these voters experience the offsetting effect identified in Proposition 7.2 and thus may still end up being more likely to vote for the Republican party. This is stated formally in the following Corollary.

Corollary 1. *Consider a voter of a fixed type* μ_j, y_j, *and suppose that there is symmetric social polarization, as well as growth in average income from* $\bar{y}_{old}$ *to* $\bar{y}_{new}$ *(voter* j's *income remains fixed). Then, the change in voter* j's *policy utility difference is given by*

$$\Delta(U^R - U^D) = [1 - w(y_j)](x_e^R - x_e^D)\left[r\left(y_j - \bar{y}_{new}\right) - r\left(y_j - \bar{y}_{old}\right)\right]$$
$$+ w(y_j)\mu\left[x_{s,new}^R - x_{s,old}^R + x_{s,old}^D - x_{s,new}^D\right]$$

Note that the first term is negative (as r is increasing, and $\bar{y}_{\text{new}} > \bar{y}_{\text{old}}$), while the second term is positive. Thus, for example, the overall effect is positive if the voter is morally very conservative (μ is large) and/ or if social polarization increases substantially. This illustrates a logic that will also be at play in our general equilibrium analysis: if social polarization increases, then morally conservative voters can become more likely to

being more likely to vote $D/$ R, respectively. Second, after the first step, a proportional increase in polarization which increases every type's probability of voting for their ideologically-preferred party. In such a mixed case, (i) realignment based on moral values takes place; and (ii) core supporters become more likely to vote for their previously preferred parties, which implies stronger voter sorting.

vote Republican *even when* their own relative income declines and their materially-optimal economic policy becomes more left-wing.

Finally, the realignment described in Propositions 6 and 7 will also affect the degree of moral polarization in the electorate, say, as measured by the distance between the moral positions of the average Republican and Democratic voters. In a realignment where social conservatives become more likely to vote Republican, and social liberals more likely to vote Democratic, the average Republican voter becomes more socially conservative, and the average Democratic voter becomes more socially liberal. Empirically, it is indeed true that the average Republican and the average Democratic voters have polarized. This is sometimes casually attributed to the distribution of moral preferences having become more extreme. In contrast, in our model – and in line with a large body of work in political science (Fiorina et al., 2006; Fiorina and Abrams, 2008; Levendusky, 2009) – this pattern is only driven by increased voter sorting based on moral values. We return to this issue in Section 6.

6 Endogenous Party Positions

6.1 Overview

In the previous section, we derived results directly from the voter utility function, taking party positions as exogenous. In this section, we endogenize party positions and allow them to respond to changes in voter income over time. Unlike all previous analyses, this requires us to model how party positions are determined.

As in traditional political economy models, an equilibrium in our model is a constellation in which a set of agents chooses which of the two parties to vote for, given that their expectations about party positions are rational. We deviate from traditional models, however, in assuming that a party's position aggregates their supporters' political preferences, i.e., the party's position is the one that maximizes the aggregate utility of the party's "supporters."

The reduced-form assumption that parties aggregate the preferences of those agents who vote for them lends itself to three interpretations that differ in what the set of agents in our model is, i.e., whose utility the parties take into account when setting their positions. A first interpretation is that the set of agents consists of all voters. Then, loosely speaking, our assumption says that each party sets policies to make those people happy who affiliate with that party. For example, party leaders could have incentives to make their own affiliates happy because they vote in the party's primaries.

A second interpretation is that the set of agents in our model is given by rich donors. Then, loosely speaking, our assumption says that each party sets social and economic policy positions to make those people happy who donate large amounts.

Finally, a third interpretation is that the agents in our model consist of the "representative" voters in each of multiple Congressional districts. This interpretation follows the spirit of the model in Polborn and Snyder (2017). They view Congressional elections as a contest in

which voters in different districts choose between local candidates. Once elected, these local candidates all have an incentive to influence their party's position in the direction preferred by their local constituents (e.g., to foster their chances in the next general election). Under this interpretation, each legislator would like their party to set policies that maximize the utility of their own district's representative (or median) voter. Then, actual party positions reflect a convex combination of the interests of each elected legislator of a party: party positions maximize the average policy utility of the voters in those districts that a party actually represents.[22] This interpretation acknowledges that parties are not cohesive organizations under the full control of a single leader, but rather that they are influenced by the demands of elected representatives who face heterogeneous reelection incentives themselves.

In summary, our model is flexible in terms of what constitutes the set of agents that influence the parties' positions. Our model is, however, very different from one where some party leadership can choose, before the election, a position and commit to implementing it if elected to a majority. It is well known that such models, with a multidimensional policy space, generically do not have a pure strategy equilibrium, and if they do have an equilibrium, it would involve complete convergence of party positions (Plott, 1967; McKelvey, 1976; Krasa and Polborn, 2012a). In contrast, our model generates platform divergence. We view this as a feature rather than a bug, especially in light of the large body of evidence in political science that documents that candidate positions do not converge even within a given district (e.g., Fiorina, 1999; Ansolabehere, Snyder and Stewart III, 2001; Burden, 2004)

While we assume that party positions are only determined by their supporters' preferences, we could assume more generally that parties maximize some convex combination of the average utility of their supporters, and the utility of the average voter in the general election (in order to capture, in addition, the parties' incentives to win the general election). This would lead to quantitatively less polarization between the parties than in our setup, but the qualitative comparative static results (i.e., the effect of increased income on social and economic polarization) would be unchanged.

6.2 Formal Equilibrium Definition

An equilibrium is a partition of voters and a set of party positions such that (i) voters maximize utility by choosing the optimal party given expected party positions, (ii) parties choose policy positions to maximize the weighted average utility of the voters which the party represents, and (iii) voters' expectations about party positions are correct.

Definition 2. *Let $\phi_t(\mu, y)$ denote the density of the voter type distribution at time t. Let S_i denote the set of voters supporting party i. A political equilibrium is defined by a tuple $(S_D, S_R, x_e^D, x_e^R, x_s^D, x_s^R)$ such that*

---

[22]Voters in each district, in turn, care about national party positions and a valence shock from the local candidates' idiosyncratic qualities.

1. *The density of S_D is given by*

$$\xi_D(\mu, y) = \frac{\phi_t(\mu, y)\Psi(\Delta(\mu, y))}{\int_{(\mu,y)} \Psi(\Delta(\mu, y))d\Phi_t(\mu, y)},$$

where

$$\Delta(\mu, y) = \frac{(1 - w(y))}{2} \left(\left[x_e^R - r(y - \bar{y}) \right]^2 - \left[x_e^D - r(y - \bar{y}) - \eta \right]^2 \right) +$$
$$\frac{w(y)}{2} \left((x_e^R - h(\mu))^2 - (x_e^D - h(\mu))^2 + \alpha(x_s^R - \mu)^2 - \alpha(x_s^D - \mu)^2 \right)$$

is the utility difference of a type (μ, y) between the Democratic and Republican policy positions. Let Ξ_D denote the corresponding cdf.

2. *Similarly, the density of S_R is given by*

$$\xi_R(\mu, y) = \frac{\phi_t(\mu, y)[1 - \Psi(\Delta(\mu, y))]}{\int_{(\mu,y)}[1 - \Psi(\Delta(\mu, y))]d\Phi_t(\mu, y)}$$

3. *The parties' policies maximize the average utility of their respective supporters. For instance, for the Democrats, we have*

$$(x_e^D, x_s^D) \in \arg\max \int_{(\mu,y)} U(x_e, x_s; \mu, y)d\Xi_D(\mu, y)$$

Like in Polborn and Snyder (2017), there is always a trivial equilibrium with policy convergence, in which both parties have positions at the average (μ, y) so that no voter perceives a difference between the parties; in this case, only the random shocks determine whether a voter ends up as Democrat or Republican, and consequently the type distributions are identical in both parties. However, by the same reasoning as in Polborn and Snyder (2017), this equilibrium is unstable if the shocks are sufficiently small, i.e., once the party positions are slightly differentiated, adjustment processes lead to an increase in the divergence between parties. We will therefore focus on the equilibrium in which the party positions are differentiated.

6.3 Analytical Approach

Our objective in this section is to understand how an increase in incomes over time (captured by β) affects the structure of the political equilibrium, in particular, party positions and voting behavior. Because voter iso-probability curves in (y, μ) space are nonlinear (see Section 4.4), it is impossible to derive the general equilibrium of our model in closed form.

Therefore, we proceed numerically, through simulations. However, we do this for a wide range of parameter values so as to analyze which results are robust and which depend on specific parameters. It turns out that the model essentially always delivers the same directional predictions.

We do not view our approach as a full structural calibration of our model, but rather as a proof-of-concept that our model plausibly generates three widely discussed comparative statics phenomena: an increase in social party polarization, increased moral polarization in the electorate, and the familiar realignment patterns emphasized in Section 5.2.2.

Parameterization of simulations. We model log income as $y \sim N(11 + \beta, 0.75)$, which is roughly calibrated to the U.S. household income distribution in 1988 (which is when the ANES data that we use start). To assess the effect of increases in aggregate income, we simulate the model separately for an "early period" with $\beta = 0$ and a "later period" with $\beta = 0.23$. This income increase corresponds to the approximate rise in median log household income in the U.S. between 1988 and 2020. Mirroring the patterns in the data, we assume that the cross-sectional distribution of moral values follows a normal distribution, and is independent of y. We normalize mean and standard deviation such that $\mu \sim N(0, 1)$.

The main degrees of freedom in the model are the parameters of the weight function (c, ρ and k) as well as the preference parameters that determine voters' materially- and morally-optimal economic policies. We assume that the latter two functions are linear, such that $h(\mu) = h_0 \cdot \mu$ and $r(y - \bar{y}) = r_0 \cdot (y - \bar{y})$. We set $k = 8.99$, which implies that the poorest voters in the early period place the minimum weight of c (which may be zero) on philosophical utility.

We then simulate the model for *a large range of economically reasonable combinations* of parameters c, ρ, r_0 and h_0. When we say "economically reasonable," we refer to the fact that some parameters are unbounded, in principle, such as $r_0, h_0 \in [0, \infty)$. However, as r_0 or h_0 go to infinity, the model collapses to one dimension. Similarly, $c = 1$ implies that even the poorest voters only care about moral issues, which eliminates the role of relative income for voting. Since our objective is to analyze how tradeoffs between economic incentives and moral preferences affect political outcomes, these parameter values are not particularly useful. We, hence, simulate over the grid: $\rho \in [0.1, 0.2, \ldots, 0.9] \times h_0 \in [1, 2, \ldots, 9] \times r_0 \in [1, 2, \ldots, 9] \times c \in (0, 0.1, \ldots, 0.4)$ for a total of 3,645 simulations. The net idiosyncratic (valence) shock is modeled as $\varepsilon \sim N(0, 1)$.

These combinations of parameters imply a broad range of relative weights and distributions of ideal policy points. To illustrate this, consider the following:

1. The parameters imply a large range of implied weight functions, see Appendix Figure 10. For instance, for $c = 0$ and $\rho = 0.1$, the moral weight $w(y)$ ranges between 0% for the poorest and 30% for the richest voters, while for $c = 0.4$ and $\rho = 0.9$ it ranges between 40% and 90%.

2. The parameters allow for large variation in how much moral values, μ, and relative income, $y - \bar{y}$, affect a voter's ideal economic policy, respectively, as Appendix Figure 11 illustrates. For instance, for $r_0 = 9$ and $h_0 = 1$ (relative income matters much more for economic conservatism than morals), the distribution of ideal economic policies (x_e^*) of top third income voters does not at all overlap with that of bottom third income

voters. For $r_0 = 1$ and $h_0 = 9$ (morals matter much more), on the other hand, the distributions of top and bottom third income voters have a large common support and are visually very similar.

3. Consider a voter with median income y and median moral values μ. Now increase the voter's moral conservatism by one standard deviation – how much would we have to decrease the voter's income to keep his ideal economic policy unaffected? For $r_0 = h_0 = 1$, the required compensation is a 1.6 SD reduction in income. For $r_0 = 9$ and $h_0 = 1$, a 0.06 SD decrease in income is required. For $r_0 = 1$, $h_0 = 9$, a 2.1 SD decrease in income is required. Thus, again, our parameters allow for a wide range of scenarios.

The simulations are conducted with $N = 90,000$ voters, drawn from the joint distribution $\Phi_{\mu,y}$.[23] We set the initial party positions to be $x_s^D = x_e^D = -0.5$ and $x_s^R = x_e^R = 0.5$, and calculate the voters' probability of voting for each party given the initial positions.[24] Then, we calculate updated Republican positions as those which maximize the weighted average utility of voters, where the weights are determined by the probability of voting Republican (and similarly for the Democrats). These new party positions induce new voting probabilities, which induce new positions, and so forth. We iterate this process until the party positions converge to a fixed point.[25]

6.4 Party Polarization on Social Issues Over Time

We again intersperse simulation "predictions" with a discussion of the available evidence.

Simulation prediction 1. *Social party polarization increases over time (as β increases).*

Discussion of simulation results. Table 1 summarizes the results of our 3,645 simulations regarding party polarization. We are interested in whether the distance between equilibrium party positions increases from the early period ($\beta = 0$) to the late period ($\beta = 0.23$). We see that social party polarization strictly increases in essentially all simulations. To illustrate the intuition behind this social polarization result, consider a scenario in which the parties are initially somewhat differentiated in both dimensions. As all voters get richer, they place a higher weight on philosophical utility, which makes their indifference curves in (y, μ) space steeper. Thus, some rich-and-morally-very-liberal voters will switch to the

---

[23]To draw from the joint distribution, we simulate on grid spaced out such that the empirical CDF matches the theoretical CDF. We simulate voter income and moral conservatism at 300 equally-spaced percentiles (from the 1st to the 99th percentile) from the analytical distributions of y and μ, then create all $300 \times 300 = 90,000$ combinations for voter primitives (y, μ).

[24]In Appendix D, we show that the simulations are robust to a wide range of initial values, subject to the constraint that the parties do not offer the same policies.

[25]We iterate each simulation 50 times, and define convergence as cases in which the average difference in party positions between subsequent iterations is less than 1%. The simulation converges in 99.8% of cases. The reported results below include the remaining 0.2% of cases, but the results are virtually identical if we exclude these cases.

