Specialized Learning and Political Polarization

Sevgi Yuksel

December, 2018

Abstract

When citizens are differentiated by how much they care about different issues informing policy, specialization allows citizens to concentrate their learning on the issues that are most important for 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 where 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 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] There were high levels of polarization in the early 20th century, which subsided after 1930. However, polarization has been increasing in the last thirty 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 in the last thirty years.[2] 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 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 mid 1980s to mid 2000s.[3] Given that TV was the dominant information source for political news for Americans in this 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] While TV remains 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 the Americans in 2013 stated internet as a main source for news compared to thirteen percent in 2001. (PEW Research Center). 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, rely more on news aggregators and social media. As of spring 2017, two thirds of Americans stated that they get at least some news on social media.[5]

---

[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 & Abrams (2008) and Abramowitz and Saunders (2008) provide arguments on either side.

[3] Nielson report on media. (2004)

[4] Internet use in the US increased was below 20 percent in the mid 1990s, and exceeds 80 percent now. (PEW Research Center)

[5] Use of social media as an political information sources is more common among Nonwhite, less educated

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 - have made it much easier for citizens to receive the type of information that is most relevant to them. While Americans consider themselves to be politically more informed than before, it is not clear if they are informed about the same issues. In this new media landscape where citizens receive curated information that is specialized on the individual level, it is much easier for citizens to concentrate their attention on different issues. These issues can be as distinct as immigration, healthcare or foreign policy.

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 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. Contrary to the previous literature on the topic, the model abstracts away from possible ideological biases in the news media or behavioral biases that might distort the decisions of citizens. The main results of the paper build on the following insight. When citizens 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. 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 that can inform policy. Citizens focus their learning on the issues that are most relevant for them in determining their position. However, as different citizens focus on different issues, they inevitably become better informed on how their ideal position differs from that of the representative agent.

or younger populations. The values are 74 percent in the Nonwhite population, 69 percent for those without a college degree and 78 percent for adults in the 18-49 age group.

While this increases the probability that they support the party whose proposed policy is closest to their ideal position, in aggregate, it has the implication that the electorate becomes less responsive to party platforms. Responsiveness here captures the marginal cost for the parties, in terms of probability of winning the election, of moving to more extreme parties. Decline in responsiveness enables ideologically motivated parties to propose more extreme policies.

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: issues relating 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 how important different issues are, and there is uncertainty on ideal policy response to these issues, party platforms in equilibrium depend on the learning strategies adopted by the citizens. The link between the learning strategies and equilibrium policies arises as follows. Parties are ideologically motivated. 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 position of the representative citizen to increase the probability of election. The way this tension is resolved depends on how the electorate is to the choice of the platform. As demonstrated by the model, that responsiveness depends on the characteristics of the learning technology.

The 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. First, when the learning technology allows for specialization, in *more fractionalized* societies, where there is greater disagreement among citizens about which issues matter the most, equilibrium policies polarize more. For example, 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 one, where all citizens care equally about both issues. Observe that more fractionalized societies, while allowing for higher polarization, look, in aggregate, the same as less fractionalized societies, putting equal weight on both issues. The level of fractionalization in the society captures the degree to which the distribution of preferences in the society differ from that of the representative citizen. Note, however, that this is not heterogeneity in the traditional sense of left-right bias, but heterogeneity that arises from different citizens putting different weights on different issues in a multi-dimensional model. Heterogeneity in the preferences of the citizens implies heterogeneity in response to policy positions 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 a citizen is able to specialize in, also increases polarization. For an electorate whose preferences remain constant, such refinements in the learning technology create further differentiation in terms of the actual learning strategies adopted by the electorate. The insight this paper builds on is that when citizens are uncertain about their position on a policy, the degree to which the underlying heterogeneity in the preferences of the citizens impacts elections depend on the level of specialization allowed by the learning technology. As different citizens take positions determined by different issues, responsiveness to party platforms declines, consequently increasing polarization. The main result in this respect also draws a connection between specialization allowed by learning technologies and 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 remain 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, citizens are assumed to be more likely receive information which could help them with their decision on which party to support. On the individual level, this has the same effect as greater ability to specialize: the probability that each citizen supports the party which is closer to her ideal position increases. However, in contrast to specialization, this has the impact of increasing their 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 indicate the ideal policy of the representative citizen to also be 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 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 on political outcomes.

