Specialized Learning and Political Polarization

Sevgi Yuksel
New York University

December 24, 2014

For latest version click on
https://files.nyu.edu/sy683/public/jmp.pdf

Abstract

This paper presents a model that demonstrates how developments in information technologies can generate increased political polarization in the absence of any changes in voter preferences. In contrast to the previous literature on the topic, the model abstracts away from possible ideological bias in the news media, and studies how specialization in information by the voters can generate political polarization. When there is heterogeneity among voters in terms of how much they care about different aspects of ideological policies, specialization in information allows for differentiation in learning strategies adopted by the electorate. Specialization allows voters to better respond to variables that influence their ideological positions, but that decreases their overall responsiveness to party platforms. In particular, equilibrium policies polarize more in fractionalized societies where there is greater disagreement about which issues matter the most. When the learning technology allows for specialization in finer subissues, it effectively transforms the society into a more fractionalized one without changing the underlying preferences of the voters, and therefore increases polarization.

JEL Classification Numbers: D72, D80, Z13

Key Words: polarization, ideology, voting behavior, elections, learning

I would like to thank Anmol Bhandari, Micael Castanheira, Catherine Hafer, Dimitri Landa, Jacopo Perego, Howard Rosenthal, Tobias Salz, Séverine Toussaert, and participants at seminars at New York University for helpful comments and suggestions. I would like to thank Guillaume Fréchette and Alessandro Lizzeri for their advice and support. I'm deeply indebted to Debraj Ray for his constant guidance and support at each stage of this project. All remaining errors are mine. Please address all correspondence to sevgi@nyu.edu.

1. INTRODUCTION

Political polarization in the US has traced a U-shaped pattern in the last century.[1] There were high levels of polarization in the early 20th century. This subsided after 1930, but has been increasing in the last thirty years. Polarization now appears to be at an all-time high. On the other hand, Fiorina and Abrams [2008] argues that the distribution of ideological preferences of the voters on a liberal-conservative scale has changed very little and that there is no evidence to indicate that voter preferences ideologically have become more extreme. This has motivated a research agenda focusing on understanding reasons for polarization in party positions beyond changes in the distribution of the ideological preferences of the electorate.

The period in which polarization has increased coincides with a period in which there have also been significant developments in information technologies. The internet dramatically reduces the cost of providing and acquiring information. It allows readers to acquire information from a wide range of sources and easily target the type of information that is most relevant to them. Fifty percent of the American public now cite the internet as a main source for national and international news. This compares dramatically with thirteen percent from 2001.[2] How the shift from traditional news sources to digital information will affect voters engagement in the political process is not yet clear. However, there are many indicators suggesting that online news consumers gather information in fundamentally different ways. They are more skeptical of the integrity of news organizations, rely more on local information and social media, and are more active in when and what kind of information they receive.[3]

In some ways, things are not so different from the local, specialized pamphlets or newspapers of a century ago. As documented by Gentzkow et al. [2014], the end of the 19th and the beginning of the 20th century is also marked by a rapid growth in the number of local newspapers. However, circulation of most of these newspapers declined sharply within 50-100 miles of the publishing center. Thus, the content covered was highly specialized to the local readership. These local newspapers died out in the midcentury as national newspapers, radio and television took their place. As the news market consolidated, specialization became more difficult.

This paper presents a model that demonstrates how changes in technologies that facilitate information acquisition can generate increased political polarization in the absence of any changes in the preferences of the voters. Contrary to the previous literature on the

[1]This is measured as the distance (on a liberal-conservative scale) between the median policy position of a Democrat and that of a Republican. (McCarthy et al. [2006])

[2]This is even higher for younger Americans at 71%. (PEW Research Center)

[3]73% say news organizations spend too much time on unimportant stories and 81% say news organizations are often influenced by powerful people, and 65% say they are politically biased. (Pew Research Center)

topic, the model abstracts away from possible ideological biases in the news media, and focuses on how optimal specialization in the type of information gathered by individual voters, allowed by the new learning technologies, can lead to further political polarization. The main results of the paper build on the following insight. When electoral competition takes place over multiple ideological issues, and voters have uncertainty about their ideal policies on these issues, party platforms in equilibrium depend on the learning strategies adopted by the voters.

A key starting point is that voters care differently about different aspects of the ideological competition taking place between the parties. Some might care more about immigration or education, while others care more about environmental policy or financial regulation. Moreover, any issue can be broken down into smaller subissues. For example, issues relating to gender equality can include gay rights, women's reproductive rights, or discrimination in the workplace. Even on the same issue, voters might differ in terms of which aspects matter more to them. For example, two voters who care equally about the environment might differ in terms of how they weigh the long-term effectiveness of a policy relative to its short-term costs.

The link between the learning strategies and equilibrium policies arises as follows. Parties propose policies on multiple ideological dimensions. Parties are ideologically motivated and receive rents from office. This creates a basic tension. The right party would like to implement right leaning policies conditional on being elected, but has an incentive to propose policies close to the median voter's to increase the probability of election. The way this tension is resolved depends on how responsive voters are to the choice of the platform. As I argue in detail below, that responsiveness depends on the learning strategies adopted by the voters.

I identify several channels through which developments in learning technologies can decrease voter responsiveness to party platforms. When voters are differentiated by how much they care about different aspects of ideological policies, specialization in information allows for differentiation in learning strategies adopted by the electorate. Voters allocate more time to learning about issues they care more about. Specialization in this model takes place not on the broad ideological dimension of left and right as it is often assumed, but across different subdimensions of ideological policies. Specialization allows voters to respond more closely to variables that influence their ideological positions on these issues. However, this decreases their responsiveness to party platforms. This implies that in *more fractionalized* societies, where there is greater disagreement among citizens about which issues matter the most, equilibrium policies polarize more. For instance, consider an environment where ideological policies can be divided generally into two: corresponding to policies on economic and social issues. A society consisting of citizens who care only about economic or social issues, even if there are equal number of each type, would be defined as more fractionalized relative to another soceity, where all citizens care equally about both ideological issues. Observe that more fractionalized societies, while allowing for higher polarization, look, in the aggregate, the same as less fractionalized societies. The distribution of the ideological points of the electorate remains constant, and heterogeneity in the distribution of weights "cancels out" over citizens.

A second issue of interest is greater specialization in learning, or the "fineness" of the learning technology. It turns out that increasing the *depth of learning*, which captures how fine the learning technology is in terms of the number of subissues a voter is able to specialize in, also increases polarization. For an electorate whose preferences remain constant, such refinements in the learning technology creates further differentiation in terms of the actual learning strategies adopted by the electorate. This makes voter preferences more noisy, decreasing their responsiveness to party platforms and consequently increasing polarization. The main result in this respect also draws a connection between specialization of the learning technologies and fractionalization of a society. As the depth of the learning technology increases, a society starts voting in a more fractionalized manner. Although the ideological types of the voters remain the same, changes in the learning technology allow for further differentiation among voters and increase the level of heterogeneity manifested in their voting behavior.

Thirdly, I consider situations where aggregate shocks do not only change the ideological positions of the electorate, but also affect which issues are more important relative to others. For instance, it can be argued that the attacks of 9/11 did not only shift the median voter's position on foreign policy and national security, but made these issues a priority over others. Similarly, the financial crisis of 2008 made differences in the economic policies proposed by the two parties particularly important for the electorate. I study the effects of *salience shocks* defined as events which change voter preferences to make certain issues more important relative to others for the electorate as a whole. When an issue becomes salient, in aggregate, the population shifts attention to this topic, and allocates more time to learning about factors that might shift ideological positions on this issue. Correlation in the learning strategies of the voters generates higher uncertainty with respect to their party preferences. This decreases their responsiveness to party platforms, which consequently leads to higher polarization.

Finally, the learning model is augmented to include a non-partisan policy where voters have identical preferences. Critically, the preferences of the parties are assumed to be different from the electorate: providing welfare improving policies on this issue is costly for the parties. Such costs could be interpreted as research and development costs associated with finding the most effective policies, or political costs associated with choosing a cabinet that might not reflect the party's own agenda. Investigating the interaction of ideological and non-partisan issues reveals that polarization of party platforms on ideological issues can have welfare consequences beyond their direct impact on the implemented ideological policies. Higher polarization on ideological issues makes voters less responsive to higher quality policies on the non-partisan issue, and hence decreases incentives for the parties to invest in better policies on this issue. Hence, changes in learning technologies negatively affect quality of policies on the non-partisan issue directly by increasing aggregate uncertainty and indirectly by generating higher polarization on ideological issues.

The paper is organized as follows. Section 2 discusses the related literature. Section 3 specifies the general model. Section 4 introduces the main forces in a single issue case. Section 5 puts more structure on the model and explores specialized learning when electoral competition takes places on multiple dimensions. Section 6 augments the learning model to include a non-partisan policy and investigates the interaction of ideological and non-partisan policies.

2. Related Literature

This paper relates to a growing literature studying political polarization. Linking political polarization among ideologically motivated parties to uncertainty about the ideal points of the electorate goes back a few decades. In a review article on spatial models of election, Duggan [2005] classifies this as the *stochastic preference model with policy motivation*.[4] Recently, McCarthy et al. [2014] provide empirical evidence that is suggestive of this causal link. Using roll-call voting behavior among state legislators, they show that the ideological distance between Democrats and Republicans from a district is correlated with the ideological heterogeneity of the electorate from that district. They interpret this to be supportive of a model where intra-district ideological polarization causes candidates to be uncertain about the ideological location of the median voter, thereby reducing their incentives for platform moderation.

This paper contributes to a literature where political polarization is linked to voter responsiveness to party platforms. To highlight this, Proposition 1 identifies the direct link between uncertainty of voter preferences and political polarization. I demonstrate that specialization in learning technologies can affect political polarization through this channel. In this sense, the paper is closest to Gul and Pesendorfer [2012] which describes a mechanism through which media concentration reduces political polarization and media competition increases polarization. They also adopt a setting in which parties are ideologically motivated and there is uncertainty about the median voter position. They model competition among profit maximizing news sources, and show that the level of competition in the media market can affect the level of information provided to the voters. A key assumption is that a monopolist media provides no information. As the media market becomes more competitive, news sources ideologically differentiate on the left-right spectrum. In their model, voters are able to learn more about which parties

[4]Hansson and Stuart [1984], Wittman [1983, 1990], Calvert [1985] and Roemer [1994] are some early papers that make this connection.

are closer to their ideal point only when there are news sources with similar ideological bias. Hence, a media market covering a greater spectrum of ideological biases increases the share of informed voters which affects aggregate uncertainty.

This paper focuses on another aspect of specialization. I assume that all voters, regardless of their ideological bias, are able to directly learn about the variables that affect their ideal points. I look at an environment where electoral competition between parties takes place over multiple issues and voters have heterogenous preferences over which issues matter the most. I study how the underlying heterogeneity in the distribution of preferences over these multiple issues affects polarization when agents have the ability to specialize in terms of the issues they are informed about.

In general, papers linking learning patterns of the electorate to political polarization have predominantly focused on the role of media bias where bias can broadly be considered as partial information disclosure or slanted reporting. The literature can be organized as emphasizing two different types of causes for media bias: supply-driven or demand driven. In supply driven models, the media, as an information source has political preferences of its own and hence strategically withholds unfavorable information or follows a biased reporting strategy. Besley and Prat [2006] study an extreme case of this: media captured by the government. Baron [2006] studies media bias resulting from the ideological bias of reporters/editors. Duggan and Martinelli [2010]: present a model of bias resulting from the reduction of multidimensional politics to one-dimensional news story.[5]

Demand driven bias results the incentives of the news sources to pander to its readers' expectations. In the model of Bernhardt et al. [2008], media consumers prefer newspapers that withhold unfavorable information about the party they support. But the media's catering to this preference is socially costly since voters become less informed and elections are less likely to correspond to the efficient outcome. Mullainathan and Shleifer [2005] similarly analyze a model where readers have a preference for news sources that confirm their prior beliefs. Media outlets confront the same trade-off between catering to the readers' priors and providing them with better information. Gentzkow and Shapiro [2006] show that the incentive to conform to the readers' prior expectations can arise endogenously when there is uncertainty about the quality of news sources. Media bias is observed in equilibrium as slanting news increases reputation for the new sources even though it makes all market participants worse off.[6]

[5] In a two-candidate election, they show that media favoring the frontrunner will focus on safe issues unlikely to deliver a surprise, while media favoring the underdog will go for risky issues that can get the underdog elected.

[6] Gentzkow and Shapiro [2010] construct a new index of media slant that measures the similarity of a news outlets language to that of a congressional Republican or Democrat. They show that for US newspapers most of the slant is driven by the demand side.

