Specialized Learning and Political Polarization

Sevgi Yuksel

July, 2021

Abstract

When citizens are differentiated by how much they care about different issues informing policy, specialization allows them to concentrate their learning on the issues that are most important to them. However, as different citizens focus on different issues, the electorate becomes less responsive to party platforms. In particular, equilibrium policies polarize more in fractionalized societies in which there is greater disagreement about which issues matter the most. When the learning technology allows for more specialization, it effectively transforms the society into a more fractionalized one without changing the underlying preferences of the electorate, thereby increasing polarization.

JEL Classification Numbers: D72, D80, Z13

Keywords: Polarization, Specialization, Voting behavior, Elections, Learning

I would like to thank the editor, anonymous referees, Anmol Bhandari, Micael Castanheira, Catherine Hafer, Dimitri Landa, Jacopo Perego, Howard Rosenthal, Tobias Salz, Jesse Shapiro, Ken Shotts, Séverine Toussaert, and participants at seminars at UC Santa Barbara, USC, UC Davis, NYU, Stanford GSB, WUSTL, HKS, McGill, EUI, TSU, HEC Paris 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 at each stage of this project. All remaining errors are mine. Economics Department, University of California, Santa Barbara, Santa Barbara, CA, 93106, sevgi.yuksel@ucsb.edu.

1 Introduction

Political polarization in the US has traced a U-shaped pattern in the last century.[1] In the early 20th century, there were high levels of polarization which subsided after 1930. However, polarization has increased in the last 30 years and now appears to be at an all-time high. At the same time, there is limited evidence to suggest that the distribution of voter preferences on a liberal-conservative scale has changed significantly over the same time period.[2] This has motivated a research agenda focusing on understanding the 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 were also significant developments in learning technologies. First, the average number of TV channels receivable by the American household increased by more than five times from the mid-1980s to the mid-2000s.[3] Given that TV was Americans' dominant information source for political news in that period, this meant a dramatic increase in the diversity of coverage and perspectives available to citizens. This massive change in the media landscape was followed by the dramatic rise in internet accessibility.[4] How the shift from traditional news sources to digital information affects citizens' engagement in the political process is not fully understood. However, there are many indicators suggesting that online news consumers gather information in fundamentally different ways. They are skeptical of the integrity of news organizations and rely more on news aggregators and social media.[5]

Most importantly, these developments in learning technologies - diversification of news on TV followed by the rise of social media and use of digital news aggregators - allow citizens to receive curated information that is specialized on the individual level. This new media landscape enables citizens to direct their attention to different issues that might be particularly important to them. These issues can be as diverse as immigration, healthcare

[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 a highly debated issue, as changes in the distribution of voter preferences are harder to track. Fiorina and Abrams (2008) and Abramowitz and Saunders (2008) provide arguments on either side.

[3]Nielson report on media. (2004)

[4]While TV continues to play an important role in how people acquire political news, the internet is becoming a primary information source for political news for many. Fifty percent of Americans in 2013 named the internet as a main source for news, compared to 13 percent in 2001. (PEW Research Center).

[5]As of spring 2017, two thirds of Americans stated that they get at least some news on social media.

and foreign policy.

In some ways, things are not so different from the local, specialized pamphlets or newspapers of a century ago. As Gentzkow et al. (2014) document, the end of the 19th and the beginning of the 20th centuries were 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. By the mid-20th century, these local newspapers had died out, as national newspapers, radio and television took their place. As the news market consolidated, specialization became more difficult.

This paper presents a simple model to illustrate how optimal specialization in the type of information gathered by individual citizens—allowed by new learning technologies—can increase political polarization in the absence of any changes in the preferences of the citizens. The paper builds on an old insight in the political economy literature on how policy polarization among ideologically motivated parties could result from uncertainty about the distribution of voter preferences.[6] Such models are built on the following tension: while ideologically motivated parties prefer to implement policies that are more extreme than the ideal position of the representative citizen conditional on being elected, parties still have incentives to moderate their policies to increase the probability of election. The way that this tension is resolved depends on how responsive the electorate is to the choice of the platform. Responsiveness here captures the marginal cost for the parties, in terms of the probability of winning the election, of moving to more extreme parties.

This paper builds on this insight by endogenizing the uncertainty over the distribution of voter preferences using a multi-dimensional framework designed to study how changes in the learning stage can affect polarization. In the model, voters are uncertain about their ideal position on a policy, but they have access to a learning technology that allows

[6]Duggan (2005) classifies this as the *stochastic preference model with policy motivation*. Hansson and Stuart (1984), Wittman (1983, 1990), Calvert (1985) and Roemer (1994) are some early papers that make this connection. In a recent paper, McCarty et al. (2018) 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 in which intra-district ideological polarization causes candidates to be uncertain about the ideological location of the median voter, thereby reducing their incentives for platform moderation.

them become informed about different issues that inform their ideal position. The model demonstrates how responsiveness to party platforms depends on the characteristics of the learning technology. Specialization enables citizens to learn about their ideal positions more effectively by allowing them to direct more attention to the issues that are more important to them. However, this also means that the underlying heterogeneity in preferences (generating disagreement about ideal policy) is more closely reflected in voting patterns. This decreases the electorate's responsiveness to party platforms, which, in turn, increases polarization.

A key starting point for the model is that citizens care differently about different aspects of policy. Some might care more about healthcare, while others care more about the environment or taxes. Moreover, any issue can be broken down into smaller subissues—for example, those related to gender equality can include women's reproductive rights and discrimination in the workplace. Even on the same issue, citizens might differ in terms of which aspects matter more to them. For example, two people 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.

When citizens have diverse views on the importance of different issues, and there is uncertainty about the ideal policy response to these issues, party platforms in equilibrium depend on the learning strategies adopted by the citizens. This paper identifies several channels through which characteristics of the learning technology interact with the distribution of preferences in the electorate to impact responsiveness to policies. In *more fractionalized* societies, where there is greater disagreement among citizens about which issues matter most, equilibrium policies become more polarized. For example, take a society consisting mostly of citizens who care only about either economic or social issues. Even if there were an equal number of each type, the society would be defined as more fractionalized relative to one in which most citizens care equally about both issues. Fractionalization would increase even further if, for example, those citizens who cared mostly about economic issues were divided in terms of precisely which economic issues (redistribution, growth, globalization, etc.) are more important. Observe that more fractionalized societies, while allowing for higher polarization, look, in aggregate, the same as less fractionalized societies, putting equal weight on all issues. The level of fractionalization in the society captures the degree to which the distribution of preferences in the society differs from that of the representative citizen. Note, however, that this is not heterogeneity in the traditional sense of left-right bias, but, rather, heterogeneity that arises from different citizens putting different weights on different issues in a multi-dimensional framework. Heterogeneity in the preferences of citizens implies heterogeneity in response to policy platforms, which decreases overall responsiveness to policies.

A second issue of interest is greater specialization in learning, or the “fineness” of the learning technology. Increasing the *depth of learning*, which captures how fine the learning technology is in terms of the subissues in which a citizen is able to specialize, also increases polarization. Note that specialization in this model takes place not on the broad ideological dimension of left and right, as is often assumed, but across different subissues that can inform policy. The paper builds on the following insight: when citizens are uncertain about their position on a policy, the degree to which the underlying heterogeneity in their preferences is reflected in their voting decisions depends on the level of specialization allowed by the learning technology. As the depth of learning increases, citizens become better at learning how their ideal position differs from that of the representative citizen. An implication of this is that voting patterns for such citizens increasingly differ from that of a representative citizen, decreasing the marginal cost associated with parties moving to more extreme policies. The main result, in this respect, also draws a connection between the specialization allowed by learning technologies and the fractionalization of a society. As the depth of the learning technology increases, a society operates in a more fractionalized manner. Although the distribution of citizens remains the same, changes in the learning technology increase the level of fractionalization manifested in how they respond to policies.

To highlight how a more informed electorate could restrain political polarization, an alternative change to learning technologies that allow for more *access* to information is considered. Keeping the level of specialization constant, with greater access, the assumption is that citizens are more likely to receive information that could help them decide which party to support. On the individual level, this has the same effect as a greater ability to specialize: the probability that each citizen supports the party that is closer to her ideal position increases. However, in contrast to specialization, this has the impact of increasing citizens’ responsiveness to policies, which decreases polarization.

Finally, the paper highlights the negative welfare implications of increased polarization. In equilibrium, parties polarize around the ideal policy of the representative citizen. Symmetry with respect to the distribution of preferences indicates that the representative citizen ’s ideal policy is the socially optimal policy. Hence, the level of polarization observed in equilibrium translates into a measure of welfare loss relative to the socially optimal policy.