Empirical evidence suggests that the specialization channel might be more important for understanding current trends. A vast majority of Americans believe 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 is only recently emerging. New papers using machine learning techniques to analyze textual data find evidence of content specialization among information sources (Angelucci et al. (2018); Nimark & Pitschner (2017); Martin & Yurukoglu (2017)).

This paper contributes 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.[6] Recently, McCarty et al. (2018) provides empirical evidence that is suggestive of this causal link.[7] This paper builds on this insight by

[6]Duggan (2005) classifies this as the *stochastic preference model with policy motivation.* Hansson & Stuart (1984), Wittman (1983, 1990), Calvert (1985) and Roemer (1994) are some early papers that make this connection.

[7]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.

endogenizing the uncertainty using a multi-dimensional framework with the aim of studying how changes in the learning stage can affect polarization.

Papers studying political polarization have predominantly focused on the role of media bias where bias can broadly be considered as partial information disclosure or slanted reporting.[8] Levy & 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 explanations include specialized candidates (Krasa & Polborn (2010, 2012)), interaction between valence and ideology (Polborn & Snyder Jr (2017), Eyster & Kittsteiner (2007), Groseclose (2001) and Ashworth & 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 & Tabellini (2017) studies policy choice when voters are rationally inattentive. Complementing 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 & Yuksel (2018) which studies the competitive provision of political information. The critical insight put forward in that paper is that competition forces information providers to specialize by becoming relatively less informative on issues that are important from a social point of view, amplifying social disagreement. Focus on specialization of information in a framework with multi-dimensional preferences link the the two papers. However, while Perego & Yuksel (2018) 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.

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

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

2 Model

Two parties indexed by $i \in \{L, R\}$ compete in an election. Each party commits to a policy y_i 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 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. $\theta := (\theta^1, ..., \theta^K)$ represents relevant facts that are considered in relation to each of these issues, such that $w \cdot \theta = \sum_{k \in \mathcal{K}} w^k \theta^k$ describes a citizen's ideal *position* on the policy.

Distribution of citizen types:

Citizens are distributed with cdf $F(w)$ on Δ^K. The assumed symmetry guarantees that the electorate puts equal weight on each issue in aggregate, but marginal distributions can vary with k. The policy position associated with the representative citizen is denoted as $y_m := w_m \cdot \theta$.[10]

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

Preferences of the citizens:

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

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

Individuals take a binary action $a \in \{0, 1\}$, which could be interpreted as voting, to show support for either party R ($a = 1$) or 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

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

context.[11] That is, for a citizen of type w, the net utility associated with supporting party R can be written as:

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

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.

Uncertainty and learning:

A citizen's decision to support either party depends on y and her ideal policy. As described above, a citizen's ideal policy position might be affected by many factors (captured by the state of the world θ) which she might care about to different degrees (captured by w). While y is publicly observed, citizens have imperfect information on θ. The state of the world on each dimension, θ^k, is assumed to be independently drawn from a normal distribution with mean 0 and precision ρ.

Each citizen is faced with a learning technology $\mathcal{L} := (\alpha, \mathcal{P})$ which imposes restrictions on her ability to independently collect information about different factors. The learning technology is characterized by an *access* parameter $\alpha \in (0, 1]$ and a *partition* $\mathcal{P}$. Formally, $\mathcal{P}$ is a partition of $\mathcal{K}$, which implies the following conditions on $\mathcal{P}$: (1) $\bigcup\limits_{A \in \mathcal{P}} A = \mathcal{K}$; (2) if $A, B \in \mathcal{P}$ such that $A \neq B$ then $A \bigcap B = \emptyset$.

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 receives a costless signal $s = \ell \cdot \theta$ where ℓ is the learning strategy chosen by the citizen 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

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

[11]See Eraslan & Ozerturk (2017), Stromberg (2004), Gentzkow & Shapiro (2006), 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 each cell of the partition corresponds to a subset of issues in $\mathcal{K}$. Each citizen chooses a learning strategy ℓ (which issues to concentrate 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.

The probability a party gets elected is assumed to be directly linked to how much support it has from the public.[12] $\Gamma(y) := \int\limits_w \mathbb{E}_\theta[z_w(y|\theta)]dF(w)$ denotes the expected support party R receives where $z_w(y|\theta)$ represents the action taken by a citizen of type w given policy choices y and state of the world θ. Note that $z_w(y|\theta)$ depends on the optimal learning strategy ℓ of the citizen.