Table 1: Simulation results regarding effects of increase in β

	Fraction of simulations in which variable in col. (1)		
	Increases	Constant	Decreases
Party polarization on social issues	99.8	0.0	0.2
Party polarization on economic issues	55.0	0.0	45.0
Moral polarization in electorate	>99.9	0	<0.1
Income polarization in electorate	<0.1	0	>99.9
Rich moral liberals voting Democratic	80.0	20.0	0.0
Poor moral conservatives voting Republican	91.0	1.0	8.0

Notes. This table presents the fraction of baseline simulations in which the variable in column 1 increases, remains constant, or decreases over time (as β increases). Social (economic) party polarization increases if the distance between party positions on social (economic) issues increases. Moral (income) polarization in the electorate increases if the difference in average moral conservatism μ (income y) between voters who vote Republican and those who vote Democrat increases. To compute this, we calculate weighted averages of moral conservatism and income in the population, where the weights are given by the probability of voting Republican and Democrat, and take the difference between the two. Rich moral liberals are defined as top 1/3 income and bottom 2/5 moral conservatism. Poor moral conservatives are defined analogously.

Democrats and some poor-and-morally-very-conservative ones to the Republicans. As parties aggregate the utilities of their supporters, and both parties' supporter sets have become more socially extreme, this induces both Democrats and Republicans to move their social policy platform to the left and right, respectively. This position movement then means that moral preferences now become even more important for vote choice because the party differences on social issues have become more salient. Ultimately, this process equilibrates at a higher level of social polarization. Our simulations show that this intuition holds true for a wide range of parameter configurations.[26]

On economic issues, the picture is considerably more mixed: in only 55% of our simulations does economic party polarization increase. Appendix Figure 12 dissects how increases or decreases in economic polarization depend on parameter values.[27]

---

[26]As discussed above in Section 6.1, our model could also assume that parties maximize some convex combination of the utility of their average supporter and of the average voter in the election. This changes the quantitative implications of the model, but not the qualitative results. For example, if weight 80% is placed on a party's voters and 20% on the average voter, our model still delivers increasing social party polarization in esentially all of our simulations.

[27]The results are intuitive: economic polarization generally increases over time when the quantity $r - h$ is not too large, especially when it is negative. To see why, note that in our model, economic party polarization increases in the variance of the distribution of voters' ideal economic policy positions. This variance, in turn, is jointly determined by voters' materially-optimal economic policies (as partly determined by r) and their morally-optimal economic policies (as partly determined by h). Thus, party polarization will tend to increase (decrease) as β increases if the increased utility weight on moral issues implies that a higher weight is placed on that utility component that has

Empirical evidence. The simulation prediction of increased social party polarization is supported by a wealth of data on elite polarization in domains such as immigration, affirmative action and minorities, abortion, religion and morality. For instance, Moskowitz et al. (2018) document using candidate surveys that increasing elite polarization in the U.S. Congress is largely driven by social and cultural issues. Gentzkow et al. (2019) and Enke (2020) provide evidence for increasing elite polarization on social and moral issues using text analyses of Congressional speeches.[28]

While the available evidence strongly suggests that increasing elite polarization is largely driven by social and moral issues, the evidence is a bit mixed regarding polarization on economic issues. For instance, while Moskowitz et al.'s data on candidate surveys show no evidence of increased economic polarization (e.g., taxes, welfare, health insurance, and employment), Gentzkow et al.'s text analyses reveal some evidence for increased elite polarization also on economic issues. Overall, it is probably fair to summarize the empirical literature as saying that economic party polarization is non-decreasing, and possibly increasing.

6.5   Sorting and Polarization Based on Moral Values

Simulation prediction 2. *Over time (as β increases):*

1. *The strength of the relationship between moral values and voting increases.*

2. *The supporters of both parties exhibit "moral polarization": the distributions of moral values of Republican and Democratic voters diverge.*

Discussion of simulation results. Regarding the first part of the prediction, we find that, in every single one of our simulations, the OLS coefficient in a regression of voting R on moral values μ strictly increases as β increases. Moreover, we also find that the raw correlation between these quantities always increases. Thus, the slope of the relationship between these variables becomes larger and the fit becomes tighter. Appendix Figure 13 illustrates by showing the distribution of the changes in regression / correlation coefficients across simulation parameterizations.

Intuitively, as anticipated in the analytical derivations in Section 5.2, two interrelated model mechanisms that point in the same direction produce this result. First, as people get richer, they place a higher utility weight on moral considerations. This directly implies that – holding party positions fixed – moral values become more strongly related to voting (Proposition 6). Second, because social polarization between parties increases, it becomes more salient for the voters' decision in the later period, which again increases the slope between voting and moral values (Proposition 7).

---

higher variance in the population. This, in turn, is determined by whether r is large relative to h.

[28]For example, Figure 6 in Gentzkow et al. (2019) shows that polarization increased more strongly in domains such as immigration, religion, or minorities than in economic domains such as money, education, health or labor.

It is instructive to compare this pattern with the strength of the relationship between income and voting. As illustrated in Appendix Figure 14, the OLS coefficient in a regression of voting on log income y strictly decreases as β increases in the vast majority of simulations (95%). The intuition is twofold. First, social party polarization generally increases. Second, voters now place a higher utility weight on their moral concerns relative to their material ones. Both of these factors attenuate the relationship between voting and income.[29]

Table 1 summarizes the second part of the Simulation prediction by focusing on polarization in the electorate: the idea that the average Republican and the average Democratic voter diverge over time. We see that in all of our simulations, moral polarization in the electorate increases. Remember that the underlying distribution of values is constant across time, so this only reflects increased values-based sorting of voters into the parties.

In contrast, income-based polarization in the electorate decreases in all of our simulations. Intuitively, because voting is now based on moral values to a larger degree, the income distributions of Republican and Democratic voters become more similar to each other.

Empirical evidence. A number of contributions have pointed to empirical regularities that are in line with these predictions. First, regarding the strength of the relationship between voting and values (income), it is widely known that in voting regressions the coefficient of moral views has increased over time, while that of income has generally decreased over time. Appendix Figure 9 visualizes these patterns for the ANES data.[30]

Second, regarding increased moral polarization (diverging distributions of voter types), a large body of work in political science has documented that increased polarization in the general public indeed reflects increased sorting, rather a change of the distribution of issue positions (e.g., Fiorina et al., 2006; Fiorina and Abrams, 2008; Levendusky, 2009). We view our simulation results as squarely in line with this well-documented empirical regularity.

6.6 Realignment of "Off-Diagonal" Voters

Simulation prediction 3. *Over time (as β increases), rich moral liberals become more likely to vote Democrat, while poor moral conservatives become more likely to vote Republican.*

Discussion of simulation results. This prediction is a corollary of the patterns described in the previous section. If the strength of the relationship between voting and values (income) increases (decreases), then a values-based realignment will take place. In particular, "off-diagonal" voters who agree with each party on only one dimension, and who voted

---

[29] The small minority of cases in which income becomes more strongly predictive of voting as β increases are simulations in which r_0 is very large relative to h_0, generally when $r_0 - h_0 \geq 8$. See Appendix Figure 15. This is (i) because economic party polarization may strongly increase in those cases, as discussed above and (ii) because high r_0 means that when the parties are strongly polarized economically the voter will vote his income rather than his values.

[30] Ansolabehere et al. (2006) already point to an increasing importance of social issues for voting in the ANES data over time.
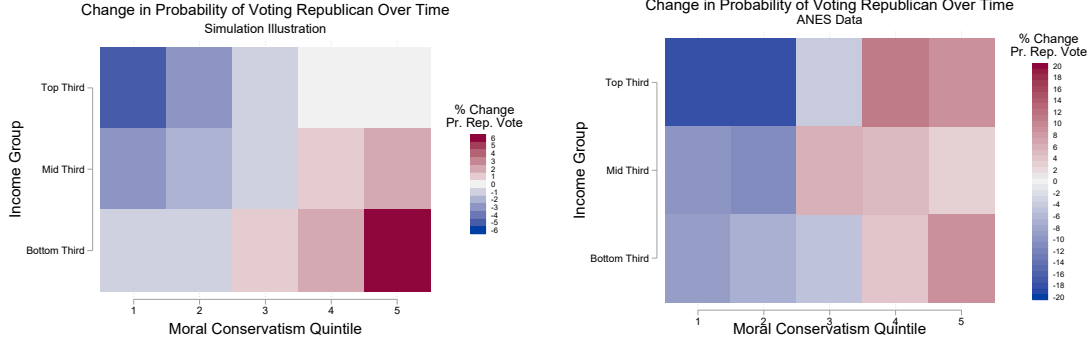
Figure 7: Percentage point change in Republican vote share over time as a function of fundamentals (y, μ). The left panel shows the result of a simulation (with $c = 0$, $\rho = 0.2$, $h_0 = 1$, and $r_0 = 1$). The right panel shows the change in Republican vote share in the ANES, from time period 1988-2000 to 2004-2020.

in line with their material incentives in the early period, become more likely to vote in line with their moral concerns in the later period. To visually illustrate these results using a specific set of parameter values ($c = 0$, $\rho = 0.2$, $h_0 = 1$, and $r_0 = 1$), the left panel of Figure 7 shows the change in voting behavior as a function of fundamentals (y, μ) from the earlier to the later period. The figure's main takeaway is that morally conservative (liberal) voters swing Republican (Democrat), in particular the "off-diagonal" voters in the bottom right and top left corners.

To document the generality with which our model simulations produce the realignment of rich moral liberals and poor social conservatives, Table 1 reports the share of simulations in which rich liberals (poor conservatives) become more likely to vote Democrat (Republican). It shows that rich moral liberals swing weakly left in all of the simulations, and poor moral conservatives swing weakly right in 91% of the simulations. The 9% of cases in which poor moral conservatives swing left are largely those cases discussed in footnote 29 in which $r - h$ is very large, such that almost exclusively relative income matters for a voter's ideal economic policy. See Appendix Figure 16.

Empirical evidence. Krasa and Polborn (2014a) use the ANES to explicitly analyze which types of voters "switched" between the 1970s and the 2000s. They find that economically-conservative-and-socially-liberal voters tended to switch from voting Republican to Democrat. Analogously, economically-liberal-and-socially-conservative voters tended to switch from voting Democrat to voting Republican. Relatedly, Gingrich and Häusermann (2015) and Gethin et al. (2021) provide evidence that the educational elite (which researchers often think of as relatively rich and morally liberal) mostly voted Republican in the 1950s-60s, yet increasingly voted Democrat over time, such that today high educational attainment is positively correlated with voting Democrat. The right panel of Figure 7 illustrates this well-known empirical pattern using ANES data. In the data, moral conservatism becomes substantially more predictive of voting over time, such that poor moral conser-

vatives swing Republican and rich moral liberals Democrat. The figure also serves as an important reminder that – in contrast to some popular narratives – it is not generically true that poor voters have trended Republican. Instead, only those poor voters who are socially conservative have done so – which is a pattern that our simulations also generate, see the left panel of Figure 7.

6.7   Income Inequality

In all simulations reported thus far, we increase incomes uniformly for all voters by adding a constant (in log space), which leaves inequality unchanged over time. However, an important feature of the American political landscape has been growing income inequality, or, more precisely, increasing incomes at the top and largely stagnating incomes at the bottom. Such inequality partly counteracts the mechanics in our simulations above because higher inequality means that people's materially-optimal economic policies become more dispersed. As a result, relative income should matter more for voting, which partly counteracts the mechanics highlighted in the previous subsections.

At the same time, as illustrated by Corollary 1 in Section 5.2.2, even with increasing inequality our model can generate increasing social party polarization and values-based realignment. There are two countervailing effects. First, because many voters get richer, they place a higher weight on their moral values, which induces social party polarization. This, in turn, leads to the realignment of poor moral conservatives and rich moral liberals. Second, because of the increase in inequality, the voters' economic conflicts increase, which, ceteris paribus, decreases social polarization and therefore reduces values-based realignment.

To examine this in the simulations, we scale the log income of all voters multiplicatively by $\kappa = 1.015$; that is, $y_{new} = \kappa y_{old}$. In contrast to our additive scaling above, this multiplicative scaling increases inequality because it implies that the logarithm of income increases more for the rich than the poor, so that the ratio of (absolute) income of the rich to income of the poor increases. We choose κ such that our setup roughly matches the change in income among the 25th percentile, median, 75th percentile, 90th percentile, and 95th percentile of households going from 1988 to 2020.

Appendix D discusses the results of these simulations in great detail. We find that in 99% of all simulations – conducted with the same parameters as in the simulations above – social party polarization increases. Moreover, we again find that the simulations generally produce strong values-based realignment. Thus, even as inequality increases, our model can generate that poor moral conservatives swing Republican.

Intuitively, the logic is that when the rich get disproportionately richer, they place a higher weight on moral considerations, which induces some rich moral liberals to swing Democratic. This induces the parties to polarize on social issues, which, in turn, "pushes" some poor moral conservatives to vote Republican. This realignment then produces further party polarization on social issues because the two voter bases have become more extreme. Thus, in our simulations, poor moral conservatives are both "pushed" to vote Republican

by social party polarization that is primarily induced by the rich, but they also "push" the parties to more extreme positions themselves.

This logic highlights that our mechanism of morals as luxury goods can account for empirically observed polarization and realignment patterns, even if income increases are concentrated at the top. Indeed, the emphasis our model places on the rich inducing social party polarization that then propagates to the poor is consistent with a large body of work in sociology that argues that rural white working class voters feel that the Democratic party "abandoned" them to cater to the moral preferences of coastal elites (e.g., Sherman, 2009; Cramer, 2016).

7   Discussion and Related Literature

Voter utility functions.   The fundamental innovation in our paper is the utility function in which voters care about both economic and moral issues, and the weight of moral issues increases in income. In contrast, most of the formal literature on electoral competition adopts a one-dimensional spatial framework. There are two major disadvantages with this approach for the questions that are in our focus. For one, the assumption that economic incentives overwhelmingly determine how people vote is empirically counterfactual, especially in recent elections (see the evidence discussed above). Second, assuming a one-dimensional policy space precludes meaningful analyses of the realignment patterns observed in practice, such as that poor-and-morally-conservative voters swung Republican, while rich-and-morally-liberal voters swung Democratic. The reason is that, in a one-dimensional framework, movement can occur only in one direction (e.g., if Democrats have a good election, some voter types who would "normally" be expected to vote for Republicans now vote for Democrats, but there are no voter types who travel in the opposite direction).