This paper is motivated by the larger public debate on how the new information landscape—governed by a variety of TV channels, social media and digital news—reshapes the political process. These information sources have dramatically reduced the cost of providing and acquiring information, but they have also endowed citizens with remarkable tools to receive specialized information. The goal of this paper is to demonstrate that while both channels (more access to information and specialization) enable citizens to learn more effectively, they can have radically different consequences for political outcomes.

Empirical evidence suggests that the specialization channel might be more important for understanding current trends. A vast majority of Americans believe that the internet allows them to be better informed on the topics that matter to them, with 75% of internet users stating that they are more informed on national news relative to five years ago. (PEW Research Center) However, studies looking at voter informedness do not find any significant improvements (Somin (2016)). Empirical evidence on how specialized voters are in their learning patterns has only recently emerged. New papers using machine learning techniques to analyze textual data find evidence of content specialization among information sources (Angelucci et al. (2020); Nimark and Pitschner (2019); Martin and Yurukoglu (2017)).

As noted before, this model builds on the insights developed by the work of Hansson and Stuart (1984), Wittman (1983, 1990), Calvert (1985) and Roemer (1994), who first made a connection between uncertainty about the distribution of voter preferences and polarization of policies proposed by ideologically motivated candidates.[7] Other papers studying political polarization have focused predominantly on the role of media bias, with bias broadly considered as partial information disclosure or slanted reporting.[8] Levy and Razin (2015) provides a model 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. Other

[7]A separate literature studies how aggregate uncertainty about the distribution of preferences moves the predictions of the rational-voter theory closer to empirical observations: See Good and Mayer (1975), Castanheira (2003) and Myatt (2012) on voter turnout, and Myatt (2007), Dewan and Myatt (2007) and Bouton and Castanheira (2012) on stable non-Duverger's Law equilibria.

[8]See Besley and Prat (2006), Duggan and Martinelli (2011), Gul and Pesendorfer (2012) for prominent examples. Bernhardt et al. (2008), Mullainathan and Shleifer (2005), and Gentzkow and Shapiro (2006) study endogenously supplied information and its impact on electoral outcomes when there is demand for biased news. Gentzkow et al. (2015) provides a recent review of the theoretical literature.

explanations include specialized candidates (Krasa and Polborn (2010, 2012)), interaction between valence and ideology (Polborn and Snyder Jr (2017), Eyster and Kittsteiner (2007), Groseclose (2001) and Ashworth and de Mesquita (2009)), informational asymmetries between voters and parties (Martinelli (2001)), and political entry deterrence when there is preference heterogeneity (Callander (2005)). This paper also contributes to a literature studying how preference heterogeneity affects public good provision.[9] Relatedly, Matějka and Tabellini (2019) studies policy choice when voters are rationally inattentive. Complementing the results in this paper, they find that divisive issues attract the most attention by voters and that this can create inefficiency in public good provision.

This paper is most closely related Perego and Yuksel (2021), which studies the competitive provision of political information. That paper's critical insight is that competition forces information providers to specialize by becoming relatively less informative on issues that are important from a social point of view, thus amplifying social disagreement. The focus on specialization of information in a framework with multi-dimensional preferences links the the two papers. However, while Perego and Yuksel (2021) studies competition among information providers, this paper takes changes in the learning environment as given, focusing on the policy choices of the parties.

The paper is organized as follows. Section 2 presents the model and discusses its key features. Section 3 defines and solves for the equilibrium. Section 4 presents the main results, and Section 5 concludes.

2 Model

Two parties indexed by $i \in \{L, R\}$ compete in an election. Each party commits to a policy $y_i \in \mathbb{R}$ that will be implemented upon election. Let $y := (y_R, y_L)$ denote the policy choices of both parties. There are $\mathcal{K} := \{1, ..., K\}$ issues with $K \geq 2$ that could influence a citizen's position on this policy. There is a continuum of citizens with diverse views on how different issues should be considered to form a position. Each citizen is characterized by a vector $w := (w^1, ..., w^K) \in \Delta^K$, which denotes how much weight she puts on each issue in evaluating her position. $\theta := (\theta^1, ..., \theta^K) \in \mathbb{R}^K$ represents the state of the the world and is

[9] See Banerjee and Pande (2007), Alesina et al. (1999), Lizzeri and Persico (2001) and Fernández and Levy (2008).

common among all citizens. Given θ, $w \cdot \theta = \sum_{k \in \mathcal{K}} w^k \theta^k$ describes the *ideal position* on the policy for a citizen of type w. For each k, $\theta^k \in \mathbb{R}$ is independently drawn from a normal distribution with mean 0 and precision ρ.

Distribution of citizen types:

Citizens are distributed with cdf $F(w)$ over the set of weights Δ^K.

Assumption 1. $F(w)$ is smooth and symmetric in aggregate—i.e., $w_m := \int_{\Delta^K} w \, dF(w) = (\frac{1}{K}, ..., \frac{1}{K})$.

The assumed symmetry guarantees that the electorate puts equal weight on each issue in aggregate, but marginal distributions can vary with k.[10] The policy position associated with the representative citizen is denoted as $y_m := w_m \cdot \theta$.[11]

Preferences of the citizens:

Individual citizens prefer policies that are closer to their ideal position. That is, conditional on θ, the payoff to citizen of type w if policy y_i is implemented can be written as:

$$-(w \cdot \theta - y_i)^2. \tag{1}$$

Each individual takes a binary action $a \in \{0, 1\}$, which could be interpreted as voting, to show support for either party R ($a = 1$) or party L ($a = 0$). Citizens are assumed to receive utility from expressing their preferences. Since no individual has an impact on the policy outcome, a direct utility associated with expressing one's preferences (perhaps rising from a sense of civic responsibility) is the most straightforward and possibly realistic assumption in this context.[12] That is, for a citizen of type w, the net utility associated with supporting party R can be written as:

$$u_w(\theta \,|\, y) := -(w \cdot \theta - y_R)^2 + (w \cdot \theta - y_L)^2. \tag{2}$$

[10]For example, suppose $K = 3$, and that half of agents care only about the first and third issue, i.e., $w = (2/3, 0, 1/3)$, and half of agents care only about the second and third issue, i.e., $w = (0, 2/3, 1/3)$. On average, the society puts equal weight on all issues: $w_m = (1/3, 1/3, 1/3)$. While all agents put the same weight on the third issue, there is heterogeneity in weights associated with the first two issues.

[11]Equivalence of y_m with the median position in the society requires stronger symmetry assumptions.

[12]See Eraslan and Ozerturk (2018), Stromberg (2004), Gentzkow and Shapiro (2006), and Baron (2006) for discussions on how political news can have instrumental value for citizens beyond voting decisions, potentially informing financial, educational, or labor supply decisions.

Note that, by symmetry, the net utility associated with supporting party L is $-u_w(\theta \mid y)$. Citizens who are indifferent between the two parties are assumed to randomize their action, supporting each party with equal probability.

Learning:

A citizen's decision to support either party depends on her beliefs about the ideal policy and y. Citizens observe y and actively engage in learning about their own ideal policy position.

All citizens are faced with the same learning technology $\mathcal{L} := (\alpha, \mathcal{P})$, which imposes restrictions on their ability to collect information about θ. The learning technology is characterized by an *access* parameter $\alpha \in [0, 1]$ and a *partition* $\mathcal{P}$.[13]

The access parameter α determines whether or not a citizen is able to gather information about θ. Conditional on being able to gather information (which happens with probability α determined independently across citizens), each citizen observes the realization of a costless signal, which is a weighted average of the multi-dimensional state: $\ell \cdot \theta$. A citizen's learning strategy is to choose these weights ℓ (which might be different from the citizen's type w). The *partition* of the learning technology imposes constraints on the learning strategies available to the citizens. Formally, $\ell \in \Delta_{\mathcal{P}}$, where

$$\Delta_{\mathcal{P}} := \{\ell \in \Delta^K \mid \ell^k = \ell^{k'} \text{ for all } k, k' \in A \text{ for all } A \in \mathcal{P}\}.$$

$\Delta_{\mathcal{P}}$ states that the set of learning strategies available to citizens is constrained by partition $\mathcal{P}$. A citizen can observe a weighted average of the multi-dimensional state in which the weights are constant within each cell of the partition. Note that each cell of the partition corresponds to a subset of issues in $\mathcal{K}$. Thus, the learning technology allows a citizen to differentiate her learning (by choosing different weights) between issues that lie in separate cells of the partition, but constrains her to learn the same amount (by imposing the same weights) on issues that lie in the same cell of the partition.[14] In this way, $\mathcal{P}$ limits the degree to which citizens can differentially get informed on separate issues that inform their ideal

[13]Formally, $\mathcal{P}$ is a partition of $\mathcal{K}$ with $|\mathcal{P}| \geq 2$, which implies the following conditions on $\mathcal{P}$: (1) $\bigcup_{A \in \mathcal{P}} A = \mathcal{K}$; (2) if $A, B \in \mathcal{P}$ such that $A \neq B$, then $A \bigcap B = \emptyset$.