Parties are ideologically motivated.[13] The payoff to party i if policy y is implemented can be written as:

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

where $y_i^\star := y_m + b_i$, the preferred policy for party i, is assumed to be biased either towards the right or left relative to the position of the representative citizen. This implies that parties put equal weight on each issue.[14] L is a loss function and is assumed to be 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 $b_R = -b_L$.

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.

Multi-dimensional uncertainty. It is reasonable to assume that different individuals evaluate

[12]Probability of winning would be an affine transformation of expected vote share, for example, if there are non-policy voters and among these voters the share voting for the R party is distributed uniformly.

[13]Including office motivation for the parties would not change the main results.

[14]This is a natural assumption to make if we take the parties to be representative of their support base.

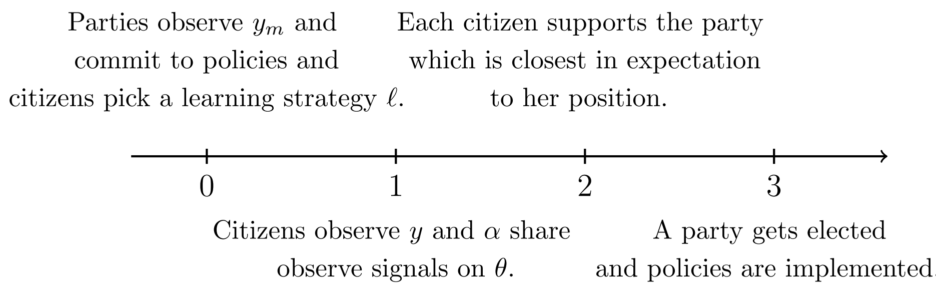

Figure 1: Timing of events

the many issues that might inform their position on a policy differently. This assumption also provides a natural framework 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 on how greater access to information (holding constant the level of specialization) affects polarization. Note that all citizens are assumed to face the same learning technology and their choice of ℓ can be interpreted as solving an attention allocation problem subject to the constraints of the learning environment.

Policy choice. At the point where 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 the probability of winning for each party changes with party positions. While there are certainly many instances where policies shift in response to aggregate shocks, a fully flexible shift in policy positions to respond to the state of the world is invariably difficult.

Baseline with a single issue. To highlight the forces that generate polarization in this context, it is useful to consider the case where $K = 1$. As noted before, this baseline model closely resembles the work of Hansson & Stuart (1984), Wittman (1983, 1990), Calvert (1985) and

Roemer (1994) who first observed a connection between uncertainty about the distribution of voter preferences and polarization of policies proposed by ideologically motivated candidates. The main contribution of this paper is to to endogenize the uncertainty using a multi-dimensional framework with the aim of studying how changes in the learning stage can affect polarization.

3 Equilibrium with Learning

The solution concept used is Perfect Bayesian Equilibrium (PBE).

Equilibrium:

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 for each type of citizen such that:

- each party maximizes its expected payoff (given the strategies of the citizens) holding constant the other party's policy choice;
- each citizen chooses a learning strategy to maximize her expected payoff; and conditional on signals each citizen chooses to support the party which 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 the learning strategies of the citizens.

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

Lemma 1. *In a symmetric equilibrium, it is optimal to choose $\ell = w_{\mathcal{P}}$ as a learning strategy.*

Lemma 1 states that a citizen, following her optimal learning strategy, concentrates attention on the issues that she cares most about. The degree to which such specialization can take place is constrained by the partition associated with the learning technology. All formal proofs (including those for the Remarks) are provided in the Appendix. I highlight the main steps and provide some intuition here. The Lemma formalizes the natural intuition that in a society where there is heterogeneity with respect to which issues are more relevant for determining the optimal policy, there will consequently be heterogeneity with respect to the learning strategies adopted. Given the information available, each citizen supports the party whose proposed policy is closest to his ideal position.