An exception in this literature is Levy and Razin [2014]. They focus on an environment in which the distribution of voters' ideal points polarize due to *correlation neglect* in learning, defined as the failure to take into account correlation in information sources. They study whether or not this leads to greater policy polarization and find a non-monotonic relationship between polarization of voter preferences and polarization of party platforms. Their main results show that when the electoral system is not too competitive (that is, there is some aggregate noise in the election outcome), presence of voters suffering from correlation neglect can generate lower levels of polarization in party platforms.

This paper also contributes to a literature studying the interaction between ideological issues and non-partisan issues in political competition. Banerjee and Pande [2007], Alesina et al. [1999], Lizzeri and Persico [2001] and Fernandez and Levy [2008] touch on this issue in different ways by investigating how preference heterogeneity affects public good provision. Banerjee and Pande [2007] examines how increased voter ethnicization, defined as greater voter preference for the party representing her ethnic group, affects legislator quality.[7] Alesina et al. [1999] present a model that links heterogeneity of preferences across ethnic groups in a city to the amount and type of public goods the city supplies.[8] [9]

Eyster and Kittsteiner [2007], Groseclose [2001] and Ashworth and de Mesquita [2009] present models which study the interaction between valence competition and party platforms. The closest to our model, Ashworth and de Mesquita [2009] study a game in which candidates first choose platforms and then invest in costly valences (e.g., engage in campaign spending). The marginal return to valence depends on platform polarization: the closer platforms are, the more valence affects the election outcome. Eyster and Kittsteiner [2007] and Groseclose [2001] have models where similar tensions are observed, but they do not directly address the above point.

3. A Model of Multidimensional Platforms

Two parties indexed by $i \in \{L, R\}$ compete in an election. There are K issues. Party platforms consist of a proposed policy vector $\mathbf{y}_i \in \mathbb{R}^K$ on these issues. Let $\mathbf{y} \equiv (\mathbf{y}_R, \mathbf{y}_L)$ denote the party platforms of both parties.

[7]The paper formalizes the intuition that when ethnic groups are of asymmetric size and quality of the candidates is random, the majority party can gets elected even with low quality candidates, and the minority party gets elected only with really high quality candidates. They also provide empirical evidence from a survey on politician corruption in North India that is consistent the theoretical predictions.

[8]In the model, voters first vote on the level of public good provision and then on the type of public good to be provided (which affects who the public good targets). Heterogeneity in preferences on the type of public good to be provided decreases the level that is voted on.

[9]Krasa and Polborn [2013] study how economic positions of the parties are influenced by the cultural positions of candidates and the distribution and intensity of non-economic preferences in the electorate.

There is a continuum of citizens with diverse views on these issues. Each citizen has an initial position $\mathbf{t} \equiv (t^1, ..., t^K) \in \mathbb{R}^K$, with weights $\mathbf{w} \equiv (w^1, ..., w^K) \in \Delta^K$ that reflect how much weight is put on each issue. t^k denotes a citizen's initial position on issue k, and w^k denotes how much weight is put on that issue. I postpone discussion of $(\mathbf{t}, \mathbf{w})$ and it's distribution to the end of this section.

Preferences of the voters. Individual citizens generally prefer policies that are closer to their initial position; that is, given any platform $\mathbf{z} \in \mathbb{R}^K$. The payoff to citizen $(\mathbf{t}, \mathbf{w})$ if platform $\mathbf{z}$ is implemented can be written as:

$$(1) \qquad -\sum_k w^k \ell(t^k + \theta^k - z^k) + \epsilon$$

where $\{\theta^k\}_{k=1,...,K}$ and ϵ are stochastic preference shifters that are described in detail below, and ℓ is a loss function that is symmetric ($\ell(x) = \ell(-x)$) and convex ($\ell'(0) = 0$, $\ell'(x) > 0$, $\ell''(x) > 0$ for $x > 0$).

Voting. I assume that citizens directly receive utility from voting for the party whose proposed policies give them a higher payoff. I, therefore, put aside the issue of why people vote. Indeed, in a model with a continuum of voters, no individual has an impact on the policy outcome. A direct utility from honest voting (perhaps rising from a sense of civic responsibility) is the most straightforward and possibly most realistic assumption in this context.

Preferences of the parties. A party gets elected if it receives more than half the share of all votes. Parties are ideologically motivated. They also receive rents from being in office. The payoff to party i if platform $\mathbf{z}$ is implemented can be written as:

$$(2) \qquad -\frac{1}{K}\sum_k L(t_i^k - z^k) + \mathbf{1}_i\Gamma$$

where $\mathbf{t}_i \in \mathbb{R}^K$ is the vector ideal points for party i, and L its loss function and is allowed to be different from the voters. L is symmetric with $L'(0) = 0$, $L'(x) > 0$ and $L''(x) > 0$ for $x > 0$. Fixing the loss function, the underlying ideological polarization in party preferences can be captured by $\mathbf{t}_R = -\mathbf{t}_L$. $\Gamma > 0$ represents rents from office. $\mathbf{1}_i$ is the indicator function which equals 1 only when party i wins the election.

Note that parties are assumed to put equal weight on each issue, an assumption we will also be making for the electorate as a whole.[10]

[10]This is a natural assumption to make if we take the parties to be representative of their support base. This is not critical for the results, but simplifies exposition.

Shocks to voter preferences. We have already introduced two types of uncertainty that potentially affect a citizen's voting decision: aggregate shocks that shift ideal points, and idiosyncratic shocks that change voter preferences for parties. Think of $\mathbf{t}$ along with $\mathbf{w}$ as the initial type of the citizen. Then, a state of the world $\theta = (\theta^1, ...\theta^K)$, common to all citizens, is realized, resulting in an *ex-post* ideal point for each citizen:

$$\hat{\mathbf{t}} = \mathbf{t} + \theta \tag{3}$$

I assume that each θ^k is independently drawn from a normal distribution with mean 0 and precision ρ^k: $\theta^k \sim \mathcal{N}\left(0, \frac{1}{\rho^k}\right)$.

The state of the world has alternative interpretations with respect to whether or not it is $\hat{t}$ or t which *truly* represents voter preferences. The first would be consistent with situations where θ represents unforeseen changes in the environment that permanently shift preferences. The economic crisis of 2008 or the events of 9/11 would fall into this category. Also, environments in which ex-ante differentiation in ideal points is due to different priors would be consistent with this model. The second interpretation would be consistent with temporary shocks in the perception of the environment that shift voter preferences without changing the welfare consequences of these policies. Here, θ could be interpreted as capturing aggregate uncertainty associated with how effective campaigns are in shifting voter preferences. The backlash to Mitt Romney's remarks about 47% of voters being dependent upon government, or Barack Obama's unsatisfactory performance in defending his agenda in the first presidential debate of the 2012 presidential election would fall into this category.

The second shifter is an idiosyncratic shock that affects a citizen's preference for parties. First, let us record a citizen's *baseline net preference*, for party R, as follows:

$$V_{\mathbf{t},\mathbf{w}}(\mathbf{y}, \theta) = -\sum_k w^k \ell(t^k + \theta^k - y_R^k) + \sum_k w^k \ell(t^k + \theta^k - y_L^k) \tag{4}$$

Whether or not a citizen votes for party R will be given by the sign of $V_{\mathbf{t},\mathbf{w}}(\mathbf{y}, \theta) + \delta$ where δ (to be interpreted as $\epsilon_R - \epsilon_L$) is a shock representing the individual's idiosyncratic preference for party R that is independent of party platforms. Throughout, I maintain the following assumption:

Assumption 1. δ is drawn independently for each voter from a distribution with "large uniform" support on $(-\infty, \infty)$; specifically from an improper uniform distribution on $\mathbb{R}$.

This assumption is not meant to be taken literally; its essence is that large baseline net preferences for parties in either direction can be reversed, albeit with small probability.

It is best to display, upfront, the expositional simplicity we obtain with it, so that the reader can judge the implication for herself.[11]

Observation 1. Suppose that δ is independently drawn across voters from any distribution H. Fix some value of θ, and define $E(\mathbf{y}) = \int V_{\mathbf{t},\mathbf{w}}(\mathbf{y},\theta)dF(\mathbf{t},\mathbf{w})$. Let $W(H, E(\mathbf{y})) \in \{0,1\}$ denote whether or not party R wins the election. For a sequence of uniform distributions H_n with support $[-\bar{\delta}_n, \bar{\delta}_n]$ such that $\bar{\delta}_n \to \infty$,

$$\begin{array}{lll} W(H_n, E(\mathbf{y})) \to 1 & \text{as } n \to \infty & \text{if } E(\mathbf{y}) > 0 \\ W(H_n, E(\mathbf{y})) \to 0 & \text{as } n \to \infty & \text{if } E(\mathbf{y}) < 0 \end{array} \tag{5}$$

Observation 1 tells us that the "sum of aggregate preference differentials", $E(\mathbf{y})$, is a sufficient statistic for predicting the victory of a party, when the distribution of idiosyncratic shocks is "diffuse enough". In short, Assumption 1 together with Observation 1 imply that the right party wins the election if and only if, given the state of the world, the aggregate welfare associated with the right party is higher than with the left party. This allows us to abstract away from the issue of how cardinal preferences are translated into vote shares, and how the curvature of this function affects election outcomes when there is type heterogeneity. This is clearly a very important question on its own, but lies outside of the focus of this paper. The linearity, implicit in Observation 1, eliminates any such forces. The reader is referred to Krishna and Morgan [2012] for an overview on the topic.[12]

Distribution of voter types. Voters are distributed with cdf $F(\mathbf{t},\mathbf{w})$ on $[-T,T]^K \times \Delta^K$.

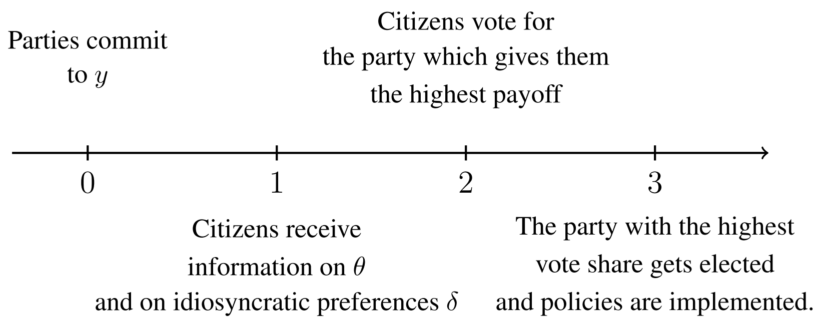
Assumption 2. $F(\mathbf{t},\mathbf{w})$ is smooth and symmetric, i.e. for any $\mathbf{w}$, $f(\mathbf{t},\mathbf{w}) = f(-\mathbf{t},\mathbf{w})$; and for any $\mathbf{w}$, $\int_t f(\mathbf{t},\mathbf{w})dt = \int_t f(\mathbf{t},\mathbf{w}_p)dt$, where $\mathbf{w}_p$ is any permutation of $\mathbf{w}$.

Assumption 2 allows citizens to put different weights on different issues. The symmetry assumptions, on the other hand, state that for any weight vector, type distribution is symmetric with respect to ex-ante ideological positions. In addition, the marginal distribution of the weight vectors is symmetric with respect to permutations. Note that this allows for the distribution of weight vectors to be correlated with ex-ante ideal positions. These assumptions guarantee that on aggregate the electorate puts equal weight on each issue, and the median ideal position ex-ante on each issue is 0.

[11] The model can easily be rewritten, with the cost of additional notation, such that the support of the idiosyncratic term is bounded. This can be done by assuming that the state of the world is drawn from a bounded distribution or by assuming that the support of δ changes with θ.

[12] Krishna and Morgan [2012] study outcomes produced by the majority rule in a setting where the intensity of preferences affects turnout. They show that the majority rule implements the welfare maximizing outcome under voluntary voting.

Timing. I assume that parties do not observe the realization of θ when they announce policy platforms. The citizens, on the other hand, get a chance to learn about θ before they cast their vote. Both the prior on θ and the information structure available to the citizens to learn about θ are common knowledge among the parties and the citizens. The main goal of this paper is to study how changes in this learning stage affect polarization of party positions in equilibrium.



It is important to note that parties are assumed to commit to their proposed policy positions. They are not able to shift policy in response to the state of the world. This friction causes ex-ante uncertainty with respect to vote shares going to different parties. There are certainly many instances where party positions shift in response to aggregate shocks. However, a fully flexible shift in policy positions to respond to the state of the world is invariably difficult. For instance, a party that has called for deregulation of the financial market for years, which has chosen a candidate who symbolizes this position, cannot easily shift gears in response to economic shocks that increase support in the electorate for more regulation. Similarly, a candidate who has vehemently preached an isolationist foreign policy position cannot support military action overseas without losing political capital even when the political environment shifts significantly.[13]Although the model examined here focuses on the full commitment case, any friction in the adjustment of the policies to the state of the world would be enough for the results of the paper.