[14]For example, consider a partition $\mathcal{P} = \{\{1, 2\}, \{3, 4\}\}$ where $K = 4$. The learning technology allows for more or less learning about the first two issues relative to the next two (setting ℓ^1 and ℓ^2 higher or lower than ℓ^3 and ℓ^4), but constrains learning to be identical between the first two and the next two issues ($\ell^1 = \ell^2$ and $\ell^3 = \ell^4$).

position. It is in this sense that the partition $\mathcal{P}$ will later be described as capturing the degree to which a citizen can be specialized in her learning.

Each citizen chooses a learning strategy ℓ (which issues to gather information on) to maximize the expected utility associated with supporting a party.

Preferences of the parties:

Before committing to policies, parties learn about the position of the representative citizen: y_m. Both the prior on θ and the learning technology available to the citizens to gather information about θ are common knowledge among the parties and the citizens. $s_R(y \,|\, \theta) := \int_w z_w(y|\theta)dF(w)$ denotes the support (vote share) that party R receives conditional on θ, where $z_w(y|\theta)$ describes whether a citizen of type w supports party R given policy choices y and state of the world θ. Note that $z_w(y|\theta)$ depends on the citizen's optimal learning strategy ℓ and the probability of receiving a signal α. $s_L(y \,|\, \theta)$ is defined naturally as $1 - s_R(y \,|\, \theta)$. Since parties commit to policies before learning the realization of θ, $s_i(y) = \mathbb{E}_\theta[s_i(y \,|\, \theta)]$ will be used to denote the unconditional support (vote share) for party i

In addition to the voters whose behavior has been described above, we assume that there is an equal measure of non-policy voters and, among these voters, the share voting for party R is distributed uniformly.[15] Under this assumption, each party's probability of winning, i.e. the probability of receiving an aggregate vote share greater than half, is equal to $s_i(y)$ (Remark 1 in the Appendix).

Parties are ideologically motivated.[16] That is, $y_i^\star := y_m + b_i$, the preferred policy for party i, is assumed to be biased either towards the right or left relative to the position of

[15] We use non-policy voters to establish equivalence of vote share among policy voters and probability of winning for the parties. The stark assumptions made for this equivalence are not necessary but are adopted to simplify the equilibrium analysis. An extensive discussion of the conditions under which maximizing vote share is equivalent to maximizing the probability of winning is provided by Banks and Duggan (2004); Patty (2005, 2007); McKelvey and Patty (2006). In general, the probability that a policy is implemented or that a candidate is elected can be a nonlinear function of the behavior of individual agents within a society. Main results of the paper could be extended to such settings at the cost of complicating the model further.

[16] The assumption of ideological motivation for parties was first used introduced in Wittman (1973) and since then has been used in many applications. The assumption captures the idea that political parties are instruments of various economic classes or, more generally, interest groups. While ideological motivation is a key force driving polarization in our model, allowing also for office motivation for the parties would not change the main comparative results.

the representative citizen.[17] The underlying ideological bias in party preferences is assumed to be symmetric—that is, $b_R = -b_L > 0$. As with citizens, parties prefer policies that are closer to their ideal position. That is, conditional on θ, the payoff to party i if policy y is implemented can be written as:

$$-(y_i^\star - y)^2 \tag{3}$$

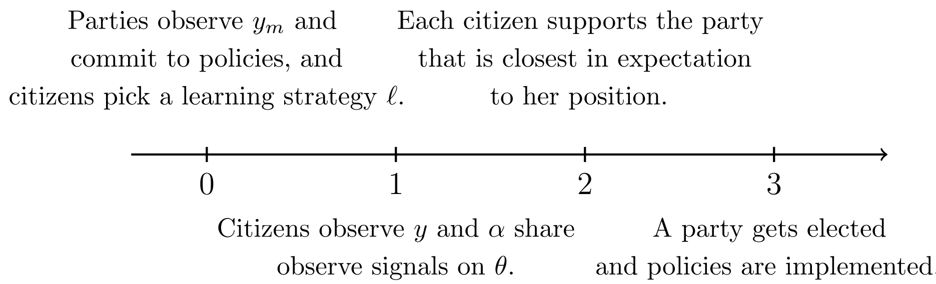

Figure 1: Timing of events

Discussion of the Model:

This section revisits the key modeling assumptions and their implications. As highlighted above, the main focus of the paper is to study how the learning environment impacts the distribution of citizens' choices and, consequently, the policies proposed by the parties.

Heterogeneity and multi-dimensional uncertainty. We assume that different individuals evaluate the many issues that might inform their position on a policy differently. That is, we study a setting in which individuals place different weights on each issue, as opposed to one in which agents hold different views on the state of world for each issue. This assumption provides a natural framework in which to study the implications of specialized learning.

Learning environment. The learning technology, with partition $\mathcal{P}$, effectively limits the degree to which citizens can differentially get informed on separate dimensions of the state and, with access parameter α, puts constraints on how much the electorate is able to learn overall. In a model with multi-dimensional uncertainty, comparative statics on the partition $\mathcal{P}$ can be used to study how the ability to specialize affects polarization. Focusing on α, on the other hand, provides insight into the way that greater access to information (holding the level of specialization constant) affects polarization. Note that all citizens are assumed to

[17] Note that this implies that parties put equal weight on each issue. This is a natural assumption to make if we take the parties to be representative of their support base.

face the same learning technology and that their choice of ℓ can be interpreted as solving an attention allocation problem, subject to the constraints of the learning environment.

Policy choice. When parties commit to their policy choices, they are assumed to know only the position of the representative citizen y_m but not the state of the world θ. This is a reasonable assumption, to the extent that acquiring information about the aggregate preferences (or the position of the median voter) is much easier for parties relative to learning about the entire distribution of preferences. The parties are not able to shift policy in response to the state of the world. This contributes to the uncertainty, with respect to how each party's probability of winning changes with the party's positions.[18]

3 Equilibrium with Learning

The solution concept used is Perfect Bayesian Equilibrium (PBE). A PBE is characterized by three components: policy choices y for the parties; a learning strategy ℓ for each type of citizen; and a decision rule a on which party to support conditional on signals and y for each type of citizen. Given citizens' learning strategies, each party maximizes its expected payoff holding the other party's policy choice constant. Each citizen chooses a learning strategy to maximize her expected payoff and chooses to support the party that gives her the highest payoff.

An equilibrium is referred to as *symmetric* if parties polarize equally to the left and right of the position of the representative citizen— i.e., $y_R = y_m + \beta$ and $y_L = y_m - \beta$ for some $\beta > 0$.

Optimal learning strategy:

The following definition will be useful in describing citizens' learning strategies.

Definition 1. $w_{\mathcal{P}}$ is a *projection* of w on partition $\mathcal{P}$, if $w_{\mathcal{P}} \in \Delta_{\mathcal{P}}$ and $\sum\limits_{k \in A} w_{\mathcal{P}}^k = \sum\limits_{k \in A} w^k$ for every $A \in \mathcal{P}$.

The vector $w_{\mathcal{P}}$ can be interpreted as the weight vector that is most similar to w under the constraint that weights must be same for issues that lie in the same cell of the partition

[18]While there are certainly instances in which policies can adjust to new information about voter preferences, a fully flexible shift in policy positions to respond to the state of the world is invariably difficult.

$\mathcal{P}$. Formally, $w_{\mathcal{P}}$ is the orthogonal projection of w onto the vector space $\Delta_{\mathcal{P}}$. That is, $w = w_{\mathcal{P}} + w_{\mathcal{P}^{\perp}}$, where $w_{\mathcal{P}} \in \Delta_{\mathcal{P}}$ and $w_{\mathcal{P}^{\perp}} \cdot \ell = 0$ for any $\ell \in \Delta_{\mathcal{P}}$. Note also that, by definition, $w_{\mathcal{P}}$ is unique.

Lemma 1. *Fix any y where $y_R = y_m + \beta_R$ and $y_L = y_m - \beta_L$ for some β_R and β_L. For any citizen, $\ell = w_{\mathcal{P}}$ is an optimal learning strategy. In equilibrium, the probability with which a citizen supports either party is the same with all optimal learning strategies.*

The first part of Lemma 1 states that for any citizen, it is optimal to follow a learning strategy that reflects the aggregate weight that she puts on the different subset of issues included in each cell of partition $\mathcal{P}$. The degree to which specialization can take place is naturally constrained by the partition associated with the learning technology. Consider, for example, a citizen with $w = (0.25, 0.75, 0, 0)$, who puts positive weight on only the first two of the four issues relevant to the policy. Assume that the learning technology is such that $\mathcal{P} = \{\{1, 2\}, \{3, 4\}\}$. This implies that the citizen can differentiate her learning between the first two and the last two issues but cannot target learning on the first issue relative to the second. By Lemma 1, $(0.5, 0.5, 0, 0)$ is an optimal learning strategy for such an agent. Lemma 1 also suggests that the learning technology, through the partition $\mathcal{P}$, constrains the degree to which heterogeneity in preferences is manifested in voting decisions. That is, conditional on receiving a signal, this citizen's decision on which party to support will be identical to that of an agent with $w = (0.5, 0.5, 0, 0)$.