Lemma 1 also suggests that the learning technology, through the partition $\mathcal{P}$, constrains the degree to which heterogeneity in preferences is manifested in citizens decisions (which party to support). Consider, for example, a citizen with $w = (\frac{1}{4}\frac{3}{4}, 0, 0)$, who only puts positive weight on the first two of the four issues relevant for 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 at the cost of the second issue. The optimal learning strategy for this citizen by Lemma 1 is $\ell = (\frac{2}{4}\frac{2}{4}, 0, 0)$. That is, conditional on receiving a signal, the decision of this citizen on which party to support will be identical to an agent with $w = (\frac{2}{4}\frac{2}{4}, 0, 0)$.

The proof of the lemma builds on the following points. For any signal s, the expected net utility associated with supporting party R (L) for a citizen of type w can be written as $\mathbb{E}_\theta[u_w(\theta, y) \,|\, s]$, $(-\mathbb{E}_\theta[u_w(\theta, y) \,|\, s])$. Since the action taken on which party to support will be optimal, the expected payoff will simply be equal to the absolute value of $\mathbb{E}_\theta[u_w(\theta, y) \,|\, s]$. Hence, a citizen will choose a learning strategy ℓ to maximize $\mathbb{E}_s\left[\,|\,\mathbb{E}_\theta[u_w(\theta, y) \,|\, s]\,|\,\right]$. The learning technology, however, puts constrains on how effectively the citizen can learn about her ideal position $w \cdot \theta$. Nonetheless, $w \cdot \theta$ can be written to consist of two components $w \cdot \theta = w_{\mathcal{P}} \cdot \theta + w_r \cdot \theta$. By construction, $w_r \cdot \theta$ is independent of $w_{\mathcal{P}} \cdot \theta$ and y. Hence, maximizing $\mathbb{E}_s\left[\,|\,\mathbb{E}_\theta[u_w(\theta, y) \,|\, s]\,|\,\right]$ becomes equivalent to maximizing $\mathrm{E}_s\left[\,|\,\mathbb{E}_\theta[u_{w_{\mathcal{P}}}(\theta, y) \,|\, s]\,|\,\right]$. Thus, $\ell = w_{\mathcal{P}}$ constitutes the most effective learning strategy.

Remark 1. Given policy choices y and the learning strategy of the citizens in a symmetric equilibrium, the probability party R is elected can be written as:

$$s(y) := \mathbb{E}(\Gamma(y, \theta)) = \alpha \int \Phi\left(\frac{-\delta}{\sigma_{\mathcal{L}}(w)}\right) dF(w) + \frac{1-\alpha}{2}, \tag{4}$$

where $\delta = \frac{y_R + y_L}{2} - y_m$[15] and $\sigma^2_{\mathcal{L}}(w) = \frac{1}{\rho} \sum_k \left(w^k_{\mathcal{P}} - w^k_m\right)^2$.

Remark 1 demonstrates how learning strategies can potentially impact the election. Aggregate uncertainty with respect to winning probability is an equilibrium object. $\sigma_{\mathcal{L}}(w)$ is different for each type w and a function of the learning technology.

Definition 2. Given policy choices y, let $\pi(y) := (\pi_R(y), \pi_L(y))$ stand for the marginal cost of deviations on policies in terms of probability of winning. This is referred as *responsiveness* to policies (evaluated at y).

In a symmetric equilibrium, there are two forces that dictate the positions of the parties. The optimality condition written for party R, for a policy strictly to the right of the ex-ante position of the representative citizen, reveals these two opposing forces:

$$-\pi_R(y) \left[L(y^\star_R - y_R) - L(y^\star_R - y_L)\right] + s(y)L'(y^\star_R - y_R) = 0 \tag{5}$$

Conditional on winning the election parties prefer policies that are closer to their ideological position (second component of Equation 5). 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 5). 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 symmetric equilibrium where these opposing forces are in balance. Note that when $y_R = y_L = y_m$, the first term in Equation 5 representing the *fear of loss effect* is zero but the *policy distortion effect* is strictly positive pushing party R to deviate to a more extreme policy. This suggests that the unique symmetric equilibrium necessarily involves strict polarization.