4. Single issue

To highlight the forces that generate polarization in this context, it is useful to start with $K = 1$. In this section, in addition, I assume that citizens perfectly observe the realization of θ before they cast their vote. Note that $\theta \sim \mathcal{N}(0, \frac{1}{\rho})$. Hence, this section formalizes, in the context of this model, the connection between uncertainty about the

[13]A potential Republican presidential candidate for 2016, Rand Paul is currently criticized heavily for changing his position on US involvement in the Middle East. In June 2014, Paul openly expressed his skepticism toward any plan to launch American air strikes against the Islamic State in a Wall Street Journal op-ed. In September 2014, he accused the Obama administration for not taking a more decisive stand in Time Magazine.

distribution of voter preferences and polarization of party platforms. This connection was first made by Hansson and Stuart [1984], Wittman [1983, 1990], Calvert [1985] and Roemer [1994].

Equilibrium. A profile of party platforms $\mathbf{y}$ constitutes an equilibrium if each party maximizes its expected payoff, given its prior on the state of the world, holding constant the other party's platform. Citizens vote for the party which gives them the highest utility given their ideal points and the idiosyncratic noise.

Proposition 1.
There is a unique symmetric equilibrium.

- **(1):** *If $\Gamma\sqrt{\rho} < L'(t_R)$, parties polarize: the proposed policies are strictly to the right and left of the ex-ante position of the median voter.*
- **(2):** *An increase in uncertainty regarding the state of the world (defined as a decline in ρ), increases polarization.*

All formal proofs (including those for the Observations) are provided in the Appendix. I highlight the main steps and provide some intuition here. The following simple observation pins down the probability of winning for each party, given platforms $\mathbf{y}$.

Observation 2. Given party platforms $\mathbf{y}$, the probability $s(\mathbf{y})$ that the right party gets elected is equal to the probability that the state of the world is higher than a threshold value, that is,

$$s(\mathbf{y}) = 1 - \Phi\left(\sqrt{\rho}\tau(\mathbf{y})\right) \tag{6}$$

where $\Phi(.)$ denotes the cdf of normal distribution with mean 0 and variance 1, $\tau(\mathbf{y})$ defines the threshold ideal point at which a voter would be indifferent between voting for the right and the left party. Given symmetry of the loss function, $\tau(\mathbf{y}) = \frac{y_R + y_L}{2}$.

Observation 2 formalizes the intuition that when citizens are symmetrically distributed in terms of ideological preferences, the party that wins the vote of the median voter also wins the election. However, given party platforms there is remaining uncertainty about which party the median voter supports since the preferences of the median voter shift conditional on the state of the world. Hence, the probability of winning the election is equivalent to the probability that, given the party platforms, the realization of the state of the world will shift median voter preferences sufficiently towards voting for a certain party.

Definition 1. Given party platforms $\mathbf{y}$, let $\pi(\mathbf{y}) \equiv (\pi_R(\mathbf{y}), \pi_L(\mathbf{y}))$ stand for the marginal cost of deviations on each platform in terms of probability of winning. This is referred to as *voter responsiveness to party platforms* (evaluated at $\mathbf{y}$).

Now we look at how voter responsiveness to party platforms change with the level of uncertainty about the state of the world and the level of polarization on the ideological issue.

Observation 3. If party platforms $\mathbf{y}$ are symmetrically polarized on the ideological issue, then for each $i \in \{R, L\}$, $\pi_i(\mathbf{y}) = -\phi(0)\frac{\sqrt{\rho}}{2}$.

Observation 3 demonstrates how voter responsiveness to party platforms depend on the perceived variance of the state of the world $1/\rho$. As ρ increases, voter preferences are less likely to be shifted by θ, and hence the probability of winning changes more directly with variations in policy. Observations 2 and 3 generate a foundation for the tensions we will be exploring in the model.

In a symmetric equilibrium, there are two forces that dictate the positions of the parties. The optimality condition written for the right party, for a platform strictly to the right of the median voter's ex-ante position, reveals these two opposing forces:

$$\pi_R(\mathbf{y})[V(t_R, \mathbf{y}) + \Gamma] + s(\mathbf{y})[L'(y_R - t_R)] = 0 \tag{7}$$

where $V(t_R, \mathbf{y}) = -L(y_R - t_R) + L(y_L - t_R)$ denotes how much the right party prefers the proposed policies of the right party over the left party. Note that in a symmetric equilibrium $s(\mathbf{y}) = \frac{1}{2}$.

Conditional on winning the election, parties prefer policies that are closer to their ideal point. This is captured by the second part of Equation 7. The marginal benefit of a more extreme proposed policy depends on the ideological preferences of the parties (captured by $L'(y_R - t_R)$). However, as seen in the first part of Equation 7, deviating from the equilibrium policy by moving closer to the party's ideal policy decreases the probability with which the party wins the election (captured by $\pi_R(\mathbf{y})$). The force of this effect increases with the level of office motivation (Γ) and polarization in policy platforms ($V(t_R, \mathbf{y})$) but decreases with voter responsiveness to party positions ($\pi_R(\mathbf{y})$ changes with ρ). I show that there can only be a unique symmetric equilibrium where these opposing forces are in balance.

The second part of Proposition 1 follows from the intuition built by the observations above. The state of the world shifts the type distribution to the right or to the left. As the level of uncertainty regarding the state of the world increases, aggregate vote shares become more noisy. This means that the marginal loss of deviating to more extreme policies - in terms of the change in probability of winning the election - declines. This pushes proposed policies closer to the ideal points of the parties, therefore increasing polarization.

One might wonder if this type of effect is general. Namely, if it is the case that increased polarization of the underlying parameters of the model always lead to higher polarization in party platforms. As an example, I consider whether a corresponding shift in polarization of the ideological types of the parties have a similar effect on the proposed policies. A closer look at Equation 7 reveals that a shift in the ideological preferences of the parties, defined as an increase in t_R (where $t_L = -t_R$), can affect equilibrium policies through two channels that have opposing effects on the outcome. The first one can be considered as a *policy distortion effect*. Conditional on winning, for the right party, the marginal benefit of implementing a more right-wing policy increases when party preferences become polarized (captured by $L'(y_R - t_R)$). This pushes policies away from the median voter. On the other hand, there is a *fear of loss effect*. As the ideological positions of the parties diverge, it becomes more costly to lose the election to a left-wing party (captured by $V(t_R, \mathbf{y})$). This increases the marginal benefit of winning the election which pushes policies towards the median voter. Both effects are a consequence of the convexity of the loss function L. It is ambiguous which effect prevails in equilibrium, and this depends on the specification of the loss function. I provide examples for both cases in the Appendix.[14]

5. Learning

A key insight from the model with perfect information was that polarization of party platforms on ideological issues is closely linked to voter responsiveness to party platforms (Proposition 1). When there is greater uncertainty with respect to aggregate voting behavior at the point when parties commit to platforms, the marginal cost of proposing more extreme policies declines, and party platforms polarize.

In this section, I enrich the baseline model such that uncertainty regarding the ideal points of the voters is not completely resolved before voting. Voters optimally pick learning strategies to receive information about the state of world. In light of the examples given in the previous section, learning about the state of the world can be interpreted as reading about the state of the economy (unemployment, growth rates, cost/effectiveness of policies implemented after the financial crises, etc.). Similarly, it can correspond to reading about the state of affairs in the Middle East (cost/effectiveness of US foreign policy, leadership change in regional countries, developments in talks with Iran, potential threats from Al-Qaeda, Isis, etc.). When there is learning, in addition to the forces present in the baseline model, voter responsiveness to party platforms also depends on the learning strategies adopted. The main results highlight how the interaction in heterogeneity of preferences and learning strategies can generate further polarization by decreasing voter responsiveness to party platforms.

[14] It can also be shown that changing the underlying type distribution by taking a mean preserving spread of ideological views (t) has no impact on political polarization.

Model. Note that the state of the world on each dimension, θ^k, is independently drawn from a normal distribution with mean 0 and precision ρ^k: $\theta^k \sim \mathcal{N}(0, \frac{1}{\rho^k})$. I assume that ρ^k is the same on all dimensions. The state of the world is not directly observable to the citizens. Instead, each voter has a unit of time that can be used for learning about the state of the world. When the policy space is multi-dimensional, given $\mathbf{y}$ and beliefs on θ on each dimension, each citizen allocates learning time optimally among the different dimensions. Note that all voters are assumed to have access to the same learning technology and simply solve an attention allocation problem given their budget constraint.[15]

For tractability of the model, we focus on the quadratic loss function.

Assumption 3.

$$\ell(x) = x^2 \quad \textit{and} \quad L(x) = x^2 \tag{8}$$

Assumption 3 allows us to obtain closed form solutions to the optimal learning strategy for each citizen and allows for aggregation over different citizen types.

Learning is modeled in the following way. A learning strategy is a vector η such that $\sum_k \eta^k \leq 1$ where η^k denotes the amount of time allocated to learning about issue k. Voters receive signals on the state of the world; and the precision of the signals directly increase with the time allocated to learning about that issue. Formally:

$$s^k \sim \mathcal{N}\left(\theta^k, \tfrac{1}{\eta^k}\right) \tag{9}$$

Hence, each signal can be considered as $s^k = \theta^k + \epsilon_i^k$ where ϵ_i^k is an idiosyncratic noise drawn independently for each citizen from $\mathcal{N}(0, \frac{1}{\eta^k})$. The next section examines environments where some of the noise associated with the signals is assumed to be correlated across the citizens.

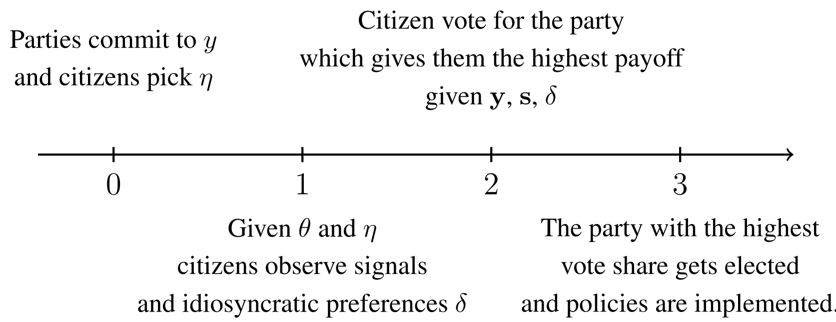
Each citizen chooses a learning strategy to maximize the probability of voting for the party whose proposed policies give them the highest payoff given the realization of the state of the world. Given Assumption 3, this is equivalent to minimizing the variance in beliefs. Hence the maximization problem can be written as follows:

$$\max_{\eta} -\mathrm{E}\left[\mathrm{E}[V_{\mathbf{t},\mathbf{w}}(\mathbf{y},\theta)|\mathbf{s}] - V_{\mathbf{t},\mathbf{w}}(\mathbf{y},\theta)\right]^2 \tag{10}$$

where $V_{\mathbf{t},\mathbf{w}}(\mathbf{y},\theta)$ as defined before denotes the baseline net preference for the right party for a citizen of type $(\mathbf{t},\mathbf{w})$ given the state of the world θ.

Timing. We modify the baseline model in the following way.

[15]Instead of assuming a budget constraint and solving an attention allocation problem, we could have alternatively written a costly information acquisition model. Qualitatively, this wouldn't change the comparative statics identified by the model.



Equilibrium with learning. A profile of party platforms $\mathbf{y} = (\mathbf{y_R}, \mathbf{y_L})$, and a learning strategy η for each type of citizen constitutes an equilibrium if each party maximizes its expected payoff (given the distribution on the state of the world and the learning strategy of the citizens) holding constant the other party's platform; and citizens choose their learning strategy to maximize the probability they vote for the party which gives them the highest utility.

Observation 4. The time allocated to learning about each ideological issue is increasing in the level of polarization on that issue and the weight put on that issue.

Observation 4 states that a citizen, following his optimal learning strategy, concentrates attention on issues that he cares more deeply about and on which there is higher polarization in party platforms. More importantly, the Observation establishes that in a society where there is heterogeneity with respect to which issues are payoff relevant, there will consequently be heterogeneity with respect to the learning strategies adopted.

Observation 5. Given party platforms $\mathbf{y}$ and and the learning strategy of the voters in a symmetric equilibrium, the probability of the right party being elected can be written as:

$$s(\mathbf{y}) = 1 - \Phi\left(\frac{\mu}{\sigma}\right) \tag{11}$$

where $\mu = \frac{1}{K}\sum_k (y_R^k + y_L^k)(y_R^k - y_L^k) = 0$, and $\sigma^2 = \frac{4}{K\rho^k}\left(\int_{\mathbf{t},\mathbf{w}} r(\mathbf{w})(y_R^k - y_L^k)^2 dF(\mathbf{t},\mathbf{w})\right)^2$, where $r(\mathbf{w}) = \frac{\hat{w}}{\rho^k + \hat{\eta}}$ where $\hat{w}$ denotes the highest weight put on any issue in $\mathbf{w}$ and $\hat{\eta}$ denotes the associated time allocated to that issue but such a citizen.