The second part of Lemma 1 states that, while there can be multiplicity with respect to the optimal learning strategy, this multiplicity has no impact on the voting patterns of the citizens in equilibrium.[19] In all future results, we'll assume that citizens choose $\ell = w_{\mathcal{P}}$ as their learning strategy. All formal proofs are provided in the Appendix.

Definition 2. Given policy choices y, let $\pi(y) = (\pi_R(y), \pi_L(y)) := \left(\frac{-\partial s_R(y)}{\partial y_R}, \frac{-\partial s_L(y)}{\partial y_L}\right)$ stand for the marginal cost of deviations on policies in terms of the probability of winning (evaluated at y). This is referred to as *responsiveness* to policies.

In an equilibrium, two forces dictate the positions of the parties. Let $\nu_i(y) = -(y_i^\star - y_i)^2 + (y_i^\star - y_{-i})^2$ represent the net utility gain for party i of implementing policy y_i over

[19] However, such learning strategies could lead to different voting patterns off the equilibrium path.

policy y_{-i}. The optimality condition written for party i reveals these two opposing forces:

$$-\pi_i(y)\nu_i(y) + s_i(y)\frac{\partial \nu_i(y)}{\partial y_i} = 0. \tag{4}$$

Conditional on winning the election, parties prefer policies that are closer to their ideological position (second component of Equation 4). This can be considered as a *policy distortion effect.* At the same time, deviating from the equilibrium policy by moving closer to the party's ideal position decreases the probability with which the party wins the election (first component of Equation 4). This can be considered as a *fear of loss effect.* The force of this effect increases with polarization in policy platforms but decreases with responsiveness to the policy choices. Standard techniques guarantee the existence of a unique equilibrium that is symmetric in which these opposing forces are in balance when there is sufficient uncertainty associated with the distribution of ideal points.[20]

Proposition 1. *Assume all citizens choose $\ell = w_{\mathcal{P}}$ as their learning strategy. For any b and F, there exists a $\bar{\rho} > 0$ such that for all $\rho < \bar{\rho}$, a unique equilibrium exists. This equilibrium is symmetric. Parties strictly polarize in equilibrium: the proposed policies are to the right and left of the position of the representative citizen.*

A key observation is that polarization is closely linked to responsiveness to policies. As seen in Equation 4, the degree to which the *fear of loss effect* moderates policy polarization depends on $\pi(y)$. When there is greater uncertainty with respect to the distribution of preferences at the point when parties commit to policies, the marginal cost of proposing more-extreme policies declines, and party platforms polarize further. The upcoming sections investigate how changes in the distribution of citizens, as well as changes in the learning technology, can affect equilibrium policies through this channel.

Fractionalized societies

The following definition introduces a partial ranking on how specialized citizens are.

Definition 3. Let i and j be two different citizens with associated weight vectors w_i and w_j. i is a *more specialized citizen* than j (relative to partition $\mathcal{P}$) if, for $\tilde{w}_{i_{\mathcal{P}}}$ and $\tilde{w}_{j_{\mathcal{P}}}$, referring

[20] Asymmetric equilibria may exist if citizens choose learning strategies other than $w_{\mathcal{P}}$, because in such cases, how citizens respond to deviations depend on off-path beliefs about y_m.

to permutations of projections $w_{i_\mathcal{P}}$ and $w_{j_\mathcal{P}}$ such that weights are in increasing order:

$$\tilde{w}_{i_\mathcal{P}} \geq_{FOSD} \tilde{w}_{j_\mathcal{P}} \tag{5}$$

when $\tilde{w}_{i_\mathcal{P}}$ and $\tilde{w}_{j_\mathcal{P}}$ are considered to be probability distributions over a discrete set of outcomes with later outcomes ranked higher.

The definition captures the idea that one citizen is more specialized than another if her weight vector is more concentrated. That is, her ideal position is highly dependent on the realization of the state of the world on a subset of issues. As an example, consider a complex policy (changes to the tax code or healthcare) that will impact society in many different ways, and, hence, the policy can be evaluated by considering each of these dimensions separately. It is natural that different citizens put different weights on these dimensions. Focusing on four dimensions, a citizen with weight vector $(1, 0, 0, 0)$, who cares about only one of the dimensions, will be more specialized than a citizen with weight vector $(0.5, 0.5, 0, 0)$, who cares equally about only the first two dimensions; and this citizen will be more specialized than a citizen with weight vector $(0.25, 0.25, 0.25, 0.25)$, who cares equally about all issues.[21]

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

Lemma 2. *Assume all citizens choose $\ell = w_\mathcal{P}$ as their learning strategy. More-specialized citizens (relative to partition $\mathcal{P}$) are less responsive to policy choices.*

The key insight is that specialized citizens are more likely to have ideal positions that differ substantially from that of the representative citizen. In forming their ideal positions, they deviate from the representative citizen: instead of considering all issues equally, they focus heavily on a subset of issues. A specialized learning strategy allows them to vote in a way that reflects these differences. Given that their ideal positions can differ substantially the that of the representative citizen around which the parties polarize, they become unlikely to change their vote as a result of small deviations from equilibrium policies. This, in turn, decreases their responsiveness to policies.

Definition 4 extends the previous ranking of specialization across citizens to societies in

[21] Actually, for such a partition, $(0.25, 0.25, 0.25, 0.25)$ represents the *least* specialized citizen, and $(1, 0, 0, 0)$ (and permutations of it) represent the *most* specialized citizens.

a natural way. A society is identified as more fractionalized than another if it is generated by a transformation of the latter, where each citizen is mapped to a more specialized one.

Definition 4. Let S and $\tilde{S}$ be two societies in which the type distribution is captured by F and $\tilde{F}$, respectively. $\tilde{S}$ is a *more fractionalized society* than S relative to partition $\mathcal{P}$ if there exists a mapping g from $\Delta^K \to \Delta^K$ such that, for any w, $g(w)$ describes a more specialized citizen, and $\tilde{F}$ corresponds to the transformed distribution using g from F.

Proposition 2. *Assume all citizens choose $\ell = w_{\mathcal{P}}$ as their learning strategy. If society $\tilde{S}$ is more fractionalized than society S relative to partition $\mathcal{P}$, there is higher polarization in proposed policies in $\tilde{S}$.*

Proposition 2 highlights how fractionalization can generate further polarization in a society. The result is a direct consequence of Lemma 2, which states that more-specialized citizens are less responsive to policy choices. As the population gets more fractionalized, the marginal cost (in terms of the decrease in the 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 position of the representative citizen.

4 Changes in the learning technology

The characterization of the learning environment lends itself to two natural comparative static exercises, one with respect to α and one with respect to the partition $\mathcal{P}$. An increase in α can be interpreted as greater access to information. Considering improvements in learning technologies that allow citizens to get specialized information about issues that are most relevant to them requires focusing on changes in the partition $\mathcal{P}$. Defining learning technologies as partitions of the underlying state space suggests a natural partial order on the level of specialization allowed by the learning technology.

Definition 5. Learning technology $\tilde{\mathcal{L}}$ *allows for more specialization* than $\mathcal{L}$ if the associated $\tilde{\mathcal{P}}$ is a finer partition of $\mathcal{K}$ than $\mathcal{P}$—i.e., for all $\tilde{A} \in \tilde{\mathcal{P}}$, $\exists A \in \mathcal{P}$ such that $\tilde{A} \subseteq A$, and there exist a $\tilde{A} \in \tilde{\mathcal{P}}$ and $A \in \mathcal{P}$ such that $\tilde{A} \subset A$.

Definition 5 provides an intuitive partial order on how much specialization a learning technology allows. Holding the access parameter constant, the partition determines the kind of signals available to the citizens and consequently, affects the degree to which citizens are able to learn about the specific issues that are most relevant to them in determining their position on a policy. A learning technology that allows for more specialization provides greater *depth of learning*, as it effectively increases the number of subissues a citizen is able to learn about differentially.

Proposition 3. *Assume all citizens choose $\ell = w_{\mathcal{P}}$ as their learning strategy. Changes in the learning technology that increase access or allow for more specialization increase the probability that any citizen w (holding y constant) will support the party whose proposed policy is closer to her position.*

Proposition 3 affirms that improvements to the learning technology that increase access to information or allow for specialization are beneficial to citizens on an individual level. Both types of changes allow citizens to learn more effectively and, consequently, increase the probability the the action they take conditional on the information available to them – to support party R or party L – matches their true preferences.