Proposition 1. *There is a unique symmetric equilibrium. Parties polarize: the proposed policies are strictly to the right and left of position of the representative citizen.*

A key observation is that polarization is closely linked to responsiveness to policies. As seen in Equation 5 which characterizes the symmetric equilibrium, the degree to which

[15] In as symmetric equilibrium, $\delta = 0$ implying the probability party R is elected to be 0.5 as expected from the symmetry. The expression is written this way to capture how deviations from the symmetric equilibrium change probability of winning for each party.

the *fear of loss effect* moderates policy polarization depends on $\pi(y)$. $\pi(y)$ captures the marginal loss of deviating to more extreme policies in terms of the change in probability of winning the election. 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. The upcoming sections investigate how changes in the distribution of citizens preferences 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 to permutations of projections $w_{i_\mathcal{P}}$ and $w_{j_\mathcal{P}}$ such that weights are in increasing order:

$$\tilde{w}_{i_\mathcal{P}} \geq_{FOSD} \tilde{w}_{j_\mathcal{P}} \tag{6}$$

According to the definition, citizen i is a *more specialized citizen* than j (relative to partition $\mathcal{P}$) if the projection of i's weight vector on $\mathcal{P}$ first order stochastically dominates j's when the weights put on different issues are reshuffled such that weights are in increasing order. The definition captures the idea that weights are more concentrated for citizen j than i. As an example, consider a policy (changes to the tax code or healthcare) which will impact society in many different ways. Assume the policy can be evaluated broadly by separately considering it's economic and social consequences. Compare a citizen who puts equal weight on both of these aspects to a citizen who determines her position by solely considering the economic aspects. The corresponding weight vectors can be represented as $(\frac{1}{2}, \frac{1}{2})$ and $(0, 1)$ where the second component corresponds to weight put on economic aspects. Since $(0, 1) \geq_{FOSD} (\frac{1}{2}, \frac{1}{2})$, the second citizen is defined to be more specialized. Actually, for such a partition, where the underlying issues can only be separated into social and economics ones, $(\frac{1}{2}, \frac{1}{2})$ represents the *least* specialized citizen, and $(0, 1)$, $(1, 0)$ represent the *most* specialized citizens.

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

Lemma 2. *More specialized citizens (relative to partition $\mathcal{P}$) are less responsive to policy choices.*

The key issue is that more specialized citizens prefer more specialized learning strategies. Relative to others, the information they gather is more concentrated. This allows them to respond more directly to aspects of the state of the world that are most relevant for them. Critically, these are aspects that differentiate them from the representative citizen. This, in return, decreases their responsiveness to policies.

Definition 4 extends the previous ranking of specialization across citizens to societies in a natural way. A society is identified to be more fractionalized than another one if it is generated by a transformation of the latter one where each citizen is mapped to a more specialized one.

Definition 4. 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 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. *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 relies on Lemma 2 which states that more specialized citizens are less responsive to policy choices. 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 position of the representative citizen.

An important note with regards to Proposition 2 is that more fractionalized societies do not look any different in aggregate. Symmetry assumptions on distribution F impose the position of the representative citizen to remain constant. In terms of the weights put on different issues, in aggregate, the society 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 citizens about which issues matter the most for determining optimal policy. Going back to the example where the partition allows for comparison along two dimensions (consisting of economic and social issues), Proposition 2 implies that a society consisting of citizens who only care about either the economic or social issues (with equal share of both) would generate higher polarization relative to a society where everyone cares equally about both issues.

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 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 for 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 is allowed by a learning technology. Holding constant the access parameter, the partition determines the kind of signals accessible by the citizens, and consequently affects the degree to which citizens are able to learn about the specific issues that are most relevant for them in determining their position on a policy. Going back to the example on healthcare, two citizens, whose position on the policy is determined exclusively by economic considerations, could disagree on which economic measures matter the most. 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 separately about.

Proposition 3. *Changes in the learning technology that increase accress or allow for more specialization increase the probability that any citizen w (holding constant y) supports the party whose proposed policy is closer to her position.*

Proposition 3 affirms that improvements in 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 L – matches their true preferences.

However, the results captured in Proposition 3 on how 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 constant $\mathcal{P}$) decreases polarization in proposed policies. In contrast, changes in $\mathcal{P}$ that allow for more specialization (holding constant α) increases polarization.*

Citizens face uncertainty about how their ideal position differs from that of the representative citizen. The learning technology impacts support for each party by constraining the type of information available to citizens on this. Polarization is linked to responsiveness to policies which is a function of this type of uncertainty. The simple model presented in this paper is intended to demonstrate in the most transparent way how more access to information vs. increased opportunities for specialization can operate to impact citizen's actions through different channels. Easier access to information (increase in α) creates a more informed electorate. For each citizen, the probability of supporting the party closer to her ideal position increases. Actions of the citizens 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. 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 closer, pressuring them to choose policies that are closer to the position of the representative citizen.