Polarization.   Our paper is related to a stream of recent theoretical and empirical work on polarization. See Moskowitz et al. (2018) for an excellent review. Most closely related to our work are papers that highlight the role of social and moral issues for realignment and polarization. Shayo (2009); Besley and Persson (2019); Grossman and Helpman (2021) and Rodrik, Ash and Mukand (2021) construct various models that view political conflict through the lens of social identity considerations. From this literature, most closely related to our focus is the model of Bonomi et al. (2021), in which voters identify either with their cultural values or their income group depending on which dimension maximizes contrast with the respective out-group. While their paper focuses on some of the same empirical regularities as ours, there are important differences. First, our model generates various new testable predictions that crucially depend on the assumption of income-dependent utility weights. Second, their model is one of standard probabilistic competition in which politicians maximize the probability of winning. As a result, their model does not produce increased social party polarization. Third, their paper is complementary to ours in that it focuses

on understanding how identity generates belief distortions, whereas in our model beliefs are always rational.[31]

The dynamic part of our paper also relates to recent work by Callander and Carbajal (2022) on the dynamics of polarization. They develop a one-dimensional behavioral model in which voting for a party moves a voter's ideal point in the direction of the party's platform in that election, and policy-motivated parties strategically choose their platforms. In contrast, in our model, voters' utility functions are constant over time, but increasing incomes change preference weights. Furthermore, parties react to changes in what voters want, and to the realignment that affects their support coalitions. We view our papers as complementary in pointing out different important factors that all contribute to polarization.

Informal narratives of realignment. While income in our model only affects the relative weight of moral values, but not values as such, some narratives assert that income and/ or higher education directly affect a voter's values. According to this argument, modernization made the elite's values more cosmopolitan and universalist,[32] which may have alienated working class voters and led to cultural backlash (Norris and Inglehart, 2019; Fukuyama, 2018).[33] Relatedly, an influential argument by authors such as Frank (2007) and Hacker and Pierson (2020) is that many working-class voters are convinced by Republican party elites to vote against their economic interests, in exchange for the socially conservative policies that the poor supposedly desire. In a nutshell, these are all stories of (cultural) conflict between rich and poor.

Our account is complementary in that we instead highlight the cultural or moral conflict between different subsets of the elite. This conflict among elites induces party polarization, which then propagates into changes in voting behavior among the poor. While this perspective may have received less attention in popular writings due its slightly higher complexity, it is backed by various pieces of evidence, such as our insight that the effect of moral values on voting and policy views is larger among the rich. Indeed, political scientists have long documented that the recent increase in the importance of social issues for voting is

---

[31]More generally, polarization is a central focus of the theoretical literature on electoral competition. Many variations of the spatial model may generate policy divergence, including policy motivation (e.g., Wittman 1983; Calvert 1985; Besley and Coate 1997; Martinelli 2001); entry deterrence (e.g., Palfrey 1984, Callander (2005)); incomplete information among voters or candidates (e.g. Castanheira 2003, Callander 2008); and candidates with differentiated abilities (e.g., Soubeyran 2009; Krasa and Polborn 2010, 2012b). The general equilibrium part of our paper relates to a literature emphasizing the importance of districts and legislatures (e.g., Snyder, 1994).

[32]For example, there is some evidence that negative economic shocks make people more socially conservative (e.g., Autor, Dorn, Hanson and Majlesi, 2020; Guriev and Papaioannou, 2020).

[33]If all voters shift left on some issue over time (e.g., gay marriage), then this only changes the level of policies, not the polarization between parties. If only some voters shift left on moral issues (such that the standard deviation of values in the population increases), social polarization may increase. However, it is not obvious why a model in which income affects *which* values people hold but not *how much values matter*, should produce our novel predictions, such as that more rich moral liberals vote against their material interests than poor moral concervatives, that within-party heterogeneity is higher on the political left, or that the income coefficient in a voting regression is larger among moral conservatives.

concentrated in the elite (Bartels, 2006), in contrast to what narratives such as "What's the matter with Kansas" suggest. Moreover, as discussed above, large focus group studies likewise identify rich people as predominantly occupied with concerns over fairness, morality and social policy topics (Hawkins et al., 2019).

While both cultural conflict among the rich and between rich and poor may be part of the overall story, we view our formal approach as a useful complement to the various qualitative narratives because our model not only explains the "big picture phenomena" of increased social polarization and realignment patterns over time, but it also makes various new micro-level predictions that are directly testable and find support in the data.

8 Conclusion

This paper has shown that the simple idea of income-dependent utility weights – which is bolstered by a large body of evidence on "modernization" or "postmaterialism" – generates a host of new testable predictions and sheds light on various widely-emphasized stylized facts about the nature of political polarization and realignment patterns in the U.S. In particular, our approach offers a new lens through which the increased salience of moral and cultural dimensions of political conflict can be understood.

An important open question is to which extent our insights are also applicable to the Western European context. As is well-known, some of the empirical regularities that motivate this paper are also visible in Europe, while others are not. On the one hand, the realignment of the educated elite towards the left (e.g., Gingrich and Häusermann, 2015; Gethin et al., 2021), and the contemporaneous importance of moral values (Enke et al., 2019) are both recurring phenomena throughout the Western world. On the other hand, measures of affective polarization have not increased nearly as much in Europe as in the U.S. (Boxell, Gentzkow and Shapiro, 2020). Moreover, there is strong anecdotal evidence that some European countries such as Germany have seen a *convergence* of the main political parties. One potential reason for this complicated pattern is differences in institutional environments. Most of Continental Europe has (pure or impure) forms of proportional representation (PR) systems rather than a majoritarian system like the U.S.

One reason why we believe that our approach could plausibly shed light also on the European context is our model prediction that the voters of left-wing parties will be more internally heterogeneous than those of right-wing parties. This prediction jives with a casual observation of the party landscape in some European countries. For example, in Germany, until the 1980s, politics was largely organized around a traditional economic left-right divide between the Social Democrats and the Christian Democrats. Then, the internal divisions within the Social Democratic Party became so large that the Green party (essentially the culturally very liberal wing of the Social Democrats) emerged. Relatedly, in Italy, a widely-discussed stylized fact is that the party landscape on the left is more fragmented than on the right, which puts left-wing parties at a "structural disadvantage" in terms of winning

elections.[34] While these are informal conjectures, we believe that a variant of our model that allows for party entry and a PR system is a promising avenue to formally explore these empirical regularities.

References

Ansolabehere, Stephen, James M Snyder, and Charles Stewart III, "Candidate positioning in US House elections," *American Journal of Political Science*, 2001, pp. 136–159.

_ , **Jonathan Rodden, and James M Snyder**, "Purple america," *Journal of Economic Perspectives*, 2006, *20* (2), 97–118.

_ , _ , **and** _ , "The strength of issues: Using multiple measures to gauge preference stability, ideological constraint, and issue voting," *American Political Science Review*, 2008, *102* (2), 215–232.

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), 3139–83.

Bade, Sophie, "Electoral competition with uncertainty averse parties," *Games and Economic Behavior*, 2011, *72* (1), 12–29.

Bartels, Larry M, "What's the Matter with What's the Matter with Kansas?," *Quarterly Journal of Political Science*, 2006, *1* (2), 201–226.

Bertrand, Marianne and Emir Kamenica, "Coming apart? Cultural distances in the United States over time," Technical Report, National Bureau of Economic Research 2018.

Besley, Timothy and Steve Coate, "An Economic Model of Representative Democracy," *Quarterly Journal of Economics*, 1997, *112*, 85–114.

_ **and Torsten Persson**, "The rise of identity politics," *LSE Documents*, 2019.

Bock, Sean, "Divided Dems, United Reps: Party strategy and within-party variation in immigration attitudes," 2020.

Bonomi, Giampaolo, Nicola Gennaioli, and Guido Tabellini, "Identity, Beliefs, and Political Conflict," *Quarterly Journal of Economics*, 2021.

Boxell, Levi, Matthew Gentzkow, and Jesse M Shapiro, "Cross-country trends in affective polarization," Technical Report, National Bureau of Economic Research 2020.

Burden, Barry C, "Candidate positioning in US congressional elections," *British Journal of Political Science*, 2004, pp. 211–227.

Bursztyn, Leonardo, Georgy Egorov, and Stefano Fiorin, "From extreme to mainstream: The erosion of social norms," *American economic review*, 2020, *110* (11), 3522–48.

Callander, Steven, "Electoral Competition in Heterogeneous Districts," *Journal of Political Economy*, 2005, *113*, 1116–1145.

_ , "Political Motivations," *Review of Economic Studies*, 2008, *75*, 671–697.

_ **and Juan Carlos Carbajal**, "Cause and Effect in Political Polarization: A Dynamic Analysis," *Journal of Political Economy*, 2022, *130* (4), 825–880.

Calvert, Randall L., "Robustness of the Multidimensional Voting Model: Candidate Motivations, Uncertainty, and Convergence," *American Journal of Political Science*, 1985, *29*, 69–95.

Castanheira, Micael, "Why Vote For Losers?," *Journal of the European Economic Association*, 2003, *1* (5), 1207–1238.

Cotofan, Maria, Lea Cassar, Robert Dur, and Stephan Meier, "Macroeconomic Conditions When Young Shape Job Preferences for Life," 2020.

Cramer, Katherine J, *The politics of resentment: Rural consciousness in Wisconsin and the rise of Scott Walker*, University of Chicago Press, 2016.

Desmet, Klaus and Romain Wacziarg, "The Cultural Divide," *Economic Journal*, 2021.

Dragu, Tiberiu and Xiaochen Fan, "An agenda-setting theory of electoral competition," *The Journal of Politics*, 2016, *78* (4), 1170–1183.

Enke, Benjamin, "Moral Values and Voting," *Journal of Political Economy*, 2020, pp. 3679–3729.

_ , **Ricardo Rodríguez-Padilla, and Florian Zimmermann**, "Moral Universalism and the Structure of Ideology," *Working Paper*, 2019.

Fiorina, Morris P, "Whatever happened to the median voter?," in "MIT Conference on Parties and Congress, Cambridge, MA," Vol. 2 1999, p. 1999.

Fiorina, Morris P. and Samuel J. Abrams, "Political polarization in the American public," *Annual Review of Political Science*, 2008, *11* (1), 563–588.

_ , _ , **and Jeremy C. Pope**, *Culture war? The myth of a polarized America*, Pearson Longman, 2006.

Frank, Thomas, *What's the matter with Kansas?: How conservatives won the heart of America*, Picador, 2007.

Fukuyama, Francis, *Identity: The demand for dignity and the politics of resentment*, Farrar, Straus and Giroux, 2018.

Gelman, Andrew, Boris Shor, Joseph Bafumi, and David Park, "Rich State, Poor State, Red State, Blue State: What's the Matter with Connecticut?," *Quarterly Journal of Political Science*, 2007, *2*, 345–367.

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

Gethin, Amory, Clara Martínez-Toledano, and Thomas Piketty, "Brahmin Left versus Merchant Right: Changing Political Cleavages in 21 Western Democracies, 1948-2020," *WID. world Working Paper*, 2021.

Gingrich, Jane and Silja Häusermann, "The decline of the working-class vote, the reconfiguration of the welfare support coalition and consequences for the welfare state," *Journal of European Social Policy*, 2015, *25* (1), 50–75.

Glaeser, Edward L, Giacomo AM Ponzetto, and Andrei Shleifer, "Why does democracy need education?," *Journal of economic growth*, 2007, *12* (2), 77–99.

Grossman, Gene and Elhanan Helpman, "Identity Politics and Trade Policy," *Review of Economic Studies*, 2021.

Guriev, Sergei and Elias Papaioannou, "The political economy of populism," 2020.

Hacker, Jacob S and Paul Pierson, *Let them Eat Tweets: How the Right Rules in an Age of Extreme Inequality*, Liveright Publishing, 2020.

Haidt, Jonathan, *The Righteous Mind: Why Good People are Divided by Politics and Religion*, Vintage, 2012.

Hawkins, Stephen, Daniel Yudkin, Miriam Juan-Torres, and Tim Dixon, "Hidden tribes: A study of America's polarized landscape," 2019.

Inglehart, Ronald, *Modernization and postmodernization: Cultural, economic, and political change in 43 societies*, Princeton university press, 1997.

_ , *Modernization and postmodernization: Cultural, economic, and political change in 43 societies*, Princeton university press, 2020.

_ **and Wayne E Baker**, "Modernization, cultural change, and the persistence of traditional values," *American sociological review*, 2000, pp. 19–51.

Kozlowski, Austin and James P Murphy, "Issue Alignment and Partisanship in the American Public: Revisiting the'Partisans without Constraint'Thesis," 2019.

Krasa, Stefan and Mattias K Polborn, "Competition between Specialized Candidates," *American Political Science Review*, 2010, *104* (4), 745–765.

 _ **and** _ , "Political competition between differentiated candidates," *Games and Economic Behavior*, 2012, *76* (1), 249–271.

 _ **and** _ , "Political Competition Between Differentiated Candidates," *Games and Economic Behavior*, 2012, *76*, 249–271.

 _ **and** _ , "Policy divergence and voter polarization in a structural model of elections," *Journal of Law and Economics*, 2014, *57* (1), 31–76.

 _ **and Mattias Polborn**, "Policy divergence and voter polarization in a structural model of elections," *Journal of Law and Economics*, 2014, *57* (1), 31–76.

 _ **and** _ , "Social ideology and taxes in a differentiated candidates framework," *American Economic Review*, 2014, *104* (1), 308–22.

Levendusky, Matthew, *The partisan sort: How liberals became Democrats and conservatives became Republicans*, University of Chicago Press, 2009.

Lipset, Seymour Martin, "Some social requisites of democracy: Economic development and political legitimacy1," *American political science review*, 1959, *53* (1), 69–105.

Martinelli, Cesar, "Elections with Privately Informed Parties and Voters," *Public Choice*, 2001, *108* (1), 147–167.

Maslow, Abraham Harold, "A theory of human motivation.," *Psychological review*, 1943, *50* (4), 370.

Mason, Lilliana, *Uncivil agreement: How politics became our identity*, University of Chicago Press, 2018.

McCarty, Nolan, Keith T Poole, and Howard Rosenthal, *Polarized America: The dance of ideology and unequal riches*, mit Press, 2016.

McKelvey, Richard D, "Intransitivities in multidimensional voting models and some implications for agenda control," *Journal of Economic theory*, 1976, *12* (3), 472–482.