Observation 5 demonstrates the additional forces introduced by learning relative to the baseline model. While aggregate uncertainty with respect to winning probability was an exogenous variable in the baseline model, with learning it becomes an equilibrium object. σ^2 depends on the learning strategies adopted by the voters (given their beliefs about party platforms) and hence on the distribution of types in the electorate.

Proposition 2. *There is a unique symmetric equilibrium:* $(\eta, \mathbf{y})$ *where parties polarize equally on every issue.*

Proposition 2 confirms that the uniqueness of the symmetric equilibrium is preserved with learning. It also establishes that parties polarize equally on every issue. The upcoming sections investigate how changes in the learning technology affect equilibrium policies.

Improving learning technology. Two types of improvements in the learning technology are considered in this section. First assume that the parameter $\alpha \geq 1$ captures developments in the learning technology such that if an agent allocated η of his attention to learning about an issue, he receives a signal of precision $\alpha\eta$. We investigate comparative statics with respect to α. Does polarization on ideological issues go down when the electorate has access to more accurate information?

Secondly, fixing the effectiveness of the leaning technology, we investigate how differentiation in news sources can affect equilibrium polarization levels. So far we had assumed that the noise associated with learning is idiosyncratic noise. Namely, if a voter allocates η^k amount of time to learning about an issue, he receives a signal $s^k = \theta^k + \epsilon_i^k$ where ϵ_i^k is independently drawn for each voter from a normal distribution with mean 0, and precision η^k. If voters are using identical information sources, it would be natural to expect some correlation in the noise term. Note that since voters are only interested in the precision of the signal they receive, any potential correlation in the noise terms of the voters has no impact on their learning strategy. However, it might impact probability of winning the election given the state of the world. Assume that signals can have systematic and idiosyncratic noise such that $s^k = \theta^k + \epsilon_s^k + \epsilon_i^k$, where both ϵ_s and ϵ_i are drawn independently from $\mathcal{N}(\theta, (\sigma_s^k)^2)$ and $\mathcal{N}(\theta, (\sigma_i^k)^2)$ correspondingly. While the first error term (ϵ_s^k) constant for all citizens (can be interpreted as noise associated with the information source), the second one (ϵ_i^k) is individual noise and declines with the time allocated to learning. Such that $(\sigma^k)^2 = (\sigma_s^k)^2 + (\sigma_i^k)^2$.

Proposition 3. *Improvements in the learning technology defined as an increase in* α *increases polarization of proposed policies. Fixing the effectiveness of the learning technology, a decline in the correlation of the error terms across voters (*$(\sigma_s^k)^2$*) (for example, due to diversification of news sources) lowers polarization.*

The first part of Proposition 3 is closely related to Proposition 1 from the single issue model. Increasing the effectiveness of the learning technology allows voters to react better to the state of world, and consequently decreases their responsiveness to party platforms. This leads to higher polarization. The second part is slightly more subtle. When error terms are correlated across the voters, whether or not a party wins the election depends not only on the state of the world, but also on the realization of the common error terms as the overall type distribution of the electorate shift with both parameters. This implies that higher correlation generates higher uncertainty with respect to winning probability which translates into higher polarization in party platforms.

Fractionalized societies. First, I introduce a partial ranking on how specialized voters are.

Definition 2. Let i and j be two different voters with associated weight vectors $\mathbf{w}_i$ and $\mathbf{w}_j$. i is a *more specialized voter* than j if for $\tilde{\mathbf{w}}_i$ and $\tilde{\mathbf{w}}_j$ referring to permutations of $\mathbf{w}_i$ and $\mathbf{w}_j$ such that $\tilde{w}^1 \leq \tilde{w}^2 ... \leq \tilde{w}^K$:

(12) $$\tilde{\mathbf{w}}_i \geq_{FOSD} \tilde{\mathbf{w}}_j$$

According to the definition, voter i is a *more specialized voter* than j if $i's$ weight vector first order stochastically dominates $j's$ when the weight put on different issues are reshuffled such that weights are increasing from issue 1 to K. The definition captures the idea that weights are more concentrated for voter j than i. As an example, consider an election where parties propose policies on two ideological dimensions corresponding generally to economic and social issues. Compare a citizen who puts equal weight on both of these issues to a citizen who conditions his vote solely on the the social issue. The corresponding weight vectors can be represented as $(\frac{1}{2}, \frac{1}{2})$ and $(0, 1)$ where the order of the issues is shuffled such that the second issue corresponds to the social issue for the second citizen. Since $(0, 1) \geq_{FOSD} (\frac{1}{2}, \frac{1}{2})$, the second citizen is defined to be more specialized. Actually, if there are only two issues, $(\frac{1}{2}, \frac{1}{2})$ represents the weight vector for the *least* specialized citizen, and $(0, 1)$, $(1, 0)$ represent the weight vectors for the *most* specialized citizens.

The following lemma demonstrates why comparing voters in terms of how specialized they are is important.

Lemma 1. When all citizens choose their learning strategy optimally, more specialized voters are less responsive to policy platforms.

It is important to note that the separation across voters in responsiveness to policies identified in Lemma 1 is entirely due to learning. When voters are perfectly informed about the state of the world, or when the same learning technology is imposed on all voters, we do not have a separation in responsiveness to policies. The key issue is that more specialized types prefer more specialized learning technologies. Relative to others, they allocate a greater share of their time to learning about fewer issues. This allows them to respond more directly to changes in the state of the world, which in return decreases their responsiveness to policies.

Consider the following partial ranking on how fractionalized the citizens of a society are.

Definition 3. Let S and $\tilde{S}$ be two societies where the type distribution is captured by F and $\tilde{F}$. $\tilde{S}$ is a *more fractionalized society* than S if there exists a mapping g from $\triangle^K \to \triangle^K$ such that for any $\mathbf{w}$, $g(\mathbf{w})$ describes a more specialized voter and $\tilde{F}$ corresponds to the transformed distribution of $(\mathbf{t}, g(\mathbf{w}))$ from F.

Definition 3 extends the previous ranking of specialization across voters to societies in a natural way. A society is identified to be more specialized than another one if it is generated by a transformation of the latter one where each voter is mapped to a more specialized voter.

Proposition 4. *If society $\tilde{S}$ is more fractionalized than society S, there is higher polarization in proposed policies in $\tilde{S}$.*

Proposition 4 highlights how fractionalization can generate further polarization in a society. The result relies on Lemma 1 which states that more specialized types are less responsive to policy platforms. As the population overall gets more fractionalized, the marginal cost (in terms of the decrease in probability of winning) of deviating to more extreme policies declines, while conditional on winning the marginal benefit of implementing policies closer to the party's ideal point remains constant. This force pushes proposed policies in equilibrium away from the median voter.

An interesting observation with regards to Proposition 4 is that more fractionalized societies in aggregate do not look any different. We've already assumed there to be no change in the type distribution with respect to ideal points. In terms of the weights put on different issues, the electorate on aggregate puts equal weight on all issues regardless of how fractionalized it is. More fractionalized societies refer to ones where there is greater disagreement among the voters about which issues matter for the election. Going back to the example where party platforms are two dimensional (consisting of economic and social policies), Proposition 4 implies that a society consisting of citizens who only care about either the economic or social issue (with equal share of both) would generate higher polarization relative to a society where everyone cares equally about both issues.

Increasing the depth of learning. Before considering changes in learning technologies that allow for more specialization, it is useful to discuss how increasing the number of issues would affect equilibrium outcomes. There are multiple interpretations to increasing the number of ideological issues. The first interpretation is about adding new ideological issues to the political agenda that were not previously considered. This would correspond to enlarging the scope of influence for policies. The second interpretation is about breaking up a general issue (such as social issues or economic issues) into smaller subissues (such as education policy, immigration policy, environmental policy etc.) The third interpretation would be about breaking up an ideological issue into different aspects such as costs vs. benefits. Considering for example an environmental policy, there could be uncertainty about its long-term effectiveness as well as short term economic costs.

Increasing the number of issues can have a direct effect on equilibrium polarization levels even without learning though its effect on aggregate uncertainty regarding the median voter's type. Increasing the level of aggregate uncertainty would decrease voter responsiveness to party platforms and lead to further polarization. Whether or not aggregate uncertainty increases or decreases with the number of issues naturally depends on which interpretation above the exercise is meant to capture. Under the different interpretations of the model aggregate uncertainty might remain constant, increase or decrease.

Hence, to focus purely on learning effects, I fix the underlying issue space on which preferences of the voters are defined and do comparative statics with respect to the ability to specialize in information allowed by the learning technology.

Assume that in an environment with perfect information there are at most $\bar{K}$ underlying issues that are payoff relevant for voters. $\bar{K}$ can be arbitrarily large and captures how finely issues can be considered as separate from each other without any restrictions on the party platforms and on the learning strategies of the voters. I define voter preferences directly on this space: $\mathbf{w}_i \in \Delta^{\bar{K}}$ can be interpreted to correspond to the true preferences of a voter. Keeping voter preferences fixed, I study, how the *finess* of the learning technology affects political polarization.

To consider developments in learning technologies that allow voters to be more finely informed about changes in the state of the world for each issue, I define the learning technology to correspond to a partition of this underlying issue space. Let $\mathcal{K} = \{1, ..., \bar{K}\}$ denote the set of underlying issues. Formally, a learning technology $\mathcal{L}$ is a symmetric *partition* of $\mathcal{K}$, which implies the following conditions on $\mathcal{L}$: (1) $\bigcup\limits_{A \in \mathcal{L}} A = \mathcal{K}$; (2) if $A, B \in \mathcal{L}$ such that $A \neq B$ then $A \bigcap B = \emptyset$ and $|A| = |B|$.

Defining learning technologies in this way brings about a natural partial order on the level of specialization allowed by the learning technology.

Definition 4. Learning technology $\tilde{\mathcal{L}}$ allows for more specialization than $\mathcal{L}$ if $\tilde{\mathcal{L}}$ is a finer partition of $\mathcal{K}$ than $\tilde{\mathcal{L}}$, i.e. for any $A \in \mathcal{L}$, $\exists \tilde{A} \in \tilde{\mathcal{L}}$ such that $\tilde{A} \subset A$.

Given the definition of a learning technology, the information structure associated with each learning technology is defined in an intuitive way. Note that each cell of the partition associated with a learning technology corresponds to a subset of issues on $\mathcal{K}$. I assume that each cell provides information that summarizes changes in the state of the world on the subset of issues included in that partition.

Assumption 4. For any $A \in \mathcal{L}$, let $\theta^A = \frac{1}{|A|} \sum\limits_{k \in A} \theta^k$.

Formally, if a citizen chooses to allocate η^A of time to learning about issue A as defined by learning technology $\mathcal{L}$, he receives a signal $s \sim \mathcal{N}\left(\theta^A, \frac{1}{\eta^A}\right)$.

Proposition 5. *If learning technology $\tilde{\mathcal{L}}$ allows for more specialization than $\mathcal{L}$, there is higher polarization of proposed policies with $\tilde{\mathcal{L}}$.*

The intuition behind Proposition 5 is closely linked to the results on specialization in the previous section. A simple example is sufficient to demonstrate this link. Consider switching from a very coarse learning technology where there is only one issue to learn about to another one where the underlying issues space is separated into economic and social issues. With the former learning technology, all voters are identical in terms of their weight vectors, namely they all put full weight on the only ideological issue available for learning as it includes all possible subissues. Moreover, the optimal learning strategy is trivially to allocate all the time to learning about this issue. When the issue space is separated into two, there will be some types who put equal weight on both issues. The corresponding optimal learning strategy for these types will be to divide their time equally between the two issues. Consider a society consisting of only such types as a benchmark. In this society polarization would not be affected by such a change in the learning technology. However, in general, there will other types of voters, who do not equally care about the two subissues. The presence of such types implies that the new society in the two issue case is more fractionalized than the benchmark society described above. In other words, increasing the number of issues amounts to increasing the level of fractionalization in a society. By Proposition 4, this generates higher polarization.

Salience shocks. So far, we've focused on societies where, on aggregate, the electorate cares equally about each issue. This is guaranteed by Assumption 2, which states that for any weight vector $\mathbf{w}$, all permutations of $\mathbf{w}$ are realized with equal probability. More importantly, we have assumed that the underlying type distribution for the electorate is common knowledge among the citizens and the parties. The only uncertainty regarding aggregate voter behavior has been due to shifts in the type distribution though the state of the world θ.

In other words, we've studied environments where unforeseen events can shift the distribution of ideal points, but we have assumed so far that the distribution of weight vectors remains constant. In this section, I relax this assumption. I consider situations where aggregate shocks not only change ideal points, but also change which issues are payoff relevant for the citizens. For instance, it can be argued that the attacks of 9/11, in addition to shifting ideal poitns on foreign policy and national security, made these issues a priority over others. Similarly, the financial crisis of 2008 made differences in the economic policies proposed by the two parties particularly important for the electorate.[16]

[16]Other examples include shift in attention towards environmental policies after extreme weather conditions, or towards gun control after violent crimes.