In contrast, increased opportunities for specialization allow individuals to concentrate their learning on issues where their preferences differ from that of the representative citizen. While this improves learning from an individual perspective, increasing the probability that the citizen supports the party which 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 where parties are trading off 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 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. 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 partitioned into economic and social issues. With the former learning technology, all citizens are identical in terms of their weight vectors, namely they all put full weight on the only issue available for learning as it includes all possible subissues. Moreover, the optimal learning strategy is trivially determined as the learning technology allows for only one strategy. When the issue space is partitioned into two, there will be some citizens who put equal weight on both issues. The corresponding optimal learning strategy for these types will be to divide their focus their learning equally between the two issues. Consider a society consisting of only such citizens as a benchmark. In this society polarization would not be affected by such a change in the learning technology. However, in general, there will be other types of citizens, who do not equally care about the two subissues. The presence of such citizens implies that the society will effectively be more fractionalized in it's learning strategies relative to the benchmark society described above. 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:

$$W(\theta) := \int -(w \cdot \theta - y)^2 dF(w) \tag{7}$$

Symmetry assumptions on F imply the position of the representative citizen, y_m, to be the welfare maximizing policy for any θ. Thus, polarization in party positions (around the position of the representative citizen) necessarily imply 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. Hence, changes in learning technologies that lead to greater polarization in policy positions of the parties imply welfare loss for the society. This observation is summarized in the Remark below.

Remark 2. For any society S, an increase in α (holding constant $\mathcal{P}$) increases welfare. In contrast, changes in $\mathcal{P}$ that allow for more specialization (holding constant α) decreases 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 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 on political outcomes. Specifically, the model illustrates how specialization in the type of information gathered by citizens 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 for them. In equilibrium, this enables citizens to learn more about how their preferences differ from that of the representative citizen. In aggregate, as different citizens focus on different issues, the electorate becomes less responsive to policies. 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 the degree to which the learning technology allows for specialization, also increases polarization by transforming the society into a more fractionalized one.

Proofs

Proof of Lemma 1

Proof. A citizen of type w strictly prefers party R if and only if $\mathbb{E}_\theta[-(w\cdot\theta-y_R)^2+(w\cdot\theta-y_L)^2|s] > 0$. In a symmetric equilibrium, $y_R = y_m+\beta$ and $y_L = y_m-\beta$. Using this, the expression inside the expectation can be written as $2w\cdot\theta(y_R-y_L)-y_R^2+y_L^2 = 4\beta(w\cdot\theta-y_m)$. The citizen's choice on which party to support will be determined by the sign of $\mathbb{E}_\theta[w\cdot\theta-y_m|s]$. Defining $w_r = w-w_{\mathcal{P}}$, we can write $w\cdot\theta-y_m = w_{\mathcal{P}}\cdot\theta-y_m+w_r\cdot\theta$. By construction, $w_r\cdot\theta$ is independent of $w_{\mathcal{P}}\cdot\theta-y_m$. To see this, we can calculate the covariance between the two variables: $\frac{1}{\rho}\sum_k w_r^k(w_{\mathcal{P}}^k-\frac{1}{K})$. By definition of $w_{\mathcal{P}}$, for any $A\in\mathcal{P}$, there exists a c_A such that for all $k\in A$, $w_{\mathcal{P}}^k = \frac{\sum\limits_{k\in A} w^k}{\sum\limits_{k\in A}\frac{1}{K}} = c_A$ and $\sum\limits_{k\in A} w_r^k = 0$. This means that the covariance can be written as

$$\frac{1}{\rho}\sum_k w_r^k(w_{\mathcal{P}}^k-\frac{1}{K}) = \frac{1}{\rho}\sum_{A\in\mathcal{P}}(c_A-\frac{1}{K})\sum_{k\in A} w_r^k = 0.$$