However, the results captured in Proposition 3—i.e., both types of changes result in more effective learning—do not translate into similar equilibrium effects on party positions. The main result of the paper, presented below, illustrates how changes in α and $\mathcal{P}$ affect the policy choices of the parties in radically different ways.

Theorem 1. *For any society S, an increase in α (holding $\mathcal{P}$ constant) decreases polarization in proposed policies. In contrast, changes in $\mathcal{P}$ that allow for more specialization (holding α constant) increases polarization.*

As demonstrated earlier, the degree to which parties polarize depends on citizens' responsiveness to party platforms. Since each citizen supports the party that is closer in expectation to her ideal position, voting patterns depend on the underlying distribution of preferences, as well as on what citizens can learn about their ideal positions through the learning technology available to them. The simple model presented in this paper is intended to demonstrate, in the most transparent way, how greater access to information vs. increased opportunities for specialization can operate to affect responsiveness to policies through different channels.

Greater access to information (increase in α) generates a more informed electorate. For each citizen (holding $\mathcal{P}$ constant), the probability of supporting the party closer to her ideal position increases. Voting patterns in aggregate reflect more closely where the parties are located relative to the position of the representative citizen, essentially making the society more responsive to the policies.[22] In this sense, comparative statics with respect to α capture our basic intuition on how a more informed electorate should be able to monitor the parties more closely, pressuring them to choose policies that are closer to the position of the representative citizen.

In contrast, increased opportunities for specialization (holding α constant) allow individuals to concentrate their learning on the issues that are most important to them. While this improves learning on the individual level—increasing the probability that each citizen will support the party that is closer to her ideal position—it makes the citizen's decision on which party to support less predictable ex-ante. Citizens with diverse views on how an ideal policy should be determined condition their decision on which party to support on different types of information. Differentiation in learning strategies makes these decisions less correlated with each other and, hence, less responsive to the party platforms. In a model in which parties trade off the probability of election with how extreme their policies are, this translates into higher polarization. As a counterpart to the results on access to information, comparative statics with respect to the depth of learning highlight that a more informed electorate does not always imply a reduction in political polarization.

The intuition behind this result is closely linked to the results on fractionalization in the previous section. The following example demonstrates this link. Assume that $\mathcal{K} = \{1, 2, 3, 4\}$, and consider switching from a coarse learning technology with $\mathcal{P} = \{\{1, 2\}, \{3, 4\}\}$ to one in which $\tilde{\mathcal{P}} = \{\{1\}, \{2\}, \{3\}, \{4\}\}$. Take, for example, a citizen with $w = (0.52, 0, 0, 0.48)$. Note that the optimal learning strategy is $w_{\mathcal{P}} = (0.26, 0.26, 0.24, 0.24)$ with the former but $w_{\tilde{\mathcal{P}}} = w$ with the latter. With a coarser learning strategy, the citizen evaluates her expected ideal position using $w_{\mathcal{P}} \cdot \theta$ which is very likely to be close to that of the representative citizen $w_m \cdot \theta$. This implies that such a citizen's decision to support party R or L is highly corre-

[22] Note that uninformed citizens are unresponsive to party platforms, in the sense that their expected ideal position is the same as that of the representative citizen's. Since voters expect the parties to be symmetrically polarized around the position of the representative citizen in equilibrium, they are indifferent between the two parties.

lated with that of the representative citizen. In contrast, with the finer learning technology, the same citizen evaluates her expected ideal position using $w \cdot \theta$ directly, which is likely to differ substantially from that of the representative citizen. Hence, her decision on which party to support is less correlated with that of the representative citizen. Now consider a deviation from a symmetric equilibrium in which parties polarize to the right and left of the position of the representative citizen. If the deviation is towards a more extreme policy, it involves a party moving farther away from the position of the representative citizen. Such a deviation necessarily results in a decrease in the likelihood that citizen of type w supports this party. However, the extent of this decrease depends on the citizen's learning strategy. A more specialized learning technology—such as $\tilde{\mathcal{P}}$ relative to $\mathcal{P}$—allows the citizen to vote in a way that reflects more closely how his ideal position differs from that of the representative citizen. Hence, the likelihood that type w supports a party that is farther away from the position of the representative citizen will be higher with $\tilde{\mathcal{P}}$ relative to $\mathcal{P}$.

The example demonstrates that the level of specialization allowed by the learning technology can impact the degree to which citizens are responsive to policies. Without any changes in the underlying preferences (as the citizen's type w is held constant in the example above), the degree to which heterogeneity in ideal positions is reflected in voting patterns depends on the learning strategies available to the citizens. Citizens choose more-specialized learning strategies when they are available, effectively making themselves more specialized in how they respond to policies. In other words, increasing the depth of learning transforms the society into a more fractionalized one without changing the underlying preferences of the electorate.

4.1 Welfare implications

This section highlights the negative welfare implications associated with polarization. The utilitarian welfare associated with implementing policy y can be written as follows:

$$W(\theta) := \int -(w \cdot \theta - y)^2 dF(w). \tag{6}$$

Symmetry assumptions on F imply that the position of the representative citizen, y_m, is the welfare-maximizing policy for any θ. Thus, polarization in party positions (around the position of the representative citizen) necessarily implies inefficiency in policy choice on the aggregate level. Higher polarization means greater divergence from the position of the representative citizen, hence greater loss in welfare relative to the optimal policy. Thus, changes in learning technologies that lead to greater polarization in the parties' policy positions imply welfare loss for the society. This observation is summarized in the Corollary below.

Corollary 1. For any society S, an increase in α (holding $\mathcal{P}$ constant) increases welfare. In contrast, changes in $\mathcal{P}$ that allow for more specialization (holding α constant) decrease welfare.

5 Conclusion

This paper is motivated by the larger public debate on how the new information landscape governed by a variety of TV channels, social media and digital news reshapes the political process. These technologies have dramatically increased access to information, but they have also endowed citizens with remarkable tools to receive specialized information. The goal of this paper is to demonstrate that, while both channels enable citizens to learn more effectively, they can have radically different consequences for political outcomes. Specifically, the model illustrates how specialization in the type of information that citizens gather can increase political polarization in the absence of any changes in the preferences of the electorate. Thus, the link between specialized learning and polarization provides a potential explanation for the U-shaped pattern traced by political polarization in the last century. When citizens have diverse views on how optimal policy should be determined, specialization allows individuals to concentrate the information they gather on the issues that are most relevant to them. In equilibrium, this enables citizens to learn more about how their ideal positions differ from that of the representative citizen. In aggregate, as different citizens focus on different issues, the electorate becomes less responsive to policies. 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 most. Increasing the depth of learning, which captures the degree to which the learning technology allows for specialization, also increases polarization by transforming the society into a more fractionalized one.

Proofs

Remark 1. *Let $s_i(y \mid \theta)$ be party i's expected vote share among policy voters. The probability party i wins the election is $s_i(y \mid \theta)$.*

Proof of Remark 1

Proof. Let ξ be the random variable that denotes the support for party i among non-policy voters. Party i wins the election if the aggregate vote share (when both policy and non-policy voters are included) is greater than half. Since the measure of non-policy voters is assumed to be the same as policy voters, this happens when $\frac{s_i(y \mid \theta)+\xi}{2} > \frac{1}{2}$. Since ξ is uniformly distributed, this happens with probability $s_i(y \mid \theta)$. □

Proof of Lemma 1

Proof. For any learning strategy ℓ, the utility a citizen of type w receives from voting is equal to $\mathbb{E}_\theta[\max\{\mathbb{E}[u_w(\theta \mid y) \mid \ell \cdot \theta], -\mathbb{E}[u_w(\theta \mid y) \mid \ell \cdot \theta]\}]$ (since the sign of $\mathbb{E}[u_w(\theta \mid y) \mid \ell \cdot \theta]$ determines the direction of the vote). If $\mathbb{E}[u_w(\theta \mid y) \mid \ell \cdot \theta]$ is normally distributed with mean μ and variance σ^2, the expression above (capturing expected utility from voting) is equal to the mean of the folded normal distribution with location and scale parameters (μ, σ^2). This mean is increasing in σ^2. Now we first show that $\mathbb{E}[u_w(\theta \mid y) \mid \ell \cdot \theta]$ is indeed normally distributed for any learning strategy ℓ.