Assume there is uncertainty with respect to which issues will be salient for the electorate. Consider an exogenous shock on the salience of issues which breaks the symmetry of Assumption 2 by making some weight vectors more likely than others.

Definition 5. Let $g^{k^*} : \triangle^K \to \triangle^K$ be a transformation of the type distribution. g^{k^*} makes issue k^* *more salient in the society* relative to other issues if for any $\mathbf{w}$,$\mathrm{g}^{k^*}(\mathbf{w}) = \mathbf{w}_p$, where $\mathbf{w}_p$ is a permutation of $\mathbf{w}$ in which $w_p^{k^*} \geq w^{k^*}$, and the ranking among other issues remain constant.

Proposition 6. *Assume it is common knowledge that after parties commit to their policies, with probability $\lambda \in [0, 1]$ a randomly selected issue becomes salient in the sense that the type distribution is transformed by g^k. Polarization of proposed policies increase with the probability of a salience shock, λ.*

Proposition 6 states that as the probability of the salience shock increases, polarization of party platforms also increase. Note that when all issues are equally salient, even though voters might specialize individually, on aggregate the electorate divides its attention equally among the different issues. When an issue become salient for the society, voters shift more attention to this issue, and on average put more weight on the information they receive about the state of the world on this issue. Hence, salience shocks make a society more specialized as a whole. This correlates how voter preferences move with the state of the world, decreasing voter responsiveness to party platforms. It should be pointed out that, unlike the previous results in this section, this effect is not entirely due to learning. Even if the state of the world was perfectly observable to all citizens, polarization in proposed policies would increase with the probability of the salience shock. However, with learning, this effect is accentuated further.

6. INTERACTION OF IDEOLOGICAL AND NON-PARTISAN ISSUES

Polarization of party platforms on ideological issues can have welfare consequence beyond their direct impact on the implemented ideological policies. To investigate the interaction of ideological and non-partisan issues, the model presented in the learning section is augmented to include a non partisan policy.

In addition to the ideological issues, party platforms include a non partisan policy. Each party's policy platform is denoted by $(x_i, \mathbf{y}_i) \in \mathrm{R}^{K+1}$ where the first is the proposed policy on the non partisan issue and the second is the proposed policies on the ideological issues as before. Preferences of the voters are adjusted to include the non-partisan issue.

$$g(x_i) + u_{\mathbf{t},\mathbf{w}}(\mathbf{y}_i, \theta) \tag{13}$$

where $u_{\mathbf{t},\mathbf{w}}(\mathbf{y}_i, \theta)$ corresponds to the utility function defined in the previous section and $g(x)$ is a smooth increasing concave function associated the non-partisan policy. While citizens inherently hold opposing preferences on ideological issues, they have common preferences over the non-partisan issue. The non-partisan issue can be interpreted to correspond to the level of public good to be provided that benefits all citizens equally.

Similarly, the preferences of the parties are modified to include the non partisan issue. The payoff for each party if platform $(x, \mathbf{y})$ is implemented can be written as:

$$\tilde{U}_i(x, \mathbf{y}) + \mathbf{1}_i[\Gamma - c(x_i)] \tag{14}$$

where $c(x)$ is a smooth increasing convex function ($c'(0) = 0$) representing additional costs borne by the party in relation to implementing policies on the non-partisan issue and $\tilde{U}_i(x, \mathbf{y})$ captures ideological preferences as defined in the previous section.

Note that even on the non partisan issue, preferences of the parties are not directly aligned with the electorate. Better policies are costly for the parties to implement. Such costs could be interpreted as research and development costs involved with finding the most effective policies or candidates. Alternatively, they could represent political costs associated with pushing for policies or candidates that are not always aligned with the ideology of the party. Note that any issue where party preferences are completely aligned with the preferences of the electorate can be dropped from the model as parties would automatically take on the same position and it will have no impact on winning probability for either party.

Lemma 2. Given party platforms $\mathbf{x}, \mathbf{y}$, let $\pi^x(\mathbf{y})$ and $\pi^y(\mathbf{y})$ stand for the marginal cost of deviations on each platform in terms of win probability. If party platforms are symmetrically polarized on the ideological issue and are the same on the non-partisan issue, for each party:

$$\begin{aligned} \pi_i^x(\mathbf{x}, \mathbf{y}) &= \frac{g'(x)}{\sigma} \\ \pi_i^y(\mathbf{x}, \mathbf{y}) &= -\frac{2y}{K\sigma} \end{aligned} \tag{15}$$

where $\sigma^2 = \frac{4}{K\rho^k}\left(\int_{\mathbf{t},\mathbf{w}} f(\mathbf{t}, \mathbf{w}) r(\mathbf{w})(y_R^k - y_L^k)^2 d\mathbf{t} d\mathbf{w}\right)^2$, with $r(\mathbf{w}) = \frac{\hat{w}}{\rho^k + \hat{\eta}}$ where $\hat{w}$ denotes the highest weight put on any issue in $\mathbf{w}$ and $\hat{\eta}$ is the associated time allocated to learning about this issue.

Lemma 2 demonstrates how voter responsiveness to party platforms depend on σ. As before, as σ decreases, voter preferences are less likely to be shifted by θ, and hence the probability of winning changes more directly with variations in policy. On the other hand, there is a tension between how ideological and non-partisan issues interact. As parties polarize further on the ideological issues, σ increases and voting behavior becomes more correlated with ideological preferences. Hence the marginal benefit of improving policies on the non-partisan issue on winning probability declines.

However, the interaction of the policies on ideological and non partisan issues in joint elections are not restricted to the forces identified by Lemma 2. Note that conditional on winning, each party enjoys implementing ideological policies closer to their ideal point and lower levels of non-partisan policies. Hence equilibrium platforms affect the marginal benefit of winning elections. Higher polarization on the ideological issue generates low levels of non-partisan policies which induces a *fear of loss* effect similar to the one identified in the first section.

Proposition 7. *There is a unique symmetric equilibrium* $(\eta, \mathbf{x}, \mathbf{y})$ *where parties polarize equally on the ideological issues and propose the same policy on the non-partisan issue. Moreover, changes in the learning technology identified in Section 5 that generate higher polarization in ideological policies also lead to lower quality policies on the non-partisan issue.*

Proposition 7 combines the intuition built in the previous sections with the insights from Lemma 2. Changes in learning technologies can lead to further polarization through its impact on aggregate uncertainty. Voters responsiveness to the party platforms decline as aggregate uncertainty with respect to winning probability increase. There is, however, an additional effect when there is a non-partisan issue. Higher polarization on ideological issues in turn makes voters less sensitive to better quality policies on the non-partisan issue, and hence decreases incentives for the parties to invest in better policies on this issue. Hence changes in the learning technologies negatively affect quality of policies on the non-partisan issue directly by increasing aggregate uncertainty and indirectly by generating higher polarization on the ideological issues.

7. Conclusion

The main results in this paper provide a potential explanation for the U-shaped pattern traced by political polarization in the last century. The main motivation of this paper has been to explore the effects of specialization in information allowed by the Internet in the last few decades. However, as mentioned before, it can be argued that the market structure of new sources allowed for a great degree of specialization in the early 20th century as well. This paper demonstrates that the U-shaped pattern can be traced back to the level of specialization in information allowed by the information sources. Importantly, it shows that the specialization of information need not be linked to ideological differentiation, but comes out naturally when electoral competition takes place over a multi-dimensional policy space, and there is heterogeneity among citizens in terms of which issues matter the most. Specialization allows voters to respond more closely to variables that influence their ideological positions on the issues that matter the most to them. This in return decreases their responsiveness to party platforms. The main results of the paper show that when information technologies allow for specialization, equilibrium policies polarize more in fractionalized societies where there is greater disagreement about which issues matter the most. Increasing the depth of learning, which captures how fine the learning technology is in terms of the number of subissues a voter is able to specialize in, also increases polarization by generating a more fractionalized society without changing the underlying preferences of the voters.

APPENDIX

Proof of Observation 1.

Proof. Fix θ and assume that $E(\mathbf{y})$ evaluated at θ is strictly higher than 0. The opposite case can be proven using a symmetric argument. We will solve the case with endogenous learning which includes the complete information case as a special case. Given vector of signals $\mathbf{s}$, for type $\mathbf{t}, \mathbf{w}$, the share of votes going to the right party can be written as $H_n(\mathrm{E}[V_{\mathbf{t},\mathbf{w}}(\mathbf{y},\theta)|\mathbf{s}])$. Integrating over all the signal realizations, and all types, the share of votes going to the right party can be written as $\int\limits_{\mathbf{t},\mathbf{w}} \int\limits_{\mathbf{s}} \phi_{\mathbf{t},\mathbf{w}}(\mathbf{s}|\theta) H_n(\mathrm{E}[V_{\mathbf{t},\mathbf{w}}(\mathbf{y},\theta)|\mathbf{s}]) ds dF(\mathbf{t},\mathbf{w})$ where $\phi_{\mathbf{t},\mathbf{w}}(\mathbf{s}|\theta)$ denotes the probability of observing vector of signals $\mathbf{s}$ given θ which depends on the learning strategy adopted by the citizen of type $\mathbf{t}, \mathbf{w}$. Define $\kappa^n = \int\limits_{\mathbf{t},\mathbf{w}} prob(|\mathrm{E}[V_{\mathbf{t},\mathbf{w}}(\mathbf{y},\theta)|\mathbf{s}]| > \bar{\delta_n}) dF(\mathbf{t},\mathbf{w})$, namely the probability that citizens will receive signals that will push their expected baseline net preference for the right party outside of the $[-\bar{\delta_n}, \bar{\delta_n}]$ range. The share of votes going to the right party is at least:

$$\int\limits_{\mathbf{t},\mathbf{w}} \int\limits_{\mathrm{E}[V_{\mathbf{t},\mathbf{w}}(\mathbf{y},\theta)|\mathbf{s}] \in [-\bar{\delta_n}, \bar{\delta_n}]} \phi_{\mathbf{t},\mathbf{w}}(\mathbf{s}|\theta) \tilde{H}_n(\mathrm{E}[V_{\mathbf{t},\mathbf{w}}(\mathbf{y},\theta)|\mathbf{s}]) d\mathbf{s} dF(\mathbf{t},\mathbf{w}) - \kappa^n$$

Here, to generate a lower bound, we assumed that each citizen votes for left party whenever $|\mathrm{E}[V_{\mathbf{t},\mathbf{w}}(\mathbf{y},\theta)|\mathbf{s}]| > \bar{\delta_n}$. It is sufficient to show that for large n, the sum must be strictly higher than $1/2$. By assumption, the first term converges to $(1-\kappa^n)\frac{\int_{\mathbf{t},\mathbf{w}} \mathrm{E}[\mathrm{E}[V_{\mathbf{t},\mathbf{w}}(\mathbf{y},\theta)|s]] + \delta^n dF(\mathbf{t},\mathbf{w})}{2\bar{\delta}_n} = (1-\kappa^n)\left(\frac{1}{2} + \frac{E(\mathbf{y})}{2\bar{\delta}_n}\right)$. Looking at the second term, we see that for each citizen $prob(|\mathrm{E}[V_{\mathbf{t},\mathbf{w}}(\mathbf{y},\theta)|\mathbf{s}]| > \bar{\delta}_n) < 2\Phi\left(\frac{c(\mathbf{t},\mathbf{w}) - \bar{\delta}_n}{\sigma(\mathbf{t},\mathbf{w})}\right)$ for some constant $c(\mathbf{t},\mathbf{w})$ and $\sigma(\mathbf{t},\mathbf{w})$. This is due to the fact that $\mathrm{E}[V_{\mathbf{t},\mathbf{w}}(\mathbf{y},\theta)|\mathbf{s}]$ is a random variable that is normally distributed. Subtracting half from the addition of the two components, and multiplying by $2\bar{\delta}_n$, we get the following expression for share of votes minus $\frac{1}{2}$: $E(\mathbf{y}) + 2\bar{\delta}_n\left(\frac{3}{2} + \frac{E(\mathbf{y})}{2\delta^n}\right) \int\limits_{\mathbf{t},\mathbf{w}} \Phi\left(\frac{c(\mathbf{t},\mathbf{w}) - \bar{\delta}_n}{\sigma(\mathbf{t},\mathbf{w})}\right) dF(\mathbf{t},\mathbf{w})$. It suffices to note that the second term converges to 0 as $\bar{\delta}_n \to \infty$. ∎

Proof of Observation 2.