Note that by construction $w_r\cdot\theta$ is also normally distributed with mean zero. This implies that $\mathbb{E}_\theta[w\cdot\theta-y_m|s] = \mathbb{E}_\theta[w_{\mathcal{P}}\cdot\theta-y_m|s]$. Calling $\tilde{w} = w\cdot\theta-y_m$ and $\tilde{s} = s-y_m$, we can write $\mathbb{E}_\theta[w\cdot\theta-\hat{y}|s] = \frac{cov(\tilde{w},\tilde{s})}{\sigma_{\tilde{s}}^2}\tilde{s}$ where $cov(\tilde{w},\tilde{s})$ captures the covariance of the two variables. For this step, we used that both $\tilde{w}$ and $\tilde{s}$ are normally distributed with mean 0. But since the optimal action relies on the sign of $\mathbb{E}_\theta[\tilde{w}|s]$, the expected value of associated with choosing learning strategy ℓ will be

$$\mathbb{E}_{\tilde{s}}\left[\left|\frac{cov(\tilde{w},\tilde{s})}{\sigma_{\tilde{s}}^2}\tilde{s}\right|\right] = \sqrt{\frac{2}{\pi}}\left(\frac{cov(\tilde{w},\tilde{s})}{\sigma_{\tilde{s}}}\right).$$

The optimal strategy for the agent is to choose ℓ to maximize $\frac{cov(\tilde{w},\tilde{s})}{\sigma_{\tilde{s}}}$. Note that $\frac{cov(\tilde{w},\tilde{s})}{\sigma_{\tilde{s}}} = \rho\sigma_{\tilde{w}}$ where ρ is the pearson correlation coefficient which takes a value in the interval $[-1,1]$. As expected, setting $\ell = w_{\mathcal{P}}$ is sufficient for maximizing this value at 1. $\square$

Proof of Remark 1

Proof. As shown in the proof of Lemma 1, a citizen of of type w supports party R whenever $s-\hat{y} > 0$ where $\hat{y} = \frac{y_R+y_L}{2}$. In a symmetric equilibrium, agents believe that $\hat{y} = y_m$. Given citizens's beliefs about party positions in a symmetric equilibrium, and the learning strategies adopted, $s-\hat{y} > 0$ whenever $(w_{\mathcal{P}}-w_m)\cdot\theta-\delta > 0$. Note that this is consistent with citizens not knowing the actual bias parameter b associated with the preferences of the parties. Thus, for any y that they observe, (given the belief about the symmetric equilibrium) they take $\beta = \frac{y_R-y_L}{2}$, and believe $\hat{y} = y_m$ while any deviation from the symmetric equilibrium would imply $\hat{y} = \hat{y} + \delta$ for some $\delta \neq 0$. δ is included here to capture how deviations from the symmetric equilibrium (which are not detectable to citizens upon observing y_R and y_L) impacts vote shares to different parties. Taking the expectation over θ gives us that type w supports party R with probability

$$\alpha\Phi\left(\frac{-\delta}{\sigma_{\mathcal{L}}(w)}\right) dF(w) + \frac{1-\alpha}{2}$$

where $\sigma^2_{\mathcal{L}}(w) = \frac{1}{\rho}\sum_k \left(w^k_{\mathcal{P}} - w^k_m\right)^2$. Integrating over types gives us the stated result. $\square$

Proof of Proposition 1

Proof. A symmetric equilibrium is pinned down by Equation 5.

$$-\pi_R(y)\left[L(y^\star_R - y_R) - L(y^\star_R - y_L)\right] + s(y)L'(y^\star_R - y_R) = 0 \tag{8}$$

Assumptions on L ensure that policy distortion effect to be positive and continuously decreasing in β reaching 0 when $\beta = b$. The fear of loss effect is always 0 when $\beta = 0$, but negative for any $\beta > 0$ and increasing in absolute value with β. Both are changing continuously which by the mean value theorem gives us the existence and uniqueness of equilibrium for some $\beta \in (0, b)$. $|\pi_R(y)|$ decreases when $\sigma_{\mathcal{L}}(w)$ increases for each type. $\square$

Proof of Lemma 2

Proof. First, we write down the marginal cost of deviations on policies in terms of the probability of winning ($\pi(y) := (\pi_R(y), \pi_L(y))$), which is referred to as responsiveness to policies (evaluated at y). We fix $y = (y_R, y_L) = (y_m + \beta, y_m - \beta)$ for some $\beta \in (0, 1)$.