Note that $u_w(\theta \mid y) = -(w \cdot \theta - y_R)^2 + (w \cdot \theta - y_L)^2 = 2(y_R - y_L)\left(w \cdot \theta - \frac{y_R+y_L}{2}\right)$. We can separate $w \cdot \theta = \tilde{w} \cdot \theta + y_m$ where $\tilde{w} = w - w_m$. Let $G(y_m \mid y)$ represent beliefs over the position of the representative citizen y_m given y. Using this we can write

$$\mathbb{E}[u_w(\theta \mid y) \mid \ell \cdot \theta] = 2(y_R - y_L)\left(\mathbb{E}[\mathbb{E}[\tilde{w} \cdot \theta \mid \ell \cdot \theta, y_m] + y_m] - \frac{y_R + y_L}{2}\right)$$

where the expectation over y_m is a function of $G(y_m \mid y)$. Focusing on the inner expectation, let $\tilde{\ell} = \ell - w_m$. The agent learns about $\tilde{w} \cdot \theta$ from $\tilde{\ell} \cdot \theta$. (That is, the agent uses $\ell \cdot \theta - y_m$ as a signal on $\tilde{w} \cdot \theta$). Thus, $\mathbb{E}[\tilde{w} \cdot \theta \mid \ell \cdot \theta, y_m] = \mathbb{E}[\tilde{w} \cdot \theta \mid \tilde{\ell} \cdot \theta]$. Note that this term is normally distributed with mean zero and both $\tilde{w} \cdot \theta$ and $\tilde{\ell} \cdot \theta$ are independent of $y_m = w_m \cdot \theta$. This also implies that the optimal learning strategy is independent of y_m.

Furthermore, this establishes that $\mathbb{E}[u_w(\theta \mid y) \mid \ell \cdot \theta]$ will indeed be normally distributed and the variance will be equivalent to the variance of $\mathbb{E}[\tilde{w} \cdot \theta \mid \tilde{\ell} \cdot \theta]$. Since expected utility associated with voting is increasing in this variance term, the citizen will choose ℓ to maximize the variance of $\mathbb{E}_{\tilde{w}\cdot\theta}[\tilde{w}\cdot\theta \,|\, \tilde{\ell}\cdot\theta]$.

Note that $\mathbb{E}[\tilde{w}\cdot\theta \,|\, \tilde{\ell}\cdot\theta] = \mathbb{E}[\tilde{w}_{\mathcal{P}}\cdot\theta \,|\, \tilde{\ell}\cdot\theta]$ (where $\tilde{w}_{\mathcal{P}} = w_{\mathcal{P}} - w_m$) since $(w - w_{\mathcal{P}})\cdot\theta$ is independent of $\tilde{\ell}\cdot\theta$ and $w_m\cdot\theta$. The variance associated with of $\mathbb{E}[\tilde{w}_{\mathcal{P}}\cdot\theta \,|\, \tilde{\ell}\cdot\theta]$ can be written as:

$$\frac{1}{\rho}\left(\frac{(\tilde{w}_{\mathcal{P}}\cdot\tilde{\ell})^2}{\tilde{\ell}\cdot\tilde{\ell}}\right) = \frac{1}{\rho}\left(\tilde{w}_{\mathcal{P}}\cdot\tilde{w}_{\mathcal{P}}\right)\cos^2(\gamma)$$

where γ is the angle between $\tilde{\ell}$ and $\tilde{w}_{\mathcal{P}}$.[23] Since $\tilde{w}_{\mathcal{P}}$ is fixed by one's type and the learning technology, the citizen maximizes this value by choosing a vector $\tilde{\ell}$ to set $\gamma = 0$ (or $\gamma = \pi/2$) where $\cos^2(\gamma)$ takes the highest possible value 1. Clearly setting $\ell = w_{\mathcal{P}}$ achieves this. While there are other vectors that also achieve this goal, the implied variance is the same for all optimal learning strategies. Thus, the implied variance with the optimal learning strategy is $\sigma^2_{\mathcal{L}} = \frac{\tilde{w}_{\mathcal{P}}\cdot\tilde{w}_{\mathcal{P}}}{\rho}$.

□

Lemma 3. Assume all citizens choose $\ell = w_{\mathcal{P}}$ as their learning strategy. Given policy choices y and the optimal learning strategy of the citizens, the probability party R is elected can be written as:

$$s_R(y) = \alpha \int \Phi\left(\frac{\mu}{\sigma_{\mathcal{L}}(w)}\right) dF(w) + \mathbb{1}_{\beta_R=\beta_L}\frac{1-\alpha}{2} + \mathbb{1}_{\beta_R>\beta_L}0 + \mathbb{1}_{\beta_R<\beta_L}(1-\alpha), \qquad (7)$$

where $\mu = w_m\cdot\theta - \frac{y_R+y_L}{2}$[24], $\sigma^2_{\mathcal{L}}(w) = \frac{\tilde{w}_{\mathcal{P}}\cdot\tilde{w}_{\mathcal{P}}}{\rho}$ (where $\tilde{w}_{\mathcal{P}} = w_{\mathcal{P}} - w_m$) and $\mathbb{1}$ is an indicator function.

Proof of Lemma 3

Proof. Fix an equilibrium (possibly asymmetric). As shown in the proof of Lemma 1, whether or not an informed citizen supports party R depends on the sign of $\mathbb{E}[\mathbb{E}[\tilde{w}\cdot\theta \,|\, \tilde{\ell}\cdot\theta] + y_m] - \frac{y_R+y_L}{2}$. We assume $\ell = w_{\mathcal{P}}$ which implies $\mathbb{E}[\tilde{w}\cdot\theta \,|\, \tilde{\ell}\cdot\theta] = w_{\mathcal{P}}\cdot\theta - y_m$. Plugging this back into the previous expression, we see that a citizen of type w supports party R if $\mathbb{E}[w_{\mathcal{P}}\cdot\theta - y_m +$

[23] Note that for any two vectors x and y, $x\cdot y = (x\cdot x)^{\frac{1}{2}}(y\cdot y)^{\frac{1}{2}}\cos(\gamma)$ where γ is the angle between x and y.

[24] In a symmetric equilibrium where $y_R - y_m = y_L - y_m$, $\mu = 0$, implying the probability party R is elected to be one half as expected from the symmetry. The expression is written this way to capture how deviations from any possible equilibrium can change the probability of winning for each party.

$y_m] - \frac{y_R+y_L}{2} = w_{\mathcal{P}} \cdot \theta - \frac{y_R+y_L}{2} > 0$. Note that $w_{\mathcal{P}} \cdot \theta$ is distributed normally with mean y_m and variance $\sigma^2_{\mathcal{L}}(w) = \frac{\tilde{w}_{\mathcal{P}} \cdot \tilde{w}_{\mathcal{P}}}{\rho}$ (where $\tilde{w}_{\mathcal{P}} = w_{\mathcal{P}} - w_m$). Thus the probability that an informed citizen of type w votes for right party can be written as $\Phi\left(\frac{\mu}{\sigma_{\mathcal{L}}(w)}\right)$ where $\mu = w_m \cdot \theta - \frac{y_R+y_L}{2}$. Integrating over types and considering probability of being informed α gives us the result.

□

Proof of Proposition 1

Proof. Consider an equilibrium where $y_R = y_m + \beta_R$ and $y_L = y_m - \beta_L$. Define $\nu_R(y) = -(y_m + b - y_R)^2 + (y_m + b - y_L)^2 = -(b - \beta_R)^2 + (b + \beta_L)^2 = (\beta_R + \beta_L)(2b - \beta_R + \beta_L)$. Given y_L, y_R must be chosen to maximize $s_R(y)\nu_R(y) - (b + \beta_L)^2$. First we show that (fixing y_L), $s_R(y)\nu_R(y)$ is concave in y_R.[25] It is sufficient to focus on the region where $\beta_R, \beta_L \in [-b, b]$ (as choosing more extreme policies is strictly dominated for the parties). The second order condition can be written as follows:[26]

$$s''_R(y)\nu_R(y) + 2s'_R(y)\nu'_R(y) + s_R(y)\nu''_R(y) < 0$$

Note that the second term is always weakly negative, and that the third term is negative with $\nu''_R(y) = -2$. The first term can be either positive or negative. But, $\nu_R(y)$ is at most $4b^2$. So to prove that the second order condition holds, it is sufficient to show (focusing just on the first and third terms) that $\left|\frac{s''_R(y)}{s_R(y)}\right| 4b^2 - 2 < 0$. Computing $s''_R(y)$ and abstracting from the sign of μ,

$$\left|\frac{s''_R(y)}{s_R(y)}\right| \leq \frac{\int \phi\left(\frac{\mu}{\sigma_{\mathcal{L}}(w)}\right) \frac{|\mu|}{4\sigma^2_{\mathcal{L}}(w)} dF(w)}{\int \Phi\left(\frac{-|\mu|}{\sigma_{\mathcal{L}}(w)}\right) dF(w)}.$$

$|\mu|$ is bounded above by b and for any $w \neq w_m$ (note that w_m types are of measure 0), $\sigma_{\mathcal{L}}(w)$ is increasing in $1/\rho$ implying $\Phi\left(\frac{-|\mu|}{\sigma_{\mathcal{L}}(w)}\right)$ to converge to 0.5 and $\phi\left(\frac{\mu}{\sigma_{\mathcal{L}}(w)}\right) \frac{|\mu|}{4\sigma^2_{\mathcal{L}}(w)}$ to converge to 0 as $1/\rho \to \infty$. Hence, for any distribution F and bias parameter b, the second order condition is satisfied whenever $1/\rho$ is high enough.