Proof. Define $V(t,\mathbf{y}) = -\ell(t-y_R) + \ell(t-y_L)$. This denotes the expected utility of having the right party rule over the left party in a symmetric equilibrium for type t when the party platforms are given by $\mathbf{y}$. Fix $y_R > y_L$. Then there exist an τ such that $|\tau - y_R| = |\tau - y_L|$. By symmetry, $V(\tau + t, \mathbf{y}) = -V(\tau - t, \mathbf{y})$. Hence the share of votes coming from types $\tau - t$ and $\tau + t$ will be equal to $1/2$. Note that if the threshold type is lower (higher) than the median voter, the right (left) party wins. Hence the probability of winning for the right party can be formulated as the probability that the threshold type is lower than the median voter. Since the median voter's type is always equal to θ: $prob(\tau < \theta) = 1 - \Phi(\sqrt{\rho}\mu_0)$. ∎

Proof of Observation 3.

Proof. Note that the assumptions on ℓ imply when (set $y_R = -y_L$):

$$\begin{aligned} V_t(t,\mathbf{y}) &= -\ell'(t-y) + \ell'(t+y) > 0 \\ V_y(t,\mathbf{y}) &= \ell'(t-y) + \ell'(t+y) > 0 \\ V_{ty}(t,\mathbf{y}) &= \ell''(t-y) + \ell''(t-y) > 0 \end{aligned} \tag{16}$$

Also when $t = 0$, $V(0, \mathbf{y}) = 0$, $V_t(0, \mathbf{y}) = 2\ell'(y)$ and $V_{ty}(0, \mathbf{y}) = 2\ell''(y)$ and $V_{y_R}(0, \mathbf{y}) = -l'(y)$ (here we only look at a deviation by the right party).

Note that the threshold type τ solves the following equation.

$$\begin{aligned} V(\tau, \mathbf{y}) &= 0 \\ \left(\tfrac{d}{dy_R}\right) V(\tau, \mathbf{y}) &= 0 \\ \tfrac{\partial \tau}{\partial y_R} &= -\tfrac{V_{y_R}(\tau, \mathbf{y})}{V_t(\tau, \mathbf{y})} \end{aligned} \tag{17}$$

Note that in a symmetric equilibrium $V(\tau, \mathbf{y}) = 0$ which implies that $\tau = 0$. Using this, for a general loss function:

$$\pi_R(\mathbf{y}) = \sqrt{\rho}\phi(0)\tfrac{V_{y_R}(\tau, \mathbf{y})}{V_t(\tau, \mathbf{y})} = -\sqrt{\rho}\phi(0)\tfrac{1}{2} \tag{18}$$

■

Proof of Proposition 1.

Proof. The maximization problem for each party can thus be written as $\max_{y_i} EP(\mathbf{y})$ where

$$EP(\mathbf{y}) = s(\mathbf{y})[\mathcal{U}_i(y_i) + \Gamma] + (1 - s(\mathbf{y}))\mathcal{U}_i(y_{-i}) \tag{19}$$

In equilibrium $\frac{\partial EP(\mathbf{y})}{\partial y_i} \leq 0$, and if $y_i > 0$, the condition holds with equality. In a symmetric equilibrium, for the right party, the optimality condition can be rewritten as:

$$\begin{aligned} \pi_R(\mathbf{y})[\mathcal{U}_R(y_R) - \mathcal{U}_R(y_L) + \Gamma] + s(\mathbf{y})\mathcal{U}_{Ry_R}(y_R) &= 0 \\ \pi_R(\mathbf{y})[V(t_R, \mathbf{y}) + \Gamma] + \tfrac{1}{2}[u_i'(y_R)] &= 0 \end{aligned} \tag{20}$$

where $V(t_R, \mathbf{y}) = \mathcal{U}_R(y_R) - \mathcal{U}_R(y_L)$. Note that while the first term is increasing in the level of polarization, the second term is decreasing. This gives us the uniqueness of the symmetric equilibrium. Now, we focus on identifying conditions under which a positive level of polarization is observed. Set $\mathbf{y} = (0, 0)$, This is an equilibrium if $\pi_R(\mathbf{y})\Gamma \geq \frac{1}{2}[u_i'(0)]$. Note that when $\mathbf{y} = (0, 0)$, $V(t_R, \mathbf{y}) = 0$. Plugging in expressions for $u_i'(0)$ and $\pi_R(\mathbf{y})$ we get, $\Gamma\sqrt{\rho} \geq L'(t_R)$.

Note that a decline in ρ lowers $\pi_R(\mathbf{y})$ which implies (focusing on Equation 20) that the marginal cost (in terms of decline in probability of winning) associated with more extreme policies declines pushing polarization higher. ■

Two examples. Assume parties are purely ideologically motivated ($\Gamma = 0$). And $\rho = 0.01$. Consider two potential loss functions for the parties.

$$L_1(y - t) = -(y - t)^2 \quad \text{and} \quad L_2(t - y) = -\exp^{(y-t)^2} \tag{21}$$

The equilibrium policies when $t_R = -t_L = 1$ are correspondingly $(0.83, -0.83)$ and $(0.57, -0.57)$ under the quadratic and exponential quadratic loss functions. When ideological preferences for the parties polarize such that $t_R = -t_L = 2$, proposed policies polarize for the first, but move towards the middle for the second: $(1.43, -1.43)$; $(0.43, -0.43)$. Hence, the quadratic loss function provides an example where the *policy distortion effect* dominates the *fear of loss effect* when party preferences polarize. This leads to increased polarization on party platforms. On the other hand, the opposite is observed for the exponential quadratic loss function: the *fear of loss effect* prevails and polarization on party platforms decrease.

Proof. Note that parties are assumed to be purely ideologically motivated ($\lambda = 0$ and $\Gamma = 0$). And $\rho = 0.01$. We consider two potential loss functions for the parties.

$$L_1(y-t) = -(y-t)^2 \quad \text{and} \quad L_2(t-y) = -\exp^{(y-t)^2} \tag{22}$$

In a symmetric equilibrium with strictly positive polarization (rewriting Equation 20): $\pi_R(\mathbf{y})[V(t_R, \mathbf{y})] + \frac{1}{2}[u_i'(y_R)] = 0$. Note that $\pi_R(\mathbf{y}) = \sqrt{\rho}/2$.

Case 1: $t_R = -t_L = 1$**:** If $\mathbf{y} = (0.83, -0.83)$: $V_1(t_R, \mathbf{y}) = 4 \times 1 \times 0.83$, and $u_1'(y_R) = 2 \times (1 - 0.83)$. If $\mathbf{y} = (0.57, -0.57)$: $V_2(t_R, \mathbf{y}) = -\exp^{(1-0.57)^2} + \exp^{(1+0.57)^2}$, and $u_2'(y_R) = 2(1-0.57)\exp^{(1-0.57)^2}$.

Case 1: $t_R = -t_L = 2$**:** If $\mathbf{y} = (1.43, -1.43)$: $V_1(t_R, \mathbf{y}) = 4 \times 2 \times 1.43$, and $u_1'(y_R) = 2 \times (2 - 1.43)$. If $\mathbf{y} = (0.43, -0.43)$: $V_2(t_R, \mathbf{y}) = -\exp^{(2-0.43)^2} + \exp^{(2+0.43)^2}$, and $u_2'(y_R) = 2(2-0.43)\exp^{(2-0.43)^2}$.

■

Observation 6. If the ideological preferences of the parties polarize symmetrically (captured by $t = t_R = -t_L$), proposed policies also polarize in equilibrium if the *policy distortion effect* dominates the *fear of loss effect*, i.e. $\sqrt{\rho}\,[L'(t+y) - L'(t-y)] < L''(t-y)$ where y denotes polarization in proposed policy positions in equilibrium. Policies move closer to the middle when the opposite happens.

Proof. Let $\phi = \pi_R(\mathbf{y})[V(t_R, \mathbf{y}) + \Gamma] + \frac{1}{2}[L(y_R - t_R)]$. Note that $\frac{\partial \phi}{\partial y} < 0$. Hence, whether or not the level of polarization in proposed policies increase depends on the sign of $\frac{\partial \phi}{\partial t_R}$.

$$\frac{\partial \phi}{\partial t_R} = (-\sqrt{\rho}[L'(t+y) - L'(t-y)] + L''(t-y)) \tag{23}$$

■

Observation 7. Let $t = t_R = -t_L$ denote polarization in the ideological positions of the parties. For all loss functions such that for all $y \in [0, t]$ $\sqrt{\rho}L'(2t) < L''(y)$, the *policy distortion effect* dominates the *fear of loss effect*.

Proof. Since $L'(0) = 0$, polarization on proposed policies are bounded by polarization in party ideological positions. Hence $L'(t+y) - L'(t-y) < L'(2t)$. ■

Proof of Observation 4.

Proof. Solving the maximization problem for each voter: $\max_\eta -\mathrm{E}\{\mathrm{E}[u_t(\mathbf{y})|\mathbf{s}] - u_t(\mathbf{y}|\theta)\}^2$ is equivalent to solving

$$\max -\textstyle\sum_k 4\frac{(w^k y^k)^2}{\rho^k + \eta^k} \tag{24}$$

subject to $\sum_k \eta^k = 1$ gives us for every $\eta^k, \eta^{k'} > 0$ and for all $\eta^{k''} = 0$:

$$\frac{4y^k w^k}{\rho^k + \eta^k} = \frac{4y^{k'} w^{k'}}{\rho^{k'} + \eta^{k'}} \geq \frac{4y^{k''} w^{k''}}{\rho^{k''}} \tag{25}$$

■

Proof of Observation 5.

Proof.
(26)

$$\begin{aligned} s(\mathbf{y}) &= \int_\theta \Pi_k \phi(\sqrt{\rho_k}\theta_k)\mathbf{1}\left\{\int_{\mathbf{t},\mathbf{w}} f(\mathbf{t},\mathbf{w})\mathrm{E}[V_{\mathbf{t},\mathbf{w}}(\mathbf{y},\theta)|\theta]d\mathbf{t},\mathbf{w} > 0\right\} d\theta \\ &= \int_\theta \Pi_k \phi(\sqrt{\rho_k}\theta_k)\mathbf{1}\left\{\int_{\mathbf{w}} f(\mathbf{w})\mathrm{E}[V_{\mathbf{0},\mathbf{w}}(\mathbf{y},\theta)|\theta]d\mathbf{w} > 0\right\} d\theta \\ &= \int_\theta \Pi_k \phi(\sqrt{\rho_k}\theta_k)\mathbf{1}\left\{\int_{\mathbf{w}} f(\mathbf{w})\sum_k [2r(\mathbf{w})\eta^k(\mathbf{w})\theta^k - w^k(y_R^k + y_L^k)](y_R^k - y_L^k)d\mathbf{w} > 0\right\} \\ &= prob\{\int_{\mathbf{w}} f(\mathbf{w})\sum_k 2r(\mathbf{w})\eta^k(\mathbf{w})(y_R^k + y_L^k)\theta^k d\mathbf{w} > \int_{\mathbf{w}} f(\mathbf{w})\sum_k w^k(y_R^k + y_L^k)(y_R^k - y_L^k)d\mathbf{w}\} \\ &= prob\{\int_{\mathbf{w}} f(\mathbf{w})\sum [2r(\mathbf{w})\tfrac{1}{K}(y_R^k + y_L^k)\theta^k d\mathbf{w} > \tfrac{1}{K}\sum_k (y_R^k + y_L^k)(y_R^k - y_L^k)\} \\ &= prob\{\Delta > \mu\} \end{aligned}$$

Note that in the second line uses the symmetry of $\mathrm{E}[V_{\mathbf{t},\mathbf{w}}(\mathbf{y},\theta)|\theta]$ with respect to t. Third line plugs in for $\mathrm{E}[V_{\mathbf{0},\mathbf{w}}(\mathbf{y},\theta)|\theta]]$. Fifth line uses the symmetry of F with respect to the permutations and assumes that party platforms are assumed to be symmetric on all issues when citizens choose their learning strategy. The last line defines a random variable Δ that is distributed normally with mean 0 and variance $\sigma^2 = \frac{1}{\rho^k}\left(\int_{\mathbf{w}} f(\mathbf{w})\sum_k \frac{2r(\mathbf{w})(y_R^k + y_L^k)}{K}d\mathbf{w}\right)^2$. And similarly, $\mu = \frac{1}{K}\sum_k (y_R^k + y_L^k)(y_R^k - y_L^k)$. Note that $\mu = 0$ and $\sigma^2 = \frac{4y^2}{K^2\rho^k}\left(\int_{\mathbf{w}} f(\mathbf{w})r(\mathbf{w})d\mathbf{w}\right)^2$ in a symmetric equilibrium. ∎

Proof of Proposition 2.

Proof. The proof follows these steps.

Claim 1. In a symmetric equilibrium, parties polarize equally on every issue.