Meltzer, Allan H and Scott F Richard, "A rational theory of the size of government," *Journal of political Economy*, 1981, *89* (5), 914–927.

Moskowitz, Daniel J, Jon C Rogowski, and James M Snyder, "Parsing Party Polarization in Congress," 2018.

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

Nunnari, Salvatore and Jan Zápal, "Dynamic elections and ideological polarization," *Political Analysis*, 2017, *25* (4), 505–534.

Palfrey, Thomas R., "Spatial Equilibrium with Entry," *Review of Economic Studies*, 1984, *51*, 139–156.

Plott, Charles R, "A notion of equilibrium and its possibility under majority rule," *American Economic Review*, 1967, pp. 787–806.

Polborn, Mattias K and James M Snyder, "Party polarization in legislatures with office-motivated candidates," *The Quarterly Journal of Economics*, 2017, *132* (3), 1509–1550.

Rodden, Jonathan A, *Why cities lose: The deep roots of the urban-rural political divide*, Hachette UK, 2019.

Rodrik, Dani, Elliot Ash, and Sharun W Mukand, "Economic Interests, Worldviews, and Identities: Theory and Evidence on Ideational Politics," *HKS Faculty Research Working Paper Series*, 2021.

Schofield, Norman, "The mean voter theorem: necessary and sufficient conditions for convergent equilibrium," *The Review of Economic Studies*, 2007, *74* (3), 965–980.

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.

Sherman, Jennifer, *Those who work, those who don't: Poverty, morality, and family in rural America*, U of Minnesota Press, 2009.

Shleifer, Andrei, "Does competition destroy ethical behavior?," *American Economic Review*, 2004, *94* (2), 414–418.

Snyder, James M, "Safe seats, marginal seats, and party platforms: The logic of platform differentiation," *Economics & Politics*, 1994, *6* (3), 201–213.

Soubeyran, Raphael, "Does a disadvantaged candidate choose an extremist position?," *Annals of Economics and Statistics/Annales d'Économie et de Statistique*, 2009, *93/94*, 327–348.

Wittman, Donald, "Candidate Motivation: A Synthesis of Alternative Theories," *American Political Science Review*, 1983, *77*, 142–157.

Wu, Alex, "Political polarization," *Undergraduate senior thesis, Harvard University*, 2020.

Xefteris, Dimitrios, "Multidimensional electoral competition between differentiated candidates," *Games and Economic Behavior*, 2017, *105*, 112–121.

ONLINE APPENDIX

A   Proofs

Proof of Proposition 1. In the text.   □

Proof of Proposition 2. Differentiating (9) with respect to μ yields

$$
\begin{aligned}
\frac{d^2 Pr(R)}{d\mu dy} =& \psi(\Delta U)w'(y)[h'(\mu)\Delta x_e + \Delta x_s] \\
& + \psi'(\Delta U)\frac{d\Delta U}{d\mu}\left\{[1-w(y)]\Delta x_e r'(y-\bar{y}) + w'(y)[\Delta x_e(h(\mu) - r(y-\bar{y})) + \Delta x_s \mu]\right\}.
\end{aligned}
\tag{13}
$$

Positivity if $|\psi'|$ is sufficiently small follows because the second term then is close to zero, and the term in the first line of (13) is positive: The density, the derivative of the weight, and both terms in the parenthesis are positive.   □

What happens if ψ is not constant, but rather single-peaked (for example, if it is normal)?

Intuitively, under normal circumstances, increases in both μ and y will increase a voter's likelihood of voting Republican.[35]

Moreover, as μ increases, the effect of an income increase *on the utility difference* ΔU increases. Thus, μ and y are complements in terms of shifting up the cutoff value of the critical ε shock (remember that voters vote Republican if their ε is smaller than the critical value).

The only case in which this complementarity in terms of shifting the cutoff does not translate into complementarity in terms of the probability of voting Republican is if the density ψ decreases rapidly.

To see this more formally, observe first that $d\Delta U/d\mu$ is positive. As we show in the proof of Proposition 4, the term in curly brackets in the second line of (13) is positive whenever the slope of the isoprobability curve is negative (see discussion in footnote 35 why this is arguably the standard case).

In this case, $\psi' \geq 0$ is sufficient (but not necessary) for $\frac{d^2 Pr(R)}{d\mu dy} > 0$ because the density effect further strengthens the shift of the cutoff. In contrast, if $\psi' << 0$, it is possible that μ and y are not complements in terms of the probability of voting Republican.

Proof of Proposition 3. The probability of voting against one's economic interest for RML is $Pr(D|RML) = 1 - \Psi(U^R - U^D) = \Psi(U^D - U^R)$, and for PMC is $Pr(R|PMC) = \Psi(U^R - $

---

[35]As shown, this is definitely the case for an increase in μ, and while an opposite effect is possible for y for very socially-liberal voters, the standard case is probably that richer people are more likely to prefer the Republicans.

U^D). Given that Ψ is symmetric around zero, the claim holds if and only if $(U^D - U^R)_{RML} > (U^R - U^D)_{PMC}$. By (6), we have for the RML

$$(U^D - U^R)_{RML} = - [1 - w(y_{RML})]\Delta x_e \left[r\left(y_{RML} - \bar{y} \right) \right]$$
$$- w(y_{RML}) \left[\Delta x_e h(\mu_{RML}) + \Delta x_s \mu_{RML} \right]$$

and for PMC, we have

$$(U^R - U^D)_{PMC} = [1 - w(y_{PMC})]\Delta x_e \left[r\left(y_{PMC} - \bar{y} \right) \right]$$
$$+ w(y_{PMC}) \left[\Delta x_e h(\mu_{PMC}) + \Delta x_e \mu_{PMC} \right].$$

Differencing these two expressions, and substituting $r\left(y_{PMC} - \bar{y} \right) = -r\left(y_{RML} - \bar{y} \right)$ and $\mu_{RML} = -\mu_{PMC}$ yields

$$(U^D - U^R)_{RML} - (U^R - U^D)_{PMC} = \tag{14}$$
$$[w(y_{RML}) - w(y_{PMC})] \left\{ \Delta x_e h(\mu_{PMC}) + \Delta x_s \mu_{PMC} + \Delta x_e r\left(y_{RML} - \bar{y} \right) \right\} > 0,$$

because $w(y_{RML}) > w(y_{PMC})$ and all terms in the curly brackets are positive. Thus, it follows that the probability that RML votes against his economic interest is larger than that PMC votes against his economic interest. $\square$

Proof of Proposition 4. 1. The first line in (10) follows directly from applying the implicit function theorem to (6). The second line follows from dividing both numerator and denominator by $[1 - w(y)]\Delta x_e$, and observing that $\frac{d}{dy}\left(\frac{w(y)}{1-w(y)} \right) = \frac{w'(y)}{(1-w(y))^2}$.

Since all terms in the numerator are positive, and both $r' > 0$ and $\frac{d}{dy}\left(\frac{w(y)}{1-w(y)} \right) > 0$, a sufficient condition for an iso-probability curve to be downward sloping is that $[h(\mu) - r(y - \bar{y}) + \frac{\Delta x_s}{\Delta x_e}\mu] \geq 0$, though, clearly, the same result would still hold if it is slightly negative.

Similarly, if $w(y) = \frac{c+\rho(y-k)}{1+\rho(y-k)}$, then (10) simplifies to

$$\frac{dy}{d\mu} = - \frac{\frac{c+\rho(y-k)}{1-c}\left[h'(\mu) + \frac{\Delta x_s}{\Delta x_e} \right]}{r'(y-\bar{y}) + \frac{\rho}{1-c}\frac{1-c}{1+\rho(y-k)}[h(\mu) - r(y-\bar{y}) + \frac{\Delta x_s}{\Delta x_e}\mu]}$$
$$= - \frac{h'(\mu) + \frac{\Delta x_s}{\Delta x_e}}{r'(\cdot)\frac{1-c}{c+\rho(y-k)} + \frac{\rho(1-c)}{[c+\rho(y-k)][1+\rho(y-k)]}[h(\mu) - r(y-\bar{y}) + \frac{\Delta x_s}{\Delta x_e}\mu]}. \tag{15}$$

Observe that the numerator (of the term in the second line) in (15) is independent of y and μ (because h is linear). If $\rho \to 0$, then (15) goes to $-\frac{c}{(1-c)r'}\left[h'(\mu) + \frac{\Delta x_s}{\Delta x_e} \right] < 0$.

2. Using $r(y) = r_0(y - \bar{y})$ and $h(\mu) = h_0\mu$ with $r_0, h_0 > 0$, we can rewrite (10) as

$$\frac{dy}{d\mu} = - \frac{h_0 + \frac{\Delta x_s}{\Delta x_e}}{\frac{1-w}{w}r_0 + \frac{w'}{w}\left(h_0\mu - r_0(y-\bar{y}) + \mu\frac{\Delta x_s}{\Delta x_e} \right)} \tag{16}$$

For downward-sloping isoprobability curves, the denominator is positive, which implies

$$h_0\mu - r_0(y - \bar{y}) + \mu\frac{\Delta x_s}{\Delta x_e} \geq -\frac{1 - w}{w'}r_0. \tag{17}$$

To show the claim, we have to show that the value of the denominator in (16) increases at μ increases (because the numerator is a constant in μ and y).

Differentiating the denominator with respect to μ, taking into account that, along an isoprobability curve, y is a function of μ, yields

$$\left[-\frac{w'}{w^2}r_0 + \frac{w''w - (w')^2}{w^2}\left(h_0\mu - r_0(y - \bar{y}) + \mu\frac{\Delta x_s}{\Delta x_e}\right) - \frac{w'}{w}r_0\right]\frac{dy}{d\mu} + \frac{w'}{w}\left(h_0 + \frac{\Delta x_s}{\Delta x_e}\right). \tag{18}$$

Clearly, the second summand is positive, so, given that $dy/d\mu < 0$, to show that (18) is positive, it is sufficient to show that the term in square brackets is negative.

By assumption, $\frac{w''w - (w')^2}{w^2} = \frac{d}{dy}\left(\frac{w'(y)}{w(y)}\right) \leq 0$. Together with (17), this implies that the term in square brackets is smaller than

$$-\frac{w'}{w^2}r_0 + \frac{w''w - (w')^2}{w^2}\left(-\frac{1 - w}{w'}r_0\right) - \frac{w'}{w}r_0. \tag{19}$$

Putting all terms in (19) in a common fraction, we have

$$-\frac{w'^2 + w''w(1 - w) - (1 - w)(w')^2 + (w')^2 w}{w^2 w'}r_0 = -\frac{w''(1 - w) + 2w'^2}{ww'}r_0 \leq 0. \tag{20}$$

Observe that, since $r_0 > 0$, (20) holds if and only if

$$w''(1 - w) + 2w'^2 \geq 0 \tag{21}$$

for all y. In particular, this is satisfied whenever w is linear or convex,[36], and when it is not too concave.

If $w(y) = \frac{c + \rho(y - k)}{1 + \rho(y - k)}$, then $w'(y) = \frac{\rho(1 - c)}{(1 + \rho(y - k))^2}$ and $w''(y) = -\frac{2\rho^2(1 - c)}{(1 + \rho(y - k))^3}$. Substituting, the left hand side of (21) is identically zero for all y. (Note that this does not mean that $w(y) = \frac{c + \rho(y - k)}{1 + \rho(y - k)}$ is a knife-edge case. After all, the second summand in (18) is strictly positive.) $\square$

Proof of Proposition 5. We have
$$\frac{\partial x_e^*}{\partial \mu} = w(y)h'(\mu)$$

As the left-hand side of (11) is evaluated at a higher income level than the right-hand side, and $w(\cdot)$ is an increasing function, the claim follows. $\square$

---

[36]Though, note that it should not be too convex, as we have used the assumption that $\frac{d}{dy}\left(\frac{w'(y)}{w(y)}\right) \leq 0$.

Proof of Proposition 6. Consider how $U^R - U^D$ changes with β. Since a voter's relative income remains unchanged by β, we have

$$\frac{\partial(U^R - U^D)}{\partial \beta} = \frac{w'(y + \ln(\beta))}{\beta} \Delta x_e \left\{ h(\mu) + \frac{\Delta x_s}{\Delta x_e} \mu - [r(y - \bar{y})] \right\} \tag{22}$$

As the first two terms are positive, the sign of this expression is the same as the sign of the term in curly brackets, as claimed. $\square$

Proof of Proposition 7.   1. Let $\Delta x_{p,t} = x_{p,t}^R - x_{p,t}^D$ be the policy polarization between Republicans and Democrats in policy $p \in \{e, s\}$ at time $t \in \{0, 1\}$. Inspection of (6) shows that, if $\frac{\Delta x_{s,1}}{\Delta x_{s,0}} = \frac{\Delta x_{e,1}}{\Delta x_{e,0}} = k > 1$, then $U_1^R - U_1^D = k(U_0^R - U_0^D)$. Thus, if $U_0^R - U_0^D > 0$ (so that $\Psi(U_0^R - U_0^D) > 1/2$), then $U_1^R - U_1^D > U_0^R - U_0^D$, and hence $\Psi(U_1^R - U_1^D) > \Psi(U_0^R - U_0^D)$. The case that $U_0^R - U_0^D < 0$ is analogous.