Now we show that there cannot be an asymmetric equilibrium. Assume for contradiction that such an equilibrium exists. Without loss of generality assume that the right party takes a more extreme position ($\beta_R > \beta_L$). It must be that the first order condition for both the

[25] Symmetric argument could be made for party L.

[26] $s'_R(y) = \frac{\partial s_R(y)}{\partial y_R}$ and $s''_R(y) = \frac{\partial^2 s_R(y)}{\partial^2 y_R}$.

right and the left parties are satisfied at these policies. Note that this implies the probability of winning for the right party to be less than 1/2. ($s_R(y) < s_L(y)$). For the right party (fixing y_L), it must be that.

$$s'_R(y)\nu_R(y) + 2s_R(y)(b - \beta_R) = 0$$

For the left party, it must be that

$$s'_L(y)\nu_L(y) + 2s_L(y)(b - \beta_L) = 0$$

First we show that $s'_R(y) = s'_L(y)$[27] $s'_R(y) = -\int \phi\left(\frac{\mu}{\sigma_{\mathcal{L}}(w)}\right)\frac{1}{2\sigma_{\mathcal{L}}(w)}dF(w) = s'_L(y)$. Since $s_R(y) < s_L(y)$, for both first order conditions to be satisfied, it must be that $\frac{b-\beta_R}{\nu_R(y)} > \frac{b-\beta_L}{\nu_L(y)}$. This implies

$$\frac{b - \beta_R}{(\beta_R + \beta_L)(2b - \beta_R + \beta_L)} > \frac{b - \beta_L}{(\beta_R + \beta_L)(2b - \beta_L + \beta_R)}$$

Cross multiplying and eliminating common terms, this reduces to $-\beta_R^2 > -\beta_L^2$ which provides the contradiction.

Now we show existence of a symmetric equilibrium by construction. In a symmetric equilibrium where $y_R = y_m + \beta$ and $y_L = y_m + \beta$ the following first order condition must be satisfied (the condition for the left party is identical)[28].

$$s'_R(y)\,(4\beta) + 2s_R(y)(b - \beta) = 0$$

The first term is 0 when $\beta = 0$, but negative for any $\beta > 0$ (note that $s'_R(y)$ is independent of β in a symmetric equilibrium) and strictly increasing in absolute value with β. The second term is positive and strictly decreasing in β reaching 0 when $\beta = b$ (note that $s_R(y) = 0.5$ is independent of β in symmetric equilibrium). Both are changing continuously, thus, by the intermediate value theorem, there exists a unique $\beta \in (0, b)$ where the equation will be satisfied. □

Proof of Lemma 2

Proof. We fix $y = (y_R, y_L) = (y_m + \beta, y_m - \beta)$ for some $\beta > 0$.

$$\pi_R(y) = \frac{\partial s_R(y)}{\partial y_R} = \alpha \int \phi\left(\frac{0}{\sigma_{\mathcal{L}}(w)}\right)\frac{-1}{2\sigma_{\mathcal{L}}(w)}dF(w) = \frac{\alpha}{\sqrt{2\pi}}\int\left(\frac{-1}{2\sigma_{\mathcal{L}}(w)}\right)dF(w) \qquad (8)$$

[27] $s'_i(y)$ always represents the marginal cost of moving to more extreme policies. Thus, for party L, $s'_L(y) = -\frac{\partial s_L(y)}{\partial y_L}$.

[28] $s'_R(y) = s'_L(y)$ and $s_R(y) = s_L(y)$.

By symmetry, $\pi_L(y)$ can be written in similar way. Note that the expression for $s_R(y)$ is given by Lemma 3. We use that $\mu = 0$ in a symmetric equilibrium and that y_R changes μ only through $\hat{y}$, implying $\frac{\partial \mu}{\partial y_R} = -\frac{1}{2}$. Note that the responsiveness of each type of voter is inversely proportional to $\sigma_{\mathcal{L}}(w)$ where, as shown in Lemma 1, $\sigma^2_{\mathcal{L}}(w) = \frac{\tilde{w}_{\mathcal{P}} \cdot \tilde{w}_{\mathcal{P}}}{\rho}$ (with $\tilde{w}_{\mathcal{P}} = w_{\mathcal{P}} - w_m$). Take two citizens such that the projection of their weight vectors on $\mathcal{P}$ is denoted as $y_{\mathcal{P}}$ and $z_{\mathcal{P}}$. Without loss of generality, assume that the weight vectors are ordered in an increasing fashion such that $z_{\mathcal{P}} \geq_{FOSD} y_{\mathcal{P}}$. This implies that for any increasing vector v (where $k > l$ implies $v^k \geq v^l$), $z_{\mathcal{P}} \cdot v \geq y_{\mathcal{P}} \cdot v$. This is equivalent to $\tilde{z}_{\mathcal{P}} \cdot v \geq \tilde{y}_{\mathcal{P}} \cdot v$. We can choose $v = \tilde{y}_{\mathcal{P}}$ which implies $\tilde{z}_{\mathcal{P}} \cdot \tilde{y}_{\mathcal{P}} \geq \tilde{y}_{\mathcal{P}} \cdot \tilde{y}_{\mathcal{P}}$. We can also choose $v = \tilde{z}_{\mathcal{P}}$ which implies $\tilde{z}_{\mathcal{P}} \cdot \tilde{z}_{\mathcal{P}} \geq \tilde{z}_{\mathcal{P}} \cdot \tilde{y}_{\mathcal{P}}$. By transitivity, we have the result. □

Proof of Proposition 2

Proof. By Lemma 2, $\sigma_{\mathcal{L}}(w)$ is higher for more specialized citizens. $\pi_R(y)$ declines (in absolute value) as F distribution shifts towards more specialized citizens. This can be seen clearly in Equation 8. □

Two intermediate claims are provided below to be used in the following proof.

Claim 1. *If $\mathcal{P}_1$ is a finer partition than $\mathcal{P}_2$, then for any citizen of type w, $w_{\mathcal{P}_1}$ is more specialized than $w_{\mathcal{P}_2}$.*

Proof. The claim is not trivial only when $w_{\mathcal{P}_1} \neq w_{\mathcal{P}_2}$. It means that there is at least one partition $A \in \mathcal{P}_2$ such that for some $k, k' \in A$, $w^k_{\mathcal{P}_1} < w^k_{\mathcal{P}_2}$ and $w^{k'}_{\mathcal{P}_1} > w^{k'}_{\mathcal{P}_2}$. Any change of this type will lead to the specialization result. □

Claim 2. *If $\mathcal{P}_1$ is a finer partition than $\mathcal{P}_2$, then for any citizen of type w, the variance of $w_r = (w - w_{\mathcal{P}}) \cdot \theta$ is weakly lower with $\mathcal{P}_1$ than $\mathcal{P}_2$.*

Proof. Note that $w = w_{\mathcal{P}_1} + w_{r_1} = w_{\mathcal{P}_2} + w_{r_2}$. Using that $w_{\mathcal{P}_1} \cdot w_{r_1} = w_{\mathcal{P}_2} \cdot w_{r_2} = 0$, $w_{\mathcal{P}_1} \cdot w_{\mathcal{P}_1} + w_{r_1} \cdot w_{r_1} = w_{\mathcal{P}_2} \cdot w_{\mathcal{P}_2} + w_{r_2} \cdot w_{r_2}$. From $w_{\mathcal{P}_1}$ being more specialized than $w_{\mathcal{P}_2}$, we have that (Lemma 2) $w_{\mathcal{P}_1} \cdot w_{\mathcal{P}_1} \geq w_{\mathcal{P}_2} \cdot w_{\mathcal{P}_2}$ which implies $w_{r_1} \cdot w_{r_1} \leq w_{r_2} \cdot w_{r_2}$. □