Proof. In equilibrium, given the learning strategies of the voters, for every issue the marginal cost of deviation (in terms of probability of winning) $\pi_R^k(\mathbf{y})[\mathcal{U}_R(y_R) - \mathcal{U}_R(y_L) + \Gamma]$ should be equal to the marginal benefit (in terms of implementing more extreme policies), $s(\mathbf{y})\mathcal{U}_{Ry_R}(y_R)$. Note that $\mathcal{U}_R(y_R) - \mathcal{U}_R(y_L) + \Gamma$ is the same for all issues. The marginal benefit is decreasing in y_R^k, and $|\pi_R^k(\mathbf{y})|$ is constant in $y = y_R = -y_L$ (as shown in the Claim 2 below). Hence, there can only be one level of polarization which equalizes these two forces. ∎

Claim 2. In a symmetric equilibrium, the marginal cost of deviating from the equilibrium policy can be written as: $\pi_R^k(\mathbf{y}) = \frac{\rho^k}{2\bar{r}(\mathbf{w})}$.

Proof.

(27)
$$\begin{aligned} \pi_R^k(\mathbf{y}) &= -\frac{\phi(0)}{\sigma}\left[\frac{\partial\mu}{\partial y_R^k} - \frac{\mu}{\sigma}\frac{\partial\sigma}{\partial y_R^k}\right] \\ &= -\frac{\phi(0)}{\sigma}\frac{2y}{K} \\ &= -\frac{\phi(0)\rho^k}{\left(\int_{\mathbf{w}} f(\mathbf{w})2r(\mathbf{w})d\mathbf{w}\right)} \end{aligned}$$

∎

We used $\mu = 0$ in a symmetric equilibrium in step 2. Claim 2 shows that $s_{y_R^k}(\mathbf{y})$ is constant in in the level of polarization in y. Hence the equilibrium is solved as before and the symmetric equilibrium must be unique. However, with multiple dimensions, we need to check that the second order conditions are also satisfied. More precisely, we check that the Hessian is negative definite at the symmetric equilibrium.

Let $\lambda^k = \pi_R^k(\mathbf{y})[V(t_R, \mathbf{y}) + \Gamma] + \frac{1}{2}[u'_R(y_R^k)]$. The Hessian is a $K \times K$ matrix where:

$$H = \begin{pmatrix} \frac{\partial \lambda_1}{\partial y_R^1} & ... & \frac{\partial \lambda_1}{\partial y_R^K} \\ ... & ... & ... \\ \frac{\partial \lambda_K}{\partial y_R^1} & .... & \frac{\partial \lambda_K}{\partial y_R^K} \end{pmatrix} \tag{28}$$

$$\frac{\partial \lambda_i}{\partial y_R^j} = \frac{\partial \pi_R^k(\mathbf{y})}{\partial y_R^j}[V(t_R, \mathbf{y}) + \Gamma] + \pi_R^k(\mathbf{y}) \frac{\partial V(t_R, \mathbf{y})}{\partial y_R^j} + \frac{\partial \Delta_i}{\partial y_R^j} \tag{29}$$

where $\Delta_i = \frac{1}{2}[u'_R(y_R^k)]$. It is easy to see that the second term is always negative and the third term is negative when $i = j$ and or equal to 0 otherwise. Hence, it is sufficient to show that $\sum\limits_j \frac{\partial \pi_R^k(\mathbf{y})}{\partial y_R^j} \leq 0$. Note that we are using here the fact that in a symmetric equilibrium parties are equally polarized on every issue. To simplify notation let $\mu_{ij} = \frac{\partial}{\partial y_R^j} \frac{\partial \mu}{\partial y_R^i}$ and $\sigma_i = \frac{\partial \sigma}{\partial y_R^i}$

$$\frac{\partial \pi_R^k(\mathbf{y})}{\partial y_R^j} = -\frac{\partial}{\partial y_R^j}\left(\frac{1}{\sigma}\mu_i - \frac{\mu}{\sigma^2}\sigma_i\right) = -\left[\frac{1}{\sigma}\mu_{ij} - \frac{1}{\sigma^2}\mu_i\sigma_j - \frac{1}{\sigma^2}\mu_j\sigma_i\right] \tag{30}$$

Note that $\mu_{ij} = 0$ if $i \neq j$. Otherwise it is equal to $\frac{2}{K}$. It can be shown that $\frac{\sigma_i}{\sigma} = \frac{1}{2Ky}$. Hence,

$$\sum_j \frac{\partial \pi_R^k(\mathbf{y})}{\partial y_R^j} = -\frac{1}{\sigma}\left[\frac{2}{K} - 2\sum_k 2\frac{y}{K}\frac{1}{2Ky}\right] = 0 \tag{31}$$

■

Proof of Proposition 3.

Proof. We look at how $\pi_R^k(\mathbf{y})$ changes with α and σ_s^2. Note that $\pi_R^k(\mathbf{y}) = -\frac{\phi(0)}{\sigma}\frac{2y}{K}$. We investigate the effect of these variables on σ_s^2. Rewriting σ_s^2 to account for these variables:

$$\sigma^2 = \left(\frac{1}{\rho^k} + \sigma_s^2\right)\left(\int_{\mathbf{w}} f(\mathbf{w}) \sum_k \frac{2r(\mathbf{w})\alpha(y_R^k + y_L^k)}{K} d\mathbf{w}\right)^2 \tag{32}$$

Clearly σ^2 is increasing in both α and σ_s^2. ■

Proof of Lemma 1.

Proof. First we prove the following claim which will be critical for the proof.

Claim 3. If $\mathbf{w}_i \geq_{FOSD} \mathbf{w}_j$, then $r_i = \frac{w_i^K}{\rho^K + \eta_i^K} \geq \frac{w_j^K}{\rho^K + \eta_j^K} = r_j$ where η_i and η_j denote the optimal learning vector in a symmetric equilibrium for type i and j.

Proof. Assume for contradiction that $r_i < r_j$. Let $\alpha^k = w_i^k - w_j^k$ and $\beta^k = \eta_i^k - \eta_j^k$. Note that $\sum_k \alpha^k = \sum_k \beta^k = 0$. Since, $r_i < r_j$, it must be that for all k, $\frac{w_j^k + \alpha^k}{\rho^k + \eta_j^k + \beta^k} < \frac{w_j^k}{\rho^k + \eta_j^k + \beta^k}$, otherwise η^K which is guaranteed to be strictly positive would be lowered to increase η^k on another dimension. This means that $\beta^k > \frac{\alpha^k(\rho^k+\eta_j^k)}{w_j^k} = \alpha^k r_j$ summing over all the issues we get $\sum_k \beta^k = 0 > \frac{1}{r_j}\sum_k \alpha^k = 0$ which generates the contradiction. Note that when moving from η_j to η_j, it must be that for each k, $0 \leq \eta_j^k + \beta_j^k \leq 1$. Hence for any α such that there exists an k such that $0 \leq \eta_j^k + \frac{\alpha_j^k}{r_j} \leq 1$ is violated $r_i > r_j$. ∎

Let i be a more specialized voter than j. Without loss of generality assume that $w_i^1 \leq w_i^2.. \leq w_i^K$ and $w_i^j \leq w_j^2.. \leq w_j^K$. Consider a collection of votes whose weight vectors are given by the permutations of $\mathbf{w}_i$. Let $s_i(\mathbf{y})_{y_R^k}$ denote the responsiveness of these voters to deviations in y_R^k. Similarly, $s_j(\mathbf{y})_{y_R^k}$ denotes the responsiveness of voters whose weight vectors are given by permutations of $\mathbf{w}_j$. By Claim 2, $s_{y_R^k}(\mathbf{y}) = -\frac{\phi(0)\rho^k}{\left(\int_{\mathbf{w}} f(\mathbf{w})2r(\mathbf{w})d\mathbf{w}\right)}$. In this case, we are integrating over types who have weight vectors that are permutations of each other. Hence, $s_{y_R^k}(\mathbf{y}) = -\frac{\phi(0)\rho^k}{2r(\mathbf{w})}$. Combining this with Claim 3 gives us the result. ∎

Proof of Proposition 4.

Proof. As before, we will show that in aggregate voters are less responsive to policy platforms in society $\tilde{S}$ relative to society S. Let $s(\mathbf{y})_{y_R^k}$ and $\tilde{s}(\mathbf{y})_{y_R^k}$ correspondingly denote the marginal cost (in terms of probability of winning the election) of deviating on issue k in the two societies. $s(\mathbf{y})_{y_R^k} = -\phi(0)\frac{1}{\sigma}\frac{\partial\mu}{\partial y_R^k}$. Note that $\sigma^2 = \sum_k 4y^2\left(\int_{\mathbf{w}} f(\mathbf{w})\frac{r(\mathbf{w})\eta^k(\mathbf{w})}{\sigma(\tilde{\mathbf{w}})}d\mathbf{w}\right)^2\frac{1}{\rho^k}$ where $r(\mathbf{w}) = \frac{\hat{w}}{\rho^K+\hat{\eta}}$ ($\hat{w}$ refers to the maximal weight put on any issue and $\hat{\eta}$ is how much time is allocated to this issue). By Claim 3, we know that for every every $\mathbf{w}$ in society S is transformed into $\tilde{\mathbf{w}}$ in society $\tilde{S}$ such that $\tilde{\mathbf{w}} = g(\mathbf{w})$, $r(\tilde{\mathbf{w}}) \geq r(\mathbf{w})$ which completes the proof.

∎

Proof of Proposition 5.

Proof. Take any learning technology $\mathcal{L}$. For any $A \in \mathcal{L}$, $\theta^A = \frac{1}{|A|}\sum_{k\in A}\theta^k$. Note that $\theta_A \sim \mathcal{N}(0,(\sigma^A)^2)$ where $(\sigma^A)^2 = \frac{1}{|A|\rho^k}$. Similarly $\theta_{\mathbf{t},\mathbf{w}}^A = \sum_{k\in A} w^k\theta^k \sim \mathcal{N}(0,(\sigma_{\mathbf{t},\mathbf{w}}^A)^2)$ where $(\sigma_{\mathbf{t},\mathbf{w}}^A)^2 = \frac{\sum_{k\in A}(w^k)^2}{\rho^k}$. Defining $s^A = \theta^A + \epsilon$ where $\epsilon \sim \mathcal{N}(1, 1/\eta^A)$.[17] Let $\rho_{\mathbf{t},\mathbf{w}}^A = \frac{\mathrm{E}[s^A\theta_{\mathbf{t},\mathbf{w}}^A]}{\sigma_{s^A}\sigma_{\mathbf{t},\mathbf{w}}^A}$ where $(\sigma^{s^A})^2 = (\sigma^A)^2 + 1/\eta^A$. For type $(\mathbf{t},\mathbf{w})$ the learning problem can be written as:

$$\max_\eta \sum_{A\in L}(\sigma_{\mathbf{t},\mathbf{w}}^A)^2(1-(\rho_{\mathbf{t},\mathbf{w}}^A)^2) \text{ subject to } \sum_{A\in L}\eta_A = 1$$

Note that $\rho_{\mathbf{t},\mathbf{w}}^A = \frac{w^A}{|A|\rho^k\sigma^{s^A}\sigma_{\mathbf{t},\mathbf{w}}^A}$ where $w^A = \sum_{k\in A} w^k$. The problem can be rewritten:

$$(33) \qquad \begin{aligned} \max_\eta \quad & \sum_{A\in L}(\sigma_{\mathbf{t},\mathbf{w}}^A)^2 - \frac{(w^A)^2}{|A|^2(\rho^k)^2(\sigma^{s^A})^2} \\ = \quad & \sum_{A\in L}(\sigma_{\mathbf{t},\mathbf{w}}^A)^2 - \frac{(w^A)^2}{|A|^2(\rho^k)^2\left(\frac{1}{|A|\rho^k}+\frac{1}{\eta^A}\right)} \end{aligned}$$

[17]We will use two properties of the bivariate normal distribution: (1) $\mathrm{E}[X|Y] = \mu_x + \frac{\rho_{xy}\sigma_x}{\sigma_y}(y-\mu_y)$; (2)$\mathrm{Var}[X|Y] = \sigma_x^2(1-\rho_{xy}^2)$ where $\rho_{xy} = \frac{\mathrm{E}[XY]}{\sigma_x\sigma_y}$.