2. If $\frac{\Delta x_{s,1}}{\Delta x_{s,0}} > 1$ while economic positions are unchanged, then inspection of (6) shows that $U^R - U^D$ increases if and only if $\mu > 0$. This immediately implies the claim. $\square$

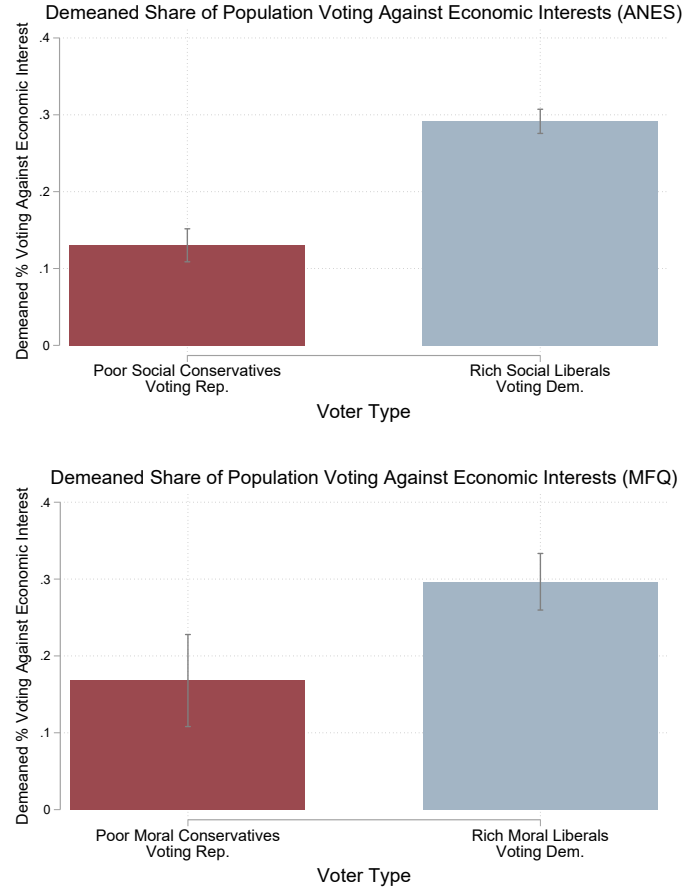# B Additional Figures for Empirical Analyses (ANES and MFQ)

Figure 8: This figure reports the demeaned share and 95% confidence interval of the share of poor social conservatives who vote Republican, and rich social liberals who vote Democrat. The demeaning is done by subtracting the overall Republican vote share from the poor conservative Republican share, and the overall Democratic vote share from the rich liberals Democratic share. Here, we define "rich" and "poor" as the top/bottom third of household income, and "socially conservative / liberal" as top/bottom third of the social / moral conservatism indices. The panel on the left uses ANES data ($N = 4,274$) and the panel on the right uses MFQ data ($N = 608$). In t-tests of the equality of means across groups, the p-value is $p < 0.01$ in both the ANES and the MFQ.
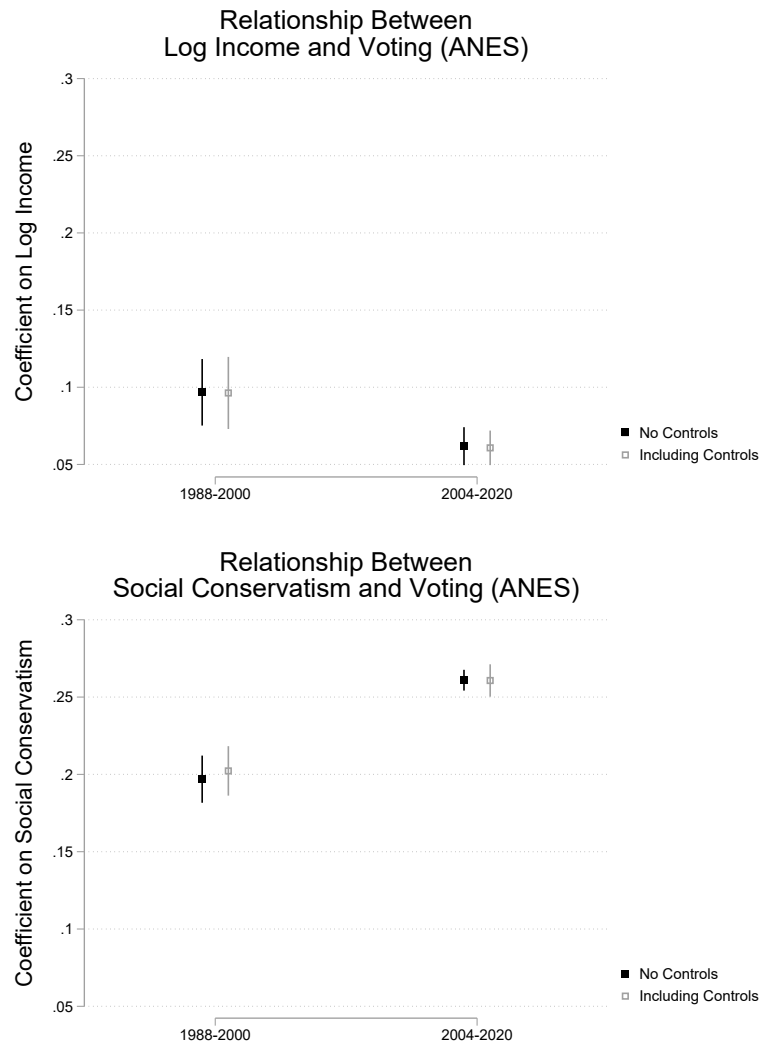
Figure 9: This figure reports coefficients from OLS regressions of an indicator for voting Republican on log income (top panel) and social conservatism (bottom panel) split by whether the survey year is pre- or post-2000. 95% confidence intervals reported. In a Wald test of the equality of coefficients between groups, the p-value is $p = 0.005$ for the top panel, and $p < 0.0001$ for the bottom panel. Data are from the ANES. Total $N = 15,505$. The conservatism index has mean zero and standard deviation one. The corresponding regression results are reported in Table 7 in Appendix C

C Additional Tables for Empirical Analyses (ANES and MFQ)

Table 2: Economic Conservatism and Income (ANES)

	Dependent variable:				
	Economic conservatism (ANES)				
	(1)	(2)	(3)	(4)	(5)
Log [HH income]	0.15***	0.21***	0.22***	0.21***	0.15***
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
Log [HH income] $\times$ Social conservatism		0.083***	0.082***	0.081***	0.068***
		(0.01)	(0.01)	(0.01)	(0.01)
Social conservatism		0.36***	0.37***	0.36***	0.38***
		(0.01)	(0.01)	(0.01)	(0.01)
State FE	No	No	Yes	Yes	Yes
Survey year FE	No	No	Yes	Yes	Yes
Baseline controls	No	No	No	Yes	Yes
Extended controls	No	No	No	No	Yes
Observations	22972	17576	17576	17576	17576
R^2	0.03	0.20	0.29	0.30	0.35

Notes. OLS estimates, clustered standard errors in parentheses. The dependent variable is the economic conservatism index, standardized into a z-score. Log income is standardized into a z-score. Baseline controls include age, age squared, and gender fixed effects. Extended controls include educational attainment fixed effects and race fixed effects. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3: Income Gradient in Voting (ANES)

| | Dependent variable: | | | | |
| | Indicator for Voting Republican (ANES) | | | | |
	(1)	(2)	(3)	(4)	(5)
Log [HH income]	0.037***	0.073***	0.072***	0.069***	0.034***
	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)
Log [HH income] × Social conservatism		0.033***	0.032***	0.030***	0.023***
		(0.00)	(0.00)	(0.00)	(0.00)
Social conservatism		0.22***	0.23***	0.23***	0.24***
		(0.00)	(0.00)	(0.01)	(0.00)
State FE	No	No	Yes	Yes	Yes
Survey year FE	No	No	Yes	Yes	Yes
Baseline controls	No	No	No	Yes	Yes
Extended controls	No	No	No	No	Yes
Observations	28554	15505	15505	15505	15505
R^2	0.01	0.26	0.29	0.29	0.39

Notes. OLS estimates, clustered standard errors in parentheses. The dependent variable is an indicator which equals 100 if a respondent votes Republican and 0 if they vote Democrat. Baseline controls include age, age squared, and gender fixed effects. Extended controls include educational attainment fixed effects and race fixed effects. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: Income Gradient in Voting (MFQ)

| | Dependent variable: | | | | |
| | Indicator for Voting Republican (MFQ) | | | | |
	(1)	(2)	(3)	(4)	(5)
Log Income [std.]	0.039***	0.025***	0.034***	0.021***	0.0047
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Log Income [std.] × Moral conservatism		0.033***	0.030***	0.030***	0.026***
		(0.01)	(0.01)	(0.01)	(0.00)
Moral conservatism		0.21***	0.20***	0.20***	0.18***
		(0.01)	(0.01)	(0.01)	(0.01)
State FE	No	No	Yes	Yes	Yes
Baseline controls	No	No	No	Yes	Yes
Extended controls	No	No	No	No	Yes
Observations	3471	3471	3471	3471	3471
R^2	0.01	0.20	0.23	0.24	0.34

Notes. OLS estimates, clustered standard errors in parentheses. The dependent variable is an indicator which equals 1 if a respondent votes Republican and 0 if they vote Democrat. Baseline controls include age, age squared, and gender fixed effects. Extended controls include educational attainment fixed effects and race fixed effects. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: Ideological Constraint over Time (ANES)

	Dependent variable:				
	Economic conservatism (ANES)				
	(1)	(2)	(3)	(4)	(5)
Social conservatism	0.36***	0.15***	0.15***	0.14***	0.17***
	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
Survey year × Social conservatism		0.011***	0.011***	0.011***	0.011***
		(0.00)	(0.00)	(0.00)	(0.00)
Survey year		0.0064***	0.0070***	0.0065***	0.0053***
		(0.00)	(0.00)	(0.00)	(0.00)
State FE	No	No	Yes	Yes	Yes
Baseline controls	No	No	No	Yes	Yes
Extended controls	No	No	No	No	Yes
Observations	18432	18432	18432	18432	18432
R^2	0.14	0.16	0.18	0.19	0.27

Notes. OLS estimates, clustered standard errors in parentheses. The dependent variable is the economic conservatism index, standardized into a z-score. "Survey year" is year of survey minus 1988. Baseline controls include age, age squared, and gender fixed effects. Extended controls include educational attainment fixed effects and race fixed effects. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: Ideological Constraint over Time by Income (ANES)

| | Rich Voters | | Poor Voters | | Full |
| | Early | Late | Early | Late | Sample |
	(1)	(2)	(3)	(4)	(5)
			Dependent variable:		
			Economic conservatism (ANES)		
Log [HH income]	0.092*	0.043	0.21***	0.20***	0.96
	(0.05)	(0.03)	(0.02)	(0.02)	(1.72)
Social conservatism	0.28***	0.55***	0.26***	0.38***	-20.1***
	(0.02)	(0.02)	(0.02)	(0.01)	(1.76)
Log [HH income] $\times$ Social conservatism					-6.65***
					(1.24)
Year					0.0045***
					(0.00)
Log [HH income] $\times$ Year					-0.00037
					(0.00)
Social conservatism $\times$ Year					0.010***
					(0.00)
Log [HH income] $\times$ Social conservatism $\times$ Year					0.0033***
					(0.00)
Observations	2010	4190	3588	7788	17576
R^2	0.100	0.29	0.13	0.16	0.22

Notes. OLS estimates, clustered standard errors in parentheses. The dependent variable is the economic conservatism index, standardized into a z-score. The first four columns present results from subsamples of the data, where "rich voters" are in the top tercile of the income distribution in a year, and "poor voters" are in the bottom tercile. "Early" and "late" split the sample by time, early being 1988-2000 and "late" being 2004-2020. The fifth column uses the full sample. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 7: Importance of Income and Social Conservatism in Voting Over Time (ANES)

| | *Dependent variable:* | | | | |
| | Indicator for Voting Republican (ANES) | | | | |
	(1)	(2)	(3)	(4)	(5)
Log [HH income]	0.070***	0.094***	0.098***	0.095***	0.040***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Social conservatism	0.24***	0.14***	0.13***	0.14***	0.14***
	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)
Survey year × Log [HH Income]		-0.00098**	-0.0011**	-0.0010*	-0.000098
		(0.00)	(0.00)	(0.00)	(0.00)
Survey year × Social conservatism		0.0047***	0.0046***	0.0045***	0.0044***
		(0.00)	(0.00)	(0.00)	(0.00)
State FE	No	No	Yes	Yes	Yes
Baseline controls	No	No	No	Yes	Yes
Extended controls	No	No	No	No	Yes
Observations	15505	15505	15505	15505	15505
R^2	0.26	0.27	0.28	0.28	0.38
Prob > F		9.1e-12	3.1e-12	5.3e-12	6.5e-13

Notes. OLS estimates, clustered standard errors in parentheses. The dependent variable is an indicator for voting Republican. "Survey year" is year of survey minus 1988. Baseline controls include age, age squared, and gender fixed effects. Extended controls include educational attainment fixed effects and race fixed effects. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