$$\pi_R(y) = \frac{ds(y)}{dy_R} = -\alpha\int\phi\left(\frac{0}{\sigma_{\mathcal{L}}(w)}\right)\frac{-1}{\sigma_{\mathcal{L}}(w)}dF(w) = \alpha\int\left(\frac{1}{\sigma_{\mathcal{L}}(w)}\right)dF(w) \tag{9}$$

By symmetry, $\pi_L(y)$ can be written in similar way. Note that $\pi_R(y)$ (responsiveness is of a citizen) decreases with $\sigma_{\mathcal{L}}(w) = \alpha\sum_k (w^k_{d_{\mathcal{L}}})^2\frac{1}{\rho}$ (with $w_{d_{\mathcal{L}}} = w_{\mathcal{P}} - w_m$). 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 $z_{d_{\mathcal{L}}} \cdot v \geq y_{d_{\mathcal{L}}} \cdot v$. We can choose $v = y_{d_{\mathcal{L}}}$ which implies $y_{d_{\mathcal{L}}} \cdot z_{d_{\mathcal{L}}} \geq ||y_{d_{\mathcal{L}}}||^2$. We can also choose $v = z_{d_{\mathcal{L}}}$ which implies $||z_{d_{\mathcal{L}}}||^2 \geq y_{d_{\mathcal{L}}} \cdot z_{d_{\mathcal{L}}}$. By transitivity, we have the result. $\square$

Proof of Proposition 2

Proof. By Lemma 2, $\sigma_{\mathcal{L}}(w)$ is higher for more specialized citizens. $\pi_R(y)$ declines as F distribution shifts towards more specialized citizens. This can be seen clearly in Equation 9. $\square$

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. $\square$

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 the variance of w_r is equal to $\frac{1}{\rho}\sum\limits_k (w^k - w^k_{\mathcal{P}})^2$. Focusing on any dimension k, it must be that $(w^k - w^k_{\mathcal{P}_1}) \leq (w^k - w^k_{\mathcal{P}_2})$ which establishes the result. $\square$

Proof of Proposition 3

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

$$\left(\frac{1}{\sigma_{\tilde{w}}}\right) \int_0^\infty \Phi\left(\frac{-\tilde{w}}{\sigma_r}\right) \phi\left(\frac{\tilde{w}}{\sigma_{\tilde{w}}}\right) d\tilde{w} = \frac{1}{2\pi}\left(\frac{\pi}{2} + \arctan\left(\frac{-\sigma_{\tilde{w}}}{\sigma_r}\right)\right) \tag{10}$$

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 10 shows that the mistake rate is decreasing in $\sigma_{\tilde{w}}$ and increasing in σ_r. So it is sufficient for the results to show that the first is increasing and the second is decreasing. The second is decreasing 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 number. 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 Remark 2

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

References

Alesina, A., Baqir, R. & Easterly, W. (1999), 'Public goods and ethnic divisions', *The Quarterly Journal of Economics* **111**(4).

Angelucci, C., Cagé, J. & Sinkinson, M. (2018), Media competition and news diets, Technical report, Mimeo.

Ashworth, S. & de Mesquita, E. D. (2009), 'Elections with platform and valence competition', *Games and Economic Behavior* **67**(1), 191–216.

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

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

Bernhardt, D., Krasa, S. & Polborn, M. (2008), 'Political polarization and the electoral effects of media bias', *Journal of Public Economics* .

Besley, T. & Prat, A. (2006), 'Handcuff's for the grabbing hand? media capture and government accountability', *American Economic Review* **96**(3).

Callander, S. (2005), 'Electoral competition in heterogeneous districts', *Journal of Political Economy* **113**(5), 1116–1145.

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

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

Duggan, J. & Martinelli, C. (2011), 'A spatial theory of media slant and voter choice', *The Review of Economic Studies* **78**(2), 640–666.

Eraslan, H. & Ozerturk, S. (2017), 'Information gatekeeping and media bias'.

Eyster, E. & Kittsteiner, T. (2007), 'Party platforms in electoral competition with heterogeneous constituencies', *Theoretical Economics* **2**, 41–70.

Fernandez, R. & Levy, G. (2008), 'Diversity and redistribution', *Journal of Public Economics* .

Fiorina, M. & Abrams, S. J. (2008), 'Political polarization in the american public', *Annu. Rev