The marginal benefit to increasing η^A can be written as: $\frac{(w^A)^2}{\left(|A|^2(\rho^k)^2\left(\frac{1}{|A|\rho^k}+\frac{1}{\eta^A}\right)\right)^2}\left(\frac{|A|^2(\rho^k)^2}{(\eta^A)^2}\right) = \left(\frac{w^A}{\eta^A+|A|\rho^k}\right)^2$.
This implies that for any A and A' such that $\eta^A, \eta^{A'} > 0$:

$$\frac{w^A}{\eta^A + |A|\rho^k} = \frac{w^{A'}}{\eta^{A'} + |A'|\rho^k}$$

Let $\tilde{\mathcal{L}}$ be a more specialized learning technology than $\mathcal{L}$. The key to the proof is to show that polarization increases for all cases except one where for each agent for any $\tilde{A}_1 \subset A$, $w^{\tilde{A}_1} = \frac{|\tilde{A}_1|w^A}{|A|}$. Note that $\mathrm{E}[\theta^A_{\mathbf{t},\mathbf{w}}|s^A] = \left(\frac{\rho^A_{\mathbf{t},\mathbf{w}}}{(\sigma^{s^A})^2}\right)s = \left(\frac{w^A}{|A|\rho^k\left(\frac{1}{|A|\rho^k}+\frac{1}{\eta^A}\right)}\right)s = \left(\frac{\eta^A w^A}{\eta^A+|A|\rho^k}\right)s$. First consider this benchmark case. We will show that $\eta^{A_1} = \frac{|\tilde{A}_1|\eta^A}{|A|}$ will be the optimal learning strategy for each agent in this case. It follows form $\frac{w^A}{\eta^A+|A|\rho^k} = \frac{\frac{|\tilde{A}_1|}{|A|}w^A}{\frac{|\tilde{A}_1|}{|A|}\eta^A+\frac{|\tilde{A}_1|}{|A|}|A|\rho^k} = \frac{w^{\tilde{A}_1}}{\frac{|\tilde{A}_1|}{|A|}\eta^A+|\tilde{A}_1|\rho^k}$. Hence, in this benchmark case, $r^A = \frac{w^A}{\eta^A+|A|\rho^k}$ for any A such that $\eta^A > 0$ remains constant when we move from learning technology $\mathcal{L}$ to $\tilde{\mathcal{L}}$. For any any agent let $\mathbf{w}_{\mathcal{L}}$ denote weight vector over the partition defined by $\mathcal{L}$. Let $\mathbf{w}_{\tilde{\mathcal{L}}^0}$ denote the weight vector over the partition defined by $\tilde{\mathcal{L}}$ generated as above where for each $\tilde{A}_1 \in \tilde{\mathcal{L}}$ and $A \in \mathcal{L}$ such that $\tilde{A}_1 \subset A$, $w^{\tilde{A}_1}_{\tilde{\mathcal{L}}^0} = \frac{|\tilde{A}_1|w^A_{\mathcal{L}}}{|A|}$. Note that if w^k is not constant for all k for all agents, there will be partition where $\mathbf{w}_{\tilde{\mathcal{L}}} \neq \mathbf{w}_{\tilde{\mathcal{L}}^0}$, which means that the defined weight vector defined over this partition will refer to a more specialized voter relative to $\mathbf{w}_{\tilde{\mathcal{L}}^0}$. Hence, when we move from $\mathcal{L}$ to $\tilde{\mathcal{L}}$, we create a more specialized voter. By Proposition 4, we have the result. ■

Proof of Proposition 6.

Proof. Again, we look at how $s_{y^k_R}(\mathbf{y})$ changes when issue k^* gains salience. There are two key components to this as before: σ^2 and $\frac{\partial\mu}{\partial y^k_R}$. Since all issues have equal probability of becoming salient, $\frac{\partial\mu}{\partial y^k_R}$ remains constant in expectation. Let $\sigma^2_{k^*}$ denote the associated variance when issue k^* gains salience. $\sigma^2_{k^*} = \frac{1}{\rho^k}\sum\limits_k \left(\int_{\mathbf{w}} f(\mathbf{w})2yr(\mathbf{w})\eta^k(g^{k^*}(\mathbf{w}))d\mathbf{w}\right)^2$. There exists an α such that $\sigma^2_{k^*} = \frac{(4y)^2}{\rho^k}\left(\int_{\mathbf{w}} f(\mathbf{w})r(\mathbf{w})\right)^2[(K-1)\left(\frac{1-\alpha}{K}\right)^2 + \left(\frac{1}{K}+\alpha\right)^2]$ where α captures how much attention is shifted to issue k^*. $\sigma^2_{k^*} > \sigma^2$ sufficient for the result.

■

Proof of Lemma 2.

Proof. Given party platforms, preference for the right party for each type can be written as $V(t,\mathbf{x},\mathbf{y}) = g(x_R)-g(x_L)-\ell(t-y_R)+\ell(t+y_L)$. In a symmetric equilibrium as before for $t=0$, $V(0,\mathbf{x},\mathbf{y}) = 2\ell(y)$, $V_t(0,\mathbf{x},\mathbf{y}) = 2\ell'(y)$ and $V_{ty}(0,\mathbf{x},\mathbf{y}) = 2\ell''(y)$ and $V_{y_R}(0,y) = -\ell'(y)$ (here we only look at a deviation by the right party). Note that the threshold type τ solves the following equation.

$$
\begin{array}{rcl}
V(\tau,\mathbf{x},\mathbf{y}) & = & 0\\
\left(\frac{d}{dx_R}\right)[V(\tau(\mathbf{x},\mathbf{y}),\mathbf{x},\mathbf{y})] & = & 0\\
V_t(\tau,\mathbf{x},\mathbf{y}))\left(\frac{\partial\tau}{\partial x}\right) & = & -g'(x)\\
\frac{\partial\tau}{\partial x} & = & -\frac{g'(x)}{V_t(\tau,\mathbf{x},\mathbf{y})}
\end{array} \tag{34}
$$

Similarly,

$$
\begin{array}{rcl}
V(\tau,\mathbf{x},\mathbf{y}) & = & 0\\
\left(\frac{d}{dy_R}\right)[V(\tau(\mathbf{x},\mathbf{y}),\mathbf{x},\mathbf{y})] & = & 0\\
\frac{\partial\tau}{\partial y_R} & = & -\frac{V_{y_R}(\tau,\mathbf{x},\mathbf{y})}{V_t(\tau,\mathbf{x},\mathbf{y})}
\end{array} \tag{35}
$$

Note that in a symmetric equilibrium $V(\tau,\mathbf{x},\mathbf{y})=0$ which implies that $\tau=0$. Using this, for a general loss function:

$$
\begin{array}{rcccl}
s_{x_R}(\mathbf{x},\mathbf{y}) & = & \sqrt{\rho}\phi(0)\frac{g'(x)}{V_t(\tau,y)} & = & \sqrt{\rho}\phi(0)\frac{g'(x)}{2l'(y)}\\
s_{y_R}(\mathbf{x},\mathbf{y}) & = & \sqrt{\rho}\phi(0)\frac{V_{y_R}(\tau,y)}{V_t(\tau,y)} & = & -\sqrt{\rho}\phi(0)\frac{1}{2}
\end{array} \tag{36}
$$

■

Proof of Proposition 7.

Proof. The maximization problem for each party can thus be written as $\max_{x_i,y_i} EP(\mathbf{x},\mathbf{y})$ where

$$EP(\mathbf{x},\mathbf{y}) = s(\mathbf{x},\mathbf{y})[\mathcal{U}_i(y_i)+\Gamma-c(x_i)]+(1-s(\mathbf{x},\mathbf{y}))\mathcal{U}_i(y_{-i}) \tag{37}$$

In equilibrium $\frac{\partial EP(\mathbf{y})}{\partial y_i}0$, and if $y_i>0$, the condition holds with equality and $\frac{\partial EP(\mathbf{y})}{\partial x_i}=0$. Assumptions on $c(x)$ ensure that $x_i>0$. In a symmetric equilibrium, for the right party, the optimality conditions can be rewritten as:

$$
\begin{array}{rcl}
s_{y_R}(\mathbf{x},\mathbf{y})[V(t_R,\mathbf{y})+\Gamma-c(x_R)]+\frac{1}{2}[u_i'(y_R)] & = & 0\\
s_{x_R}(\mathbf{x},\mathbf{y})[V(t_R,\mathbf{y})+\Gamma-c(x_R)]+\frac{1}{2}[-c'(x)] & = & 0
\end{array}
$$

From the proof for the one-dimensional case, we know already know that for a given x a unique y exists that solves the first optimality condition. Note that fixing y, there is a unique x that satisfy the second equation. The first term is decreasing in x, while the second term is increasing in x. Now we focus on the uniqueness of symmetric $(\mathbf{x},\mathbf{y})$ equilibrium. It is clear that y is increasing in x. On the other hand whether or not x increases or decreases with y depends on how $s_{x_R}(\mathbf{x},\mathbf{y})[V(t_R,\mathbf{y})+\Gamma-c(x_R)]$ changes with y. As shown in Lemma 2, $s_{x_R}(\mathbf{x},\mathbf{y})$ is decreasing in y, which implies that x is decreasing in y which guarantees the uniqueness of the symmetric equilibrium.

Let $\Phi=V(t_R,\mathbf{y})+\Gamma-c(x_R)$. There are two possible cases, either $\rho\Phi$ increases or decreases in equilibrium as a result of the changes in the learning technology. First, assume for contradiction that $\rho\Phi$ increases in equilibrium. It is easy to see that if $\rho\Phi$ increases, then y must decrease (as the marginal benefit to deviating towards the median voter increases). This implies that that x must have decreased (since y and ρ have decreased). But this combination implies that $s_{x_R}(\mathbf{x},\mathbf{y})\Phi$ must have increased as well. This would imply a higher x. We have contradiction. Hence $\rho\Phi$ must weakly decrease. This guarantees y to be increasing and x to be decreasing.

■

References

Alberto Alesina, Reza Baqir, and William Easterly. Public goods and ethnic divisions. *The Quarterly Journal of Economics*, 111(4), 1999.

Scott Ashworth and Ethan Dueno de Mesquita. Elections with platform and valence competition. *Games and Economic Behavior*, 67(1):191–216, 2009.

Abhijit Banerjee and Rohini Pande. Parochial politics: Ethnic preferences and politician corruption. *mimeo*, 2007.

David Baron. Persisten media bias. *Journal of Public Economics*, 90:1–36, 2006.

Dan Bernhardt, Stefan Krasa, and Mattias Polborn. Political polarization and the electoral effects of media bias. *Journal of Public Economics*, 2008.

Timothy Besley and Andrea Prat. Handcuff's for the grabbing hand? media capture and government accountability. *American Economic Review*, 96(3), 2006.

Randall Calvert. Robustness of the multi-dimensional voting model: Candidate motivations, uncertainty and convergence. *American Journal of Political Science*, 29, 1985.

John Duggan. A survey of equilibrium analysis in spatial models of elections. 2005.

John Duggan and Cezar Martinelli. A spatial theory of media slant and voter choice. *Working paper*, 2010.

Erik Eyster and Thomas Kittsteiner. Party platforms in electoral competition with heterogeneous constituencies. *Theoretical Economics*, 2:41–70, 2007.

Raquel Fernandez and Gilat Levy. Diversity and redistribution. *Journal of Public Economics*, 2008.

Morris Fiorina and Samuel J. Abrams. Political polarization in the american public. *Annu. Rev. Polit. Sci.*, 11(563-588), 2008.

Matthew Gentzkow and Jesse Shapiro. Media bias and reputation. *Journal of Political Economy*, 2006.

Matthew Gentzkow and Jesse Shapiro. What drives media slant? evidence from u.s. daily newspapers. *Econometrica*, 78(1), 2010.

Matthew Gentzkow, Jesse Shapiro, and Michael Sinkinson. Competition and ideological diversity: Historical evidence from us newspapers. *American Economic Review*, 104 (10), 2014.

Tim Groseclose. A model of candidate location when one candidate has a valence advantage. *American Journal of Political Science*, 45(4):862–886, 2001.

Faruk Gul and Wolfgang Pesendorfer. Media and policy. 2012.

Ingemar Hansson and Charles Stuart. Voting competitions with interested politicians: Platforms do not converge to the preferences of the median voter. *Public Choice*, 44 (3):431–441, 1984.

Stefan Krasa and Mattias Polborn. Social ideology and taxes in a differentiated candidates framework. *American Economic Review*, 2013.

Vijay Krishna and John Morgan. Majority rule and utilitarian welfare. Working paper, 2012.

Gilat Levy and Ronny Razin. Does polarization of opinions lead to polarization of platforms? the case of correlation neglect. Working paper, 2014.

Alessandro Lizzeri and Nicola Persico. The provision of public goods under alternative electoral incentives. *American Economic Review*, 91(1), 2001.

Nolan McCarthy, Keith T. Poole, and Howard Rosenthal. *Polarized America: The dance of ideology and unequal riches*. Mit Press, 2006.

Nolan McCarthy, Jonathan Rodden, Boris Shor, Chris Tausanovitch, and Chris Warshaw. Geography, uncertainty, and polarization. 2014.

Sendhil Mullainathan and Andrei Shleifer. The market for news. *American Economic Review*, 95(4), 2005.

John E. Roemer. A theory of policy differentiation in single issue electoral politics. *Social Choice and Welfare*, 11(4):355–380, 1994.

Donald Wittman. Candidate motivation: A synthesis of alternative theories. *The American Political Science Review*, 142-157, 1983.

Donald Wittman. Spatial strategies when candidates have policy preferences. *Advances in the spatial theory of voting*, (66-98), 1990.