D   Details and Analyses for Model Simulations

D.1   Figures for Baseline Simulations
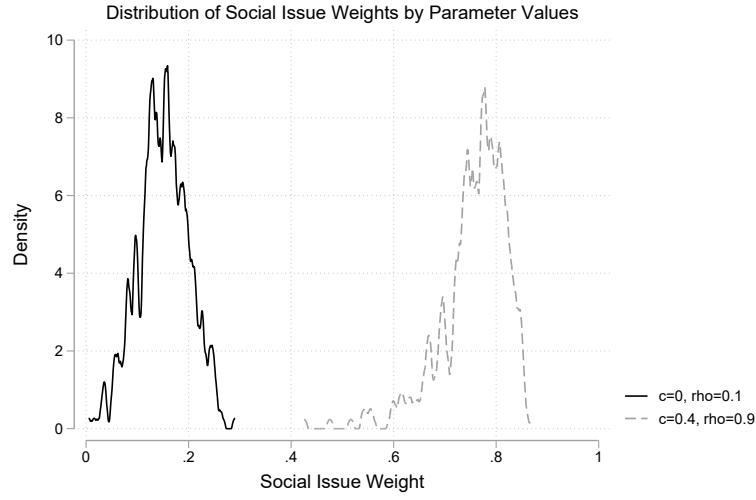
Figure 10: This figure plots the distribution of weights that voters place on social issues for two parameter combinations, $c = 0, \rho = 0.1, k = 9$ and $c = 0.4, \rho = 0.9, k = 9$.
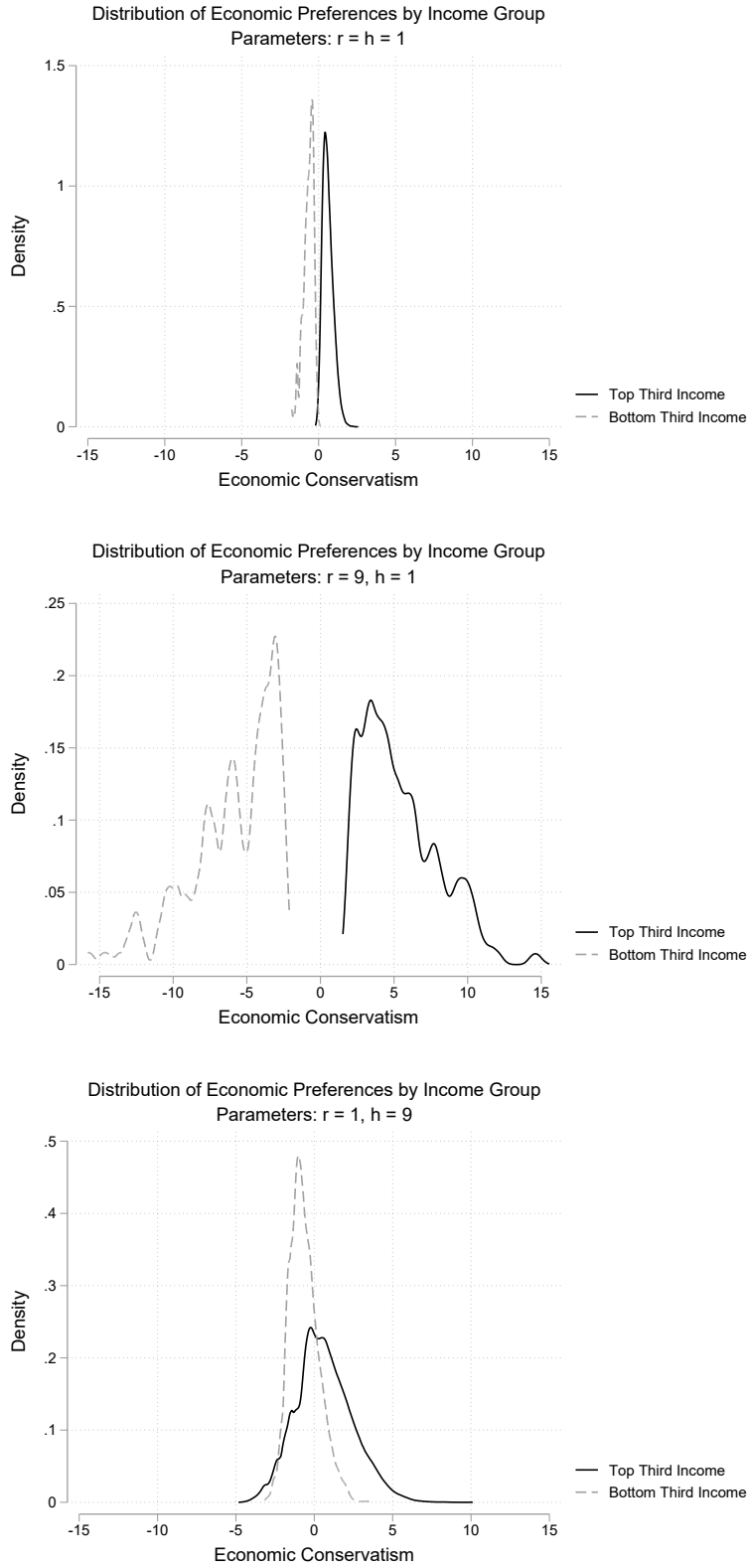
Figure 11: This figure plots the distribution of voters' economic views, split by whether voters are in the top or bottom third of the income distribution. The parameters are $r_0 = h_0 = 1$ in the top panel (income and moral values are equally important for economic views), $r_0 = 9$ and $h_0 = 1$ (income is significantly more important than moral values) in the middle panel, and $r_0 = 1$ and $h_0 = 9$ (moral values are significantly more important than income)in the bottom panel. Across all three panels, $c = 0$, $\rho = 0.1$, and $k = 9$.
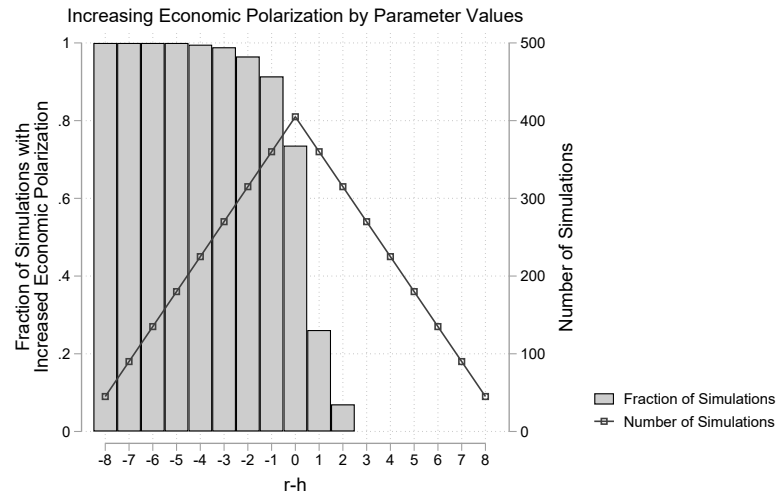
Figure 12: This bar graph plots the share of simulations in which economic polarization increases over time, separately for different values of $r_0 - h_0$. The line graph with the y-axis on the right plots the number of simulations for each $r_0 - h_0$ value.
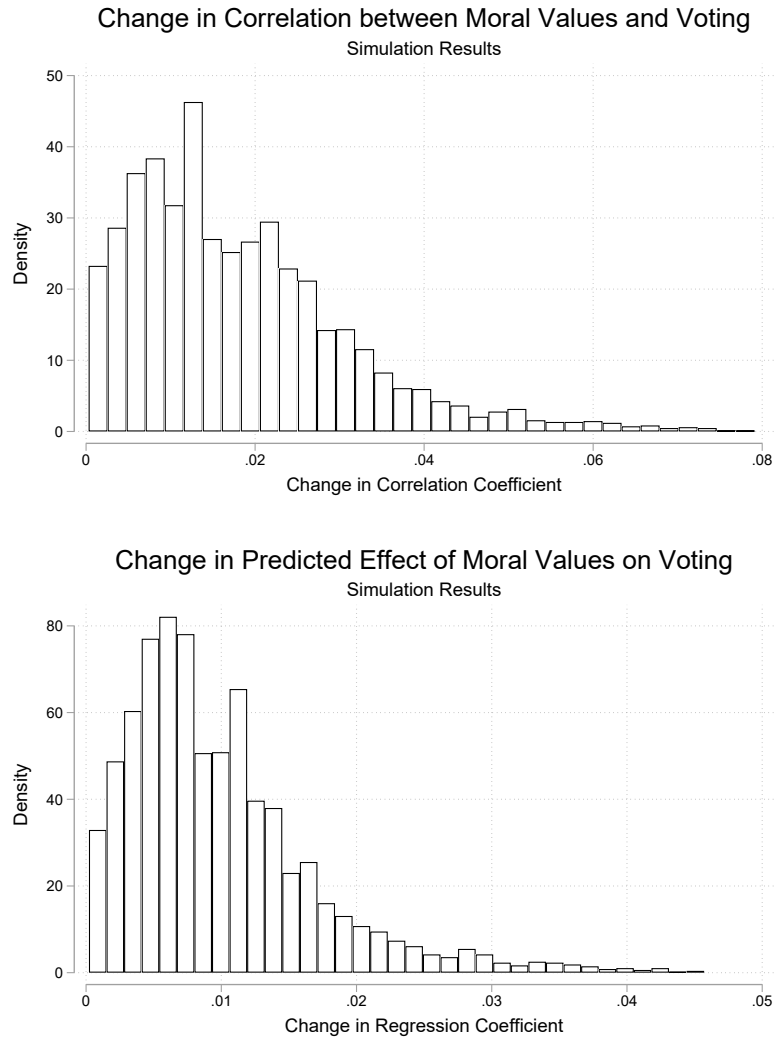
Figure 13: To analyze how the relationship between moral values and voting changes over time in the simulations, we calculate two measures of the relationship in the "early" ($\beta = 0$) and "late" ($\beta = 0.23$) periods of the simulation. First, we compute the correlation between moral conservatism and an indicator for voting Republican (observations are individual voters with varying levels of income and moral conservatism). Then, we calculate the change in the correlation from the early period to the late period, where positive values mean the correlation between moral values and voting increases over time. The top panel plots the distribution of the change in correlation between moral conservatism and voting Republican, where variation in changes comes from different parameters used in the simulation. The second measure we calculate is the coefficient on moral conservatism in a linear regression of an indicator for voting Republican on moral conservatism. We calculate the change in this regression coefficient from the early period to the late period, and the bottom panel plots the distribution of changes across simulation parameters.
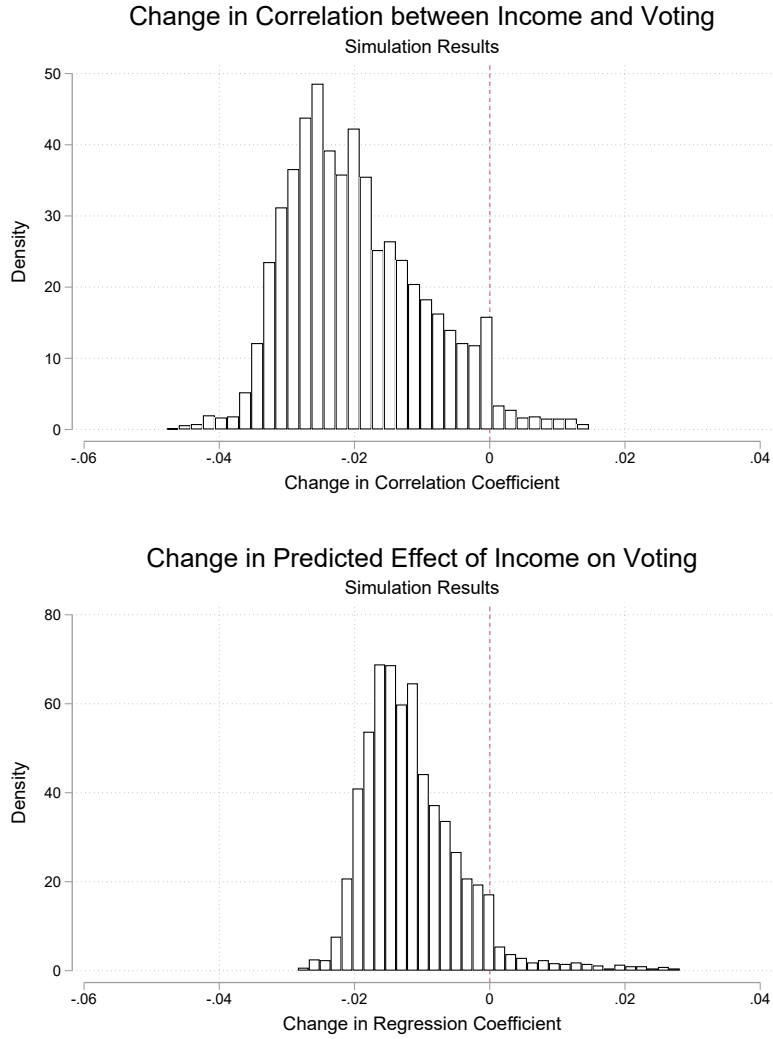
Figure 14: To analyze how the relationship between (log) income and voting changes over time in the simulations, we calculate two measures of the relationship in the "early" ($\beta = 0$) and "late" ($\beta = 0.23$) periods of the simulation. First, we compute the correlation between log income and an indicator for voting Republican (observations are individual voters with varying levels of log income and moral conservatism). Then, we calculate the change in the correlation from the early period to the late period, where negative values mean the correlation between log income and voting decreases over time. The top panel plots the distribution of the change in correlation between log income and voting Republican, where variation in changes comes from different parameters used in the simulation. The second measure we calculate is the coefficient on log income in a linear regression of an indicator for voting Republican on log income. We calculate the change in this regression coefficient from the early period to the late period, and the bottom panel plots the distribution of changes across simulation parameters.
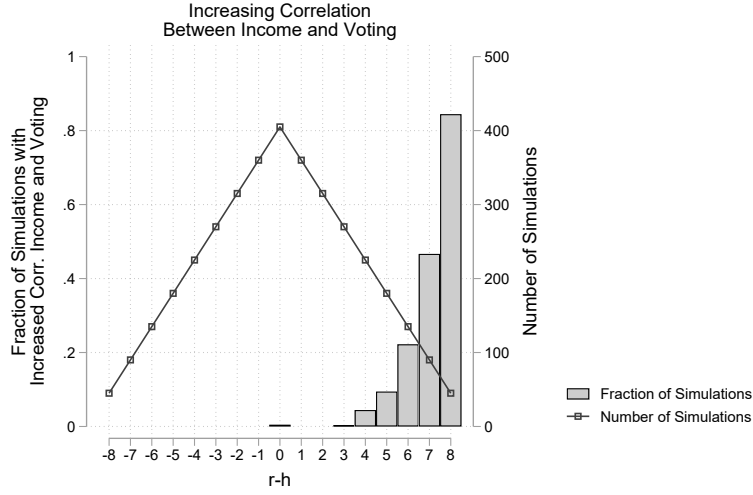
Figure 15: This figure plots the fraction of simulations in which the correlation between log income and an indicator for voting Republican increaes over time, separately for each value of $r_0 - h_0$. The line graph plots the number of simulations with a given value of $r_0 - h_0$.
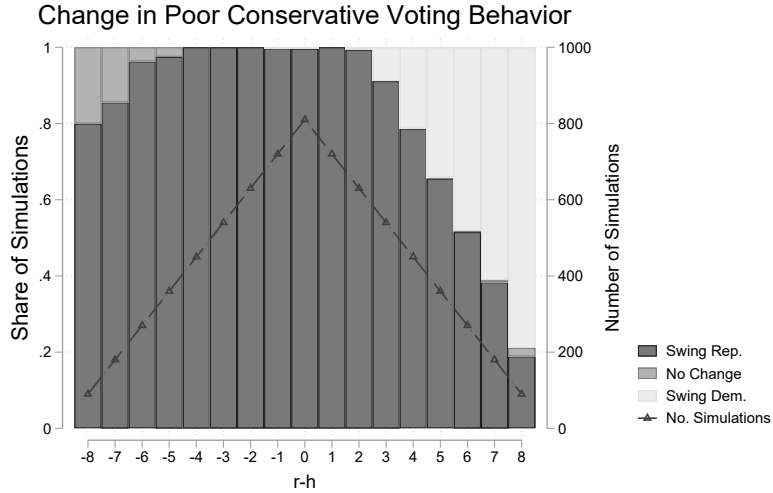

Figure 16: The bar graph plots the fraction of cases in which poor conservatives (bottom 1/3 income, top 2/5 moral conservatism) are more likely, equally as likely, or less likely to vote Republican over time as income increases. This is done separately for each value of $r_0 - h_0$. The line graph plots the number of simulations within each $r_0 - h_0$ value.