Proof of Proposition 3

Proof. As shown the proof of Lemma 1 (focusing on the case for a symmetric equilibrium), conditional on receiving a signal, a citizen's decision on which party to support depends on the sign of $\tilde{w}_{\mathcal{P}} \cdot \theta$ (where as before $\tilde{w}_{\mathcal{P}} = w_{\mathcal{P}} - w_m$). The citizen's true preference over the parties is governed by $(w - w_m) \cdot \theta$. Since all these variables are normally distributed around 0, without loss of generality we focus on the case where $\tilde{w}_{\mathcal{P}} \cdot \theta > 0$. The citizen supports the incorrect party when $\tilde{w}_{\mathcal{P}} \cdot \theta + w_r \cdot \theta < 0$ (where $w_r = w - w_{\mathcal{P}}$). This happens with probability $\Phi\left(\frac{-\tilde{w}_{\mathcal{P}} \cdot \theta}{\sigma_r^2}\right)$ where σ_r^2 is the variance associated with $w_r \cdot \theta$. (Note that we are using the independence of $\tilde{w}_{\mathcal{P}} \cdot \theta$.) Integrating over all realizations of $\tilde{w}_{\mathcal{P}} \cdot \theta$ taking into account it is normally distributed with mean 0, and variance $\sigma_{\mathcal{L}}^2 = \frac{\tilde{w}_{\mathcal{P}} \cdot \tilde{w}_{\mathcal{P}}}{\rho}$ the mistake probability can be written as:

$$\left(\frac{1}{\sigma_{\mathcal{P}}}\right) \int_0^\infty \Phi\left(\frac{-\tilde{w}_{\mathcal{P}} \cdot \theta}{\sigma_r}\right) \phi\left(\frac{\tilde{w}_{\mathcal{P}} \cdot \theta}{\sigma_{\mathcal{L}}}\right) d(\tilde{w}_{\mathcal{P}} \cdot \theta) = \frac{1}{2\pi}\left(\frac{\pi}{2} + \arctan\left(\frac{-\sigma_{\mathcal{L}}}{\sigma_r}\right)\right) \tag{9}$$

The last line is based on the following identity associated with normal distributions:

$$\int_0^\infty \Phi(bx)\phi(ax)dx = \frac{1}{2\pi|a|}\left(\frac{\pi}{2} + \arctan\left(\frac{b}{|a|}\right)\right)$$

Equation 9 shows that the mistake rate is decreasing in $\sigma_{\mathcal{L}}$ and increasing in σ_r. So it is sufficient for the results to show that the first is increasing and the second is decreasing. Note that both of them are a consequence of increased opportunities for specialization. The second is decreasing due to Claim 2 above. The first is increasing due to Claim 1 and Lemma 2.

Note that those citizens who don't receive a signal only support the party that is closest to their ideal position with one half probability. And those who receive signals are always able to increase this probability. The result on α automatically follows from this observation. □

Proof of Theorem 1

Proof. An increase in α increases responsiveness to party platforms (through its direct effect on $\pi(y)$), while increased opportunities for specialization (a finer partition $\mathcal{P}$) imply that the learning strategies (the projection of w on the partition) become more specialized as shown by Claim 1. By Proposition 2, this increases polarization. □

Proof of Corollary 1

Proof. First order conditions reveal that at the optimal policy is characterized by $y = \int w \cdot \theta \, dF(w)$. By linearity $y = \left(\int w dF(w)\right) \cdot \theta = w_m \cdot \theta = y_m$. $\square$

References

Abramowitz, A. I. and K. L. Saunders (2008): "Is polarization a myth?" *The Journal of Politics*, 70, 542–555.

Alesina, A., R. Baqir, and W. Easterly (1999): "Public goods and ethnic divisions," *The Quarterly journal of economics*, 114, 1243–1284.

Angelucci, C., J. Cagé, and M. Sinkinson (2020): "Media Competition and News Diets," *Working paper*.

Ashworth, S. and E. D. de Mesquita (2009): "Elections with platform and valence competition," *Games and Economic Behavior*, 67, 191–216.

Banerjee, A. and R. Pande (2007): "Parochial politics: Ethnic preferences and politician corruption," *mimeo*.

Banks, J. and J. Duggan (2004): "Probabilistic voting in the spatial model of elections: The theory of office-motivated candidates," in *Social Choice and Strategic Decisions*, ed. by D. Austen-Smith and J. Duggan, Springer.

Baron, D. (2006): "Persisten media bias," *Journal of Public Economics*, 90, 1–36.

Bernhardt, D., S. Krasa, and M. Polborn (2008): "Political polarization and the electoral effects of media bias," *Journal of Public Economics*, 92, 1092–1104.

Besley, T. and A. Prat (2006): "Handcuffs for the grabbing hand? Media capture and government accountability," *American economic review*, 96, 720–736.

Bouton, L. and M. Castanheira (2012): "One person, many votes: Divided majority and information aggregation," *Econometrica*, 80, 43–87.

Callander, S. (2005): "Electoral competition in heterogeneous districts," *Journal of Political Economy*, 113, 1116–1145.

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

Castanheira, M. (2003): “Victory margins and the paradox of voting,” *European Journal of Political Economy*, 19, 817–841.

Dewan, T. and D. P. Myatt (2007): “Leading the party: Coordination, direction, and communication,” *American Political Science Review*, 101, 827–845.

Duggan, J. (2005): “A survey of equilibrium analysis in spatial models of elections,” .

Duggan, J. and C. Martinelli (2011): “A spatial theory of media slant and voter choice,” *The Review of Economic Studies*, 78, 640–666.

Eraslan, H. and S. Ozerturk (2018): “Information Gatekeeping and Media Bias,” *Working paper*.

Eyster, E. and T. Kittsteiner (2007): “Party platforms in electoral competition with heterogeneous constituencies,” *Theoretical Economics*, 2, 41–70.

Fernández, R. and G. Levy (2008): “Diversity and redistribution,” *Journal of public economics*, 92, 925–943.

Fiorina, M. and S. J. Abrams (2008): “Political polarization in the American public,” *Annu. Rev. Polit. Sci.*, 11.

Gentzkow, M., J. Shapiro, and M. Sinkinson (2014): “Competition and Ideological Diversity: Historical Evidence from US Newspapers,” *American Economic Review*, 104.

Gentzkow, M. and J. M. Shapiro (2006): “Media bias and reputation,” *Journal of political Economy*, 114, 280–316.

Gentzkow, M., J. M. Shapiro, and D. F. Stone (2015): “Media bias in the marketplace: Theory,” in *Handbook of media economics*, Elsevier, vol. 1, 623–645.

Good, I. J. and L. S. Mayer (1975): “Estimating the efficacy of a vote,” *Behavioral Science*, 20, 25–33.

Groseclose, T. (2001): “A Model of Candidate Location When One Candidate Has a Valence Advantage,” *American Journal of Political Science*, 45, 862–886.

Gul, F. and W. Pesendorfer (2012): “Media and Policy,” *mimeo*.

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

Krasa, S. and M. Polborn (2010): "Competition between specialized candidates," *American Political Science Review*, 104, 745–765.

Krasa, S. and M. K. Polborn (2012): "Political competition between differentiated candidates," *Games and Economic Behavior*, 76, 249–271.

Levy, G. and R. Razin (2015): "Does polarisation of opinions lead to polarisation of platforms?: The case of correlation neglect," *Quarterly Journal of Political Science*, 10, 321–355.

Lizzeri, A. and N. Persico (2001): "The provision of public goods under alternative electoral incentives," *American Economic Review*, 91, 225–239.

Martin, G. J. and A. Yurukoglu (2017): "Bias in cable news: Persuasion and polarization," *American Economic Review*, 107, 2565–99.

Martinelli, C. (2001): "Elections with privately informed parties and voters," *Public Choice*, 108, 147–167.

Matějka, F. and G. Tabellini (2019): "Electoral competition with rationally inattentive voters," .

McCarthy, N., K. T. Poole, and H. Rosenthal (2006): *Polarized America: The dance of ideology and unequal riches*, Mit Press.

McCarty, N., J. Rodden, B. Shor, C. Tausanovitch, and C. Warshaw (2018): "Geography, uncertainty, and polarization," *Political Science Research and Methods*, 1–20.

McKelvey, R. D. and J. W. Patty (2006): "A theory of voting in large elections," *Games and Economic Behavior*, 57, 155 – 180.

Mullainathan, S. and A. Shleifer (2005): "The Market for News," *American Economic Review*, 95.

Myatt, D. P. (2007): "On the theory of strategic voting," *The Review of Economic Studies*, 74, 255–281.

——— (2012): "A rational choice theory of voter turnout," *London Business School.*

Nimark, K. P. and S. Pitschner (2019): "News media and delegated information choice," *Journal of Economic Theory*, 181, 160–196.

Patty, J. W. (2005): "Local equilibrium equivalence in probabilistic voting models," *Games and Economic Behavior*, 51, 523 – 536.

——— (2007): "Generic Difference of Expected Vote Share and Probability of Victory Maximization in Simple Plurality Elections with Probabilistic Voters," *Social Choice and Welfare*, 29, 149–173.

Perego, J. and S. Yuksel (2021): "Media Competition and Social Disagreement," *Working paper.*

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

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

Somin, I. (2016): *Democracy and political ignorance: Why smaller government is smarter*, Stanford University Press.

Stromberg, D. (2004): "Radio's Impact on Public Spending," *Quarterly Journal of Economics*, 119, 189–221.

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

——— (1990): "Spatial strategies when candidates have policy preferences," *Advances in the spatial theory of voting.*

Wittman, D. A. (1973): "Parties as utility maximizers," *American Political Science Review*, 67, 490–498.