D.2 Details for Income Inequality Extension Simulations

D.2.1 Procedure

As in the main analyses, we model log income in the "early" period as $y \sim N(11, 0.75)$, which is roughly calibrated to the U.S. household income distribution in 1988. To assess the effect of increases in income inequality (largely driven by the growing incomes of the rich), we scale log income multiplicately by a scaling factor κ. Thus, log income is distributed

$\kappa \cdot y \sim N(\kappa \cdot 11, \kappa^2 \cdot 0.75)$ in the "late period". Since the scaling is multiplicative, the absolute log income of rich voters increases more than the absolute log incomes of the poor; thus, the difference between their log incomes increases, i.e., the ratio of the absolute income of the rich relative to the income of the poor grows. This is in contrast to the additive scaling in the analyses in the main text, which preserves relative income in the population.

We choose $\kappa = 1.015$, which corresponds to the approximate change in absolute income for the 25th percentile, median, 75th percentile, 90th percentile, and 95th percentile of U.S. households from 1988 to 2020. Table 8 illustrates this by presenting data on the U.S. household income distribution, along with the corresponding incomes in our simulated dataset.

Table 8: Income Distribution in U.S. Data and Inequality Simulation

	1988		2020	
Income Percentile	U.S. Data	Simulated Data	U.S. Data	Simulated Data
5th Pctile	$7,000	$14,500	$9,400	$16,800
10th Pctile	$17,600	$20,000	$16,000	$23,000
25th Pctile	$32,000	$33,000	$34,000	$39,000
Median	$59,500	$59,500	$68,000	$70,500
75th Pctile	$90,000	$107,000	$123,000	$128,000
90th Pctile	$176,000	$180,000	$203,000	$208,000
95th Pctile	$206,000	$246,000	$273,000	$276,000

This table presents data from the Census Bureau on the U.S. household income distribution in 1988 and 2020, along with the simulated income distribution implied by model parameters. In the "early" period of the simulation (1988), log income is distributed $N(11, 0.75)$. In the "late" period of the simulation (2020), log income is distributed $N(\kappa \cdot 11, \kappa^2 \cdot 0.75)$ where $\kappa = 1.015$. This table reports the corresponding levels of income associated with various percentiles of both log income distributions.

The increase in income inequality has meaningful effects on voters' policy preferences. Consider a voter in the "early" period with log income equal to the 30th percentile of the early income distribution. For this voter to be indifferent between the parties under the standard set of parameters used to illustrate our simulation in the main text and fixed party positions,[37] they would need to have moral values equal to the 75th percentile of the moral conservatism distribution. Next, consider that same voter, who now lives in a society with the "late" income distribution. Although this voter has the same level of absolute income as before, he is relatively poorer because the rich have become richer. He now has a greater economic incentive to vote Democrat, and for him to be indifferent between parties, he needs to have moral values equal to the 90th percentile of the moral conservatism distribution.

---

[37]The parameters are $c = 0$, $\rho = 0.2$, $k = 9$, $h_0 = 1$, and $r_0 = 1$. The Democrat positions are -0.5 on both social and economic issues, and the Republican positions are 0.5 on both issues.

D.2.2 Results

Table 9: Simulation results regarding effects of increase in κ

	Fraction of simulations in which variable in col. (1)		
	Increases	Constant	Decreases
Party polarization on social issues	99.2	0.0	0.8
Party polarization on economic issues	71.0	0.0	29.0
Moral polarization in electorate	>99.9	0	<0.1
Income polarization in electorate	20.0	0	80.0
Rich moral liberals voting Democratic	79.0	21.0	0.0
Poor moral conservatives voting Republican	92.0	1.0	7.0

Notes. This table presents the fraction of inequality simulations in which the variable in column 1 increases, remains constant, or decreases over time (as κ increases). Social (economic) party polarization increases if the distance between party positions on social (economic) issues increases. Moral (income) polarization in the electorate increases if the difference in average moral conservatism μ (income y) between voters who vote Republican and those who vote Democrat increases. To compute this, we calculate weighted averages of moral conservatism and income in the population, where the weights are given by the probability of voting Republican and Democrat, and take the difference between the two. Rich moral liberals are defined as top 1/3 income and bottom 2/5 moral conservatism. Poor moral conservatives are defined analogously.

Table 9 shows that social polarization increases in essentially all of the simulations and economic polarization increases in 71% of simulations. We find more cases of economic polarization in the inequality simulations than in the baseline simulations (71% vs. 55%) - this is because growing income inequality increases the likelihood that voters sort along income rather than moral values, and parties respond by setting more extreme economic policy to cater to their supporters.

Table 9 shows that in all of our simulations, the average Republican and the average Democratic voter diverge on moral issues, and moral polarization in the electorate increases. Income-based polarization in the electorate decreases in 80% of the inequality simulations - this is less than the 100% of cases in the baseline simulation, and it is driven by cases where the growing income inequality outweighs increasing social polarization for a subset of voters.

On average across all the simulations, rich liberals become more likely to vote Democrat

and poor conservatives become more likely to vote Republican. However, the extent of sorting of poor conservatives into the Republican party is less extreme relative to the baseline simulations. This is because income inequality makes poor conservatives more economically liberal, which pushes against their moral conservatism.
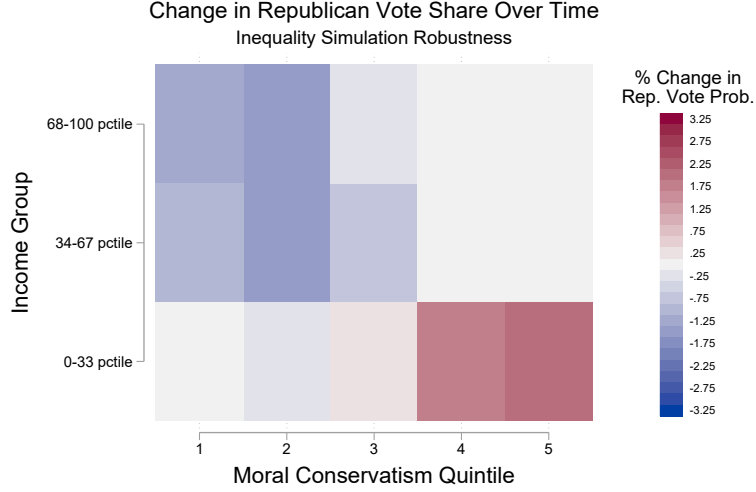

Figure 17: Percentage point change in simulated Republican vote share as a function of fundamentals (y, μ) between the early period and late period in which incomes are scaled multiplicatively by $\beta = 1.015$. This figure shows average results across all simulations.

Figure 18 shows that moral values always become more important for voting over time as incomes rise and income inequality increases. Figure 19 shows that income almost always becomes less important for voting, except for a handful of cases. As in the baseline simulations, these cases have large values of $r_0 - h_0$. Intuitively, when voters care more about income than moral values, rising income inequality leads poor voters (and in particular poor conservatives) to vote Democrat, which produces an stronger correlation between income and voting Republican.
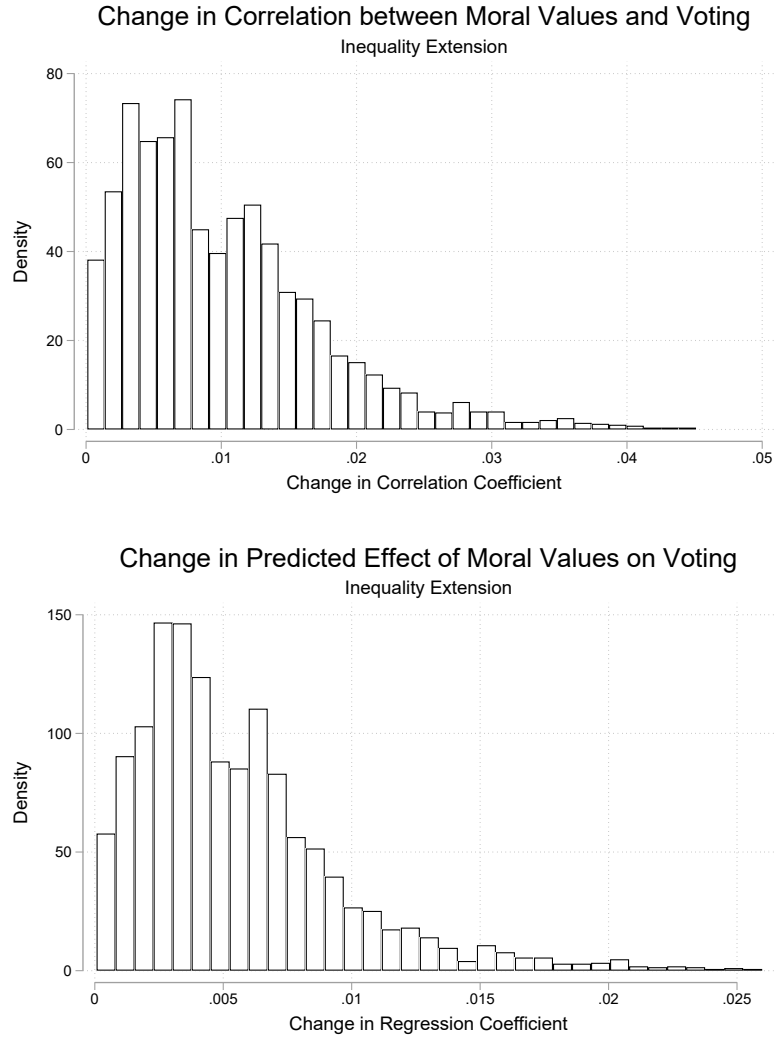
Figure 18: To analyze how the relationship between moral values and voting changes over time in the inequality simulations, we calculate two measures of the relationship in the "early" ($\beta = 1$) and "late" ($\beta = 1.015$) periods of the simulation. First, we compute the correlation between moral conservatism and an indicator for voting Republican (observations are individual voters with varying levels of income and moral conservatism). Then, we calculate the change in the correlation from the early period to the late period, where positive values mean the correlation between moral values and voting increases over time. The top panel plots the distribution of the change in correlation between moral conservatism and voting Republican, where variation in changes comes from different parameters used in the simulation. The second measure we calculate is the coefficient on moral conservatism in a linear regression of an indicator for voting Republican on moral conservatism. We calculate the change in this regression coefficient from the early period to the late period, and the bottom panel plots the distribution of changes across simulation parameters.
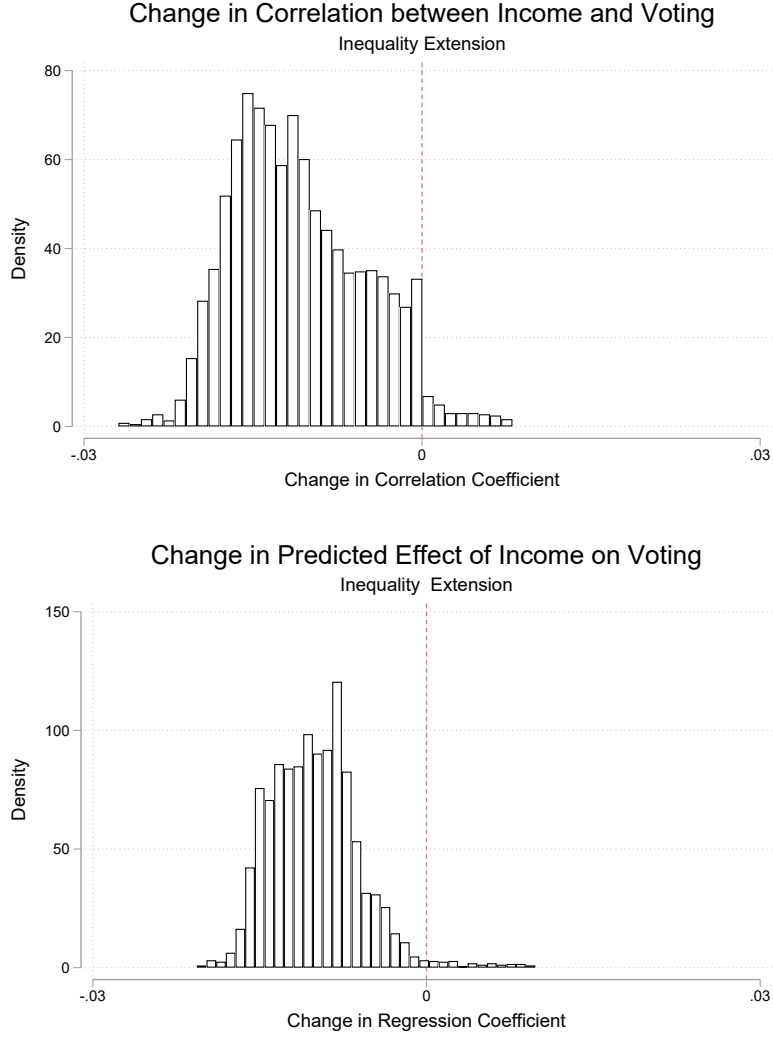
Figure 19: To analyze how the relationship between log income and voting changes over time in the inequality simulations, we calculate two measures of the relationship in the "early" and "late" ($\kappa = 1.015$) periods of the simulation. First, we compute the correlation between log income and an indicator for voting Republican (observations are individual voters with varying levels of log income and log income). Then, we calculate the change in the correlation from the early period to the late period, where positive values mean the correlation between moral values and voting increases over time. The top panel plots the distribution of the change in correlation between log income and voting Republican, where variation in changes comes from different parameters used in the simulation. The second measure we calculate is the coefficient on log income in a linear regression of an indicator for voting Republican on log income. We calculate the change in this regression coefficient from the early period to the late period, and the bottom panel plots the distribution of changes across simulation parameters.
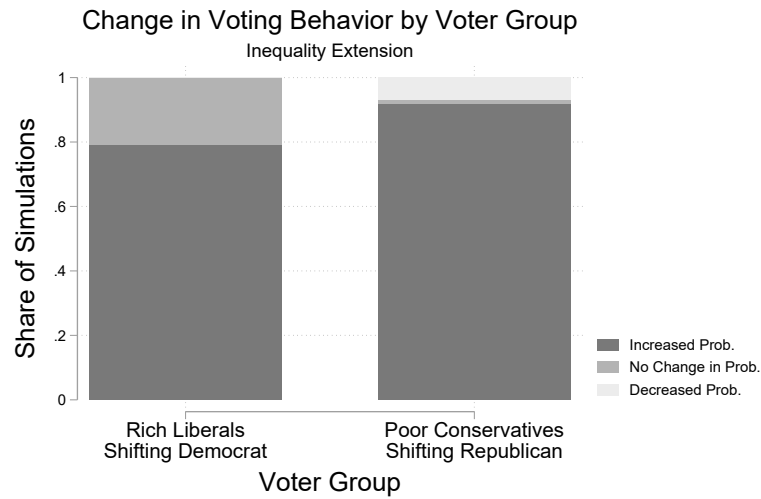
Figure 20: This figure plots the share of inequality robustness simulations in which rich liberals (top 1/3 income and bottom 2/5 moral conservatism) become more likely to vote Democrat and poor conservatives (bottom 1/3 income and top 2/5 moral conservatism) become more likely to vote Republican.
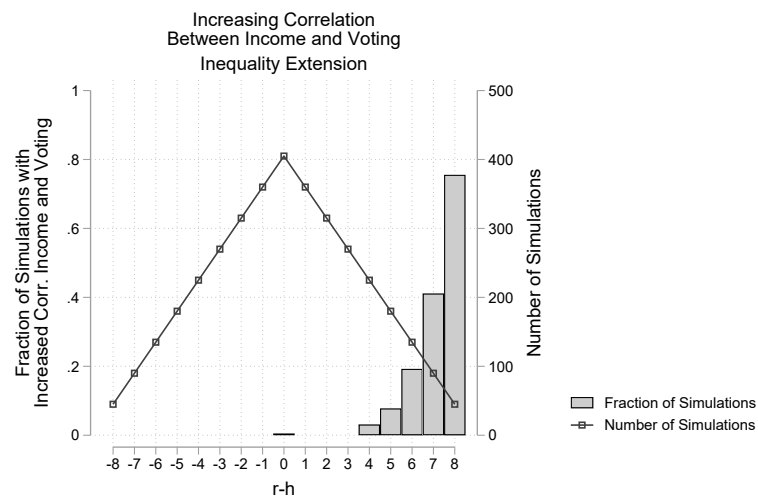

Figure 21: This figure plots the fraction of inequality simulations in which the correlation between log income and an indicator for voting Republican increaes over time, separately for each value of $r_0 - h_0$. The line graph plots the number of simulations with a given value of $r_0 - h_0$.
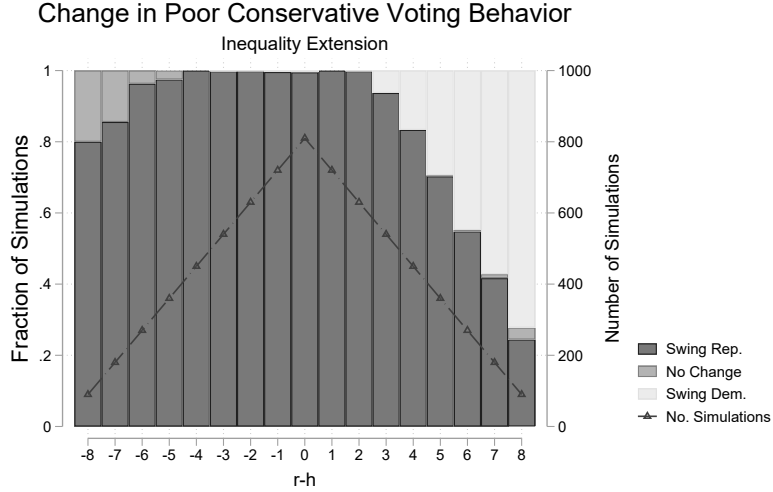
Figure 22: This figure plots the distribution of $r_0 - h_0$, split by whether poor conservatives (bottom 1/3 income, top 2/5 moral conservatism) are more likely, equally as likely, or less likely to vote Republican over time as income inequality increases.

D.3 Robustness to Initial Party Positions

In this main text, we simulate equilibrium party positions starting from the initial positions $(x_s^D, x_e^D) = (-0.5, -0.5)$ and $(x_s^R, x_e^R) = (0.5, 0.5)$. In this section, we show that our simulation results are robust to a range of initial party positions - that is, the simulation delivers the same equilibrium party positions irrespective of the starting points.

To do so, we simulate over a grid of initial positions: $x_s^D \in [-5, -1, -0.5, -0.1, 0, 0.1, 0.5, 1, 5] \times x_e^D \in [-5, -1, -0.5, -0.1, 0, 0.1, 0.5, 1, 5] \times x_s^R \in [-5, -1, -0.5, -0.1, 0, 0.1, 0.5, 1, 5] \times x_e^R \in [-5, -1, -0.5, -0.1, 0, 0.1, 0.5, 1, 5]$ for a total of 6561 simulations. These positions cover a wide range of initial cases for $(x_s^D, x_e^D, x_s^R, x_e^R)$:

1. Parties converge at various points. For example, $(0, 0, 0, 0)$ or $(5, 5, 5, 5)$.

2. Parties polarize symmetrically to varying degrees. For example, $(-0.1, -0.1, 0.1, 0.1)$ or $(-5, -5, 5, 5)$.

3. One party is more extreme than the other on one or both issues. For example, $(-0.5, -0.5, 5, 0)$ implies that the Republicans are extreme on social issues and moderate on economic issues, while the Democrats are somewhat liberal on both.

4. Parties cater to "off-diagonal" voters who are rich and liberal or poor and conservative. For example, $(-1, 1, 1, -1)$. Importantly, these cases show that our main simulation results which deliver one liberal party and one conservative party (on both issues) do not depend on starting from such points.

Note that in cases where the parameter configurations are symmetric (e.g. $(-1, -1, 1, 1)$ and $(1, 1, -1, -1)$), one of the cases can be ignored without loss of generality, as we simply

68

rename the party with larger (i.e. more conservative) positions as the Republican party. More generally, this logic applies to cases where the equilibrium positions chosen by the Democratic party are more conservative than those chosen by the Republican party. We label the party with the more conservative policies as the Republican party, since the relevant result from the simulation is that there exists a party which optimally chooses the positions, not the identity of the party.

To implement this robustness simulation, we require a set of values for the remaining model parameters, $[c, \rho, k, r_0, h_0]$. If we use the full grid from the simulation in the main text, we quickly face a curse of dimensionality since 6,561 initial positions $\times$ 3,645 parameter values generates nearly 24 million simulations. Instead, we select a handful of parameter values which cover an economically meaningful range of voter preferences and run the initial party position robustness with those parameters.

Table 10 summarizes the results. For each set of parameters, we report the mean, standard deviation, maximum, and minimum equilibrium party position across all initial party positions. Regardless of the initial position, the parties always converge very closely to the equilibrium position. This table excludes the 81 initial positions where both parties offer the same policies - in all of these cases (and only in these cases), both parties offer the exact same equilibrium positions as one another.

Table 10: Simulation Robustness to Initial Party Positions

	Parameter Set 1				Parameter Set 2				Parameter Set 3			
Variables	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max
Dem. Social Policy	-0.508	0.000132	-0.508	-0.504	-0.320	0.000373	-0.321	-0.319	-0.625	0.000303	-0.627	-0.622
Dem. Econ Policy	-0.534	0.000120	-0.534	-0.531	-1.455	0.000283	-1.456	-1.455	-3.223	0.000504	-3.225	-3.221
Rep. Social Policy	0.477	0.000186	0.471	0.477	0.176	0.000178	0.175	0.176	0.845	0.000260	0.845	0.846
Rep. Econ Policy	0.447	0.000150	0.440	0.447	0.782	0.000187	0.782	0.783	4.153	0.000627	4.152	4.155

This table presents summary statistics for the equilibrium party positions reached from baseline simulations which are initialized at 6,480 unique starting values. "Parameter Set 1" refers to the model parameters $\beta = 0$, $c = 0.1$, $\rho = 0.2$, $h_0 = 2$, and $r_0 = 1$. "Parameter Set 2" refers to the model parameters $\beta = 0$, $c = 0.1$, $\rho = 0.3$, $h_0 = 1$, and $r_0 = 3$. "Parameter Set 3" refers to the model parameters $\beta = 0$, $c = 0.2$, $\rho = 0.5$, $h = 8$, and $r = 3$.

E Data Description

E.1 American National Election Studies

Economic and social conservatism indices. Since 1988, the ANES has consistently included questions on economic issues such as social security and federal spending on public schools, and on social issues such as gay rights and the importance of family values. We adopt the methodology of Ansolabehere et al. (2006) in aggregating survey questions to construct indices of economic and social conservatism. We classify issues as either economic or moral, then perform factor analyses on each set of questions. The factor analysis allows us to construct a weighted average of the questions, where the weight placed on each question is determined its factor loading. We standardize the factor scores to have mean zero and standard deviation one, and arrange them such that higher values correspond to more conservative issue positions. Across years, our sample spans up to 18,432 respondents. The economic conservatism scale is constructed from survey responses to the following ten items:

1. "Some people think the government should provide fewer services, even in areas such as health and education, in order to reduce spending. (2004: Suppose these people are at one end of a scale, at point 1.) Other people feel that it is important for the government to provide many more services even if it means an increase in spending. (2004: Suppose these people are at the other end, at point 7. And of course, some other people have opinions somewhere in between, at points 2,3,4,5, or 6.) Where would you place yourself on this scale, or haven't you thought much about this?"

2. "There is much concern about the rapid rise in medical and hospital costs. Some (1988,1994-LATER: people) feel there should be a government insurance plan which would cover all medical and hospital expenses (1984 AND LATER: for everyone). (1996,2004: Suppose these people are at one end of a scale, at point 1). Others feel that (1988,1994-1996: all) medical expenses should be paid by individuals, and through private insurance (1984 AND LATER: plans) like Blue Cross (1984-1994: or [1996:some] other company paid plans). (1996,2004: Suppose these people are at the other end, at point 7. And of course, some people have opinions somewhere in between at points 2,3,4,5 or 6.) Where would you place yourself on this scale, or haven't you thought much about this?"

3. "Some people feel that the government in Washington should see to it that every person has a job and a good standard of living. (1972-1978,1996-LATER: Suppose these people are at one end of a scale, at point 1). Others think the government should just let each person get ahead on his/their own. (1972-1978,1996: Suppose these people are at the other end, at point 7. And, of course, some other people have opinions somewhere in between, at pints 2,3,4,5 or 6.). Where would you place yourself on this scale, or haven't you thought much about this?"

4. "Should federal spending on improving and protecting the environment (2000,2002: environmental protection; 2008,2012,2016: protecting the environment) be increased, decreased, or stay the same?"

5. "If you had a say in making up the federal budget this year, for which (1986 AND LATER: of the following) programs would you like to see spending increased and for which would you like to see spending decreased:"

    (a) Child care

    (b) Welfare programs

    (c) Social security

    (d) Public schools

6. "We'd also like to get your feelings about some groups in American society. When I read the name of a group, we'd like you to rate it with what we call a feeling thermometer. Ratings between 50 degrees-100 degrees mean that you feel favorably and warm toward the group; ratings between 0 and 50 degrees mean that you don't feel favorably towards the group and that you don't care too much for that group. If you don't feel particularly warm or cold toward a group you would rate them at 50 degrees. If we come to a group you don't know much about, just tell me and we'll move on to the next one."

    (a) Labor unions

The social conservatism scale is constructed from survey responses to the following nine items:

1. "Do you favor or oppose laws to protect [homosexuals/gays and lesbians] against job discrimination? Do you favor/oppose such laws strongly or not strongly?"

2. "Now, I am going to read several statements. After each I would like you to tell me whether you agree strongly, agree somewhat, neither agree nor disagree, disagree somewhat or disagree strongly with this statement."

    (a) The newer lifestyles are contributing to the breakdown of our society.

    (b) The world is always changing and we should adjust our view of moral behavior to those changes.

    (c) This country would have many fewer problems if there were more emphasis on traditional family ties.

    (d) We should be more tolerant of people who choose to live according to their own moral standards, even if they are very different from our own.

3. "Do you think gay or lesbian couples, in other words, homosexual couples, should be legally permitted to adopt children?"

4. "There has been some discussion about abortion during recent years. (RESPONDENT BOOKLET) Which one of the opinions on this page best agrees with your view? You can just tell me the number of the opinion you choose."

   (a) By law, abortion should never be permitted.

   (b) The law should permit abortion only in case of rape, incest, or when the woman's life is in danger.

   (c) The law should permit abortion for reasons other than rape, incest, or danger to the woman's life, but only after the need for the abortion has been clearly established.

   (d) By law, a woman should always be able to obtain an abortion as a matter of personal choice.

   (e) DK; other

5. "Which of these statements comes closest to describing your feelings about the Bible?"

   (a) The Bible is the actual Word of God and is to be taken literally, word for word

   (b) The Bible is the Word of God but not everything in it should be taken literally, word for word

   (c) The Bible is a book written by men and is not the Word of God

   (d) Other; DK

6. "Do you consider religion to be an important part of your life, or not?"

In our baseline specification, we allow the set of questions used to calculate the indices to vary across years, as some questions are added and removed from the survey over time. Two advantages of this approach are that it (i) incorporates the broadest amount of information possible and (ii) accounts for the fact that some issues which were important in 2020 were not included in the survey in 1988. A concern with this approach is that it skews comparisons across time because liberals and conservatives in 2020 are defined over a different set of questions than their counterparts in 1988. To address this, we construct indices from a factor analysis on a restricted set of issues which are available in all years from 1988 to 2020. The correlation between the baseline and restricted indices is $\rho = 0.95 - 0.97$, and we verify that all of the empirical results are consistent across specifications.

Income. The income variable in the ANES consists of five buckets that partition the population into income percentiles. These buckets are the 0-16th percentile, 17-33rd percentile, 34-67th percentile, 68-95th percentile, and 96-100th percentile. The ANES also provides the

lower and upper bounds of the lower four buckets. To impute (log) income from the buckets, we take the average of the upper and lower bounds for the four lowest buckets. To estimate income in the top bucket, we use data from the Tax Policy Center on average US incomes in the 96th to 99th percentile and the top 1% for each year. We take a weighted average of these incomes to get average income within the 96th-100th percentile group.

E.2   Moral Foundations Questionnaire

See Enke (2020) for a detailed description of the dataset and how the moral conservatism index (relative importance of communal vs. universalist moral values) is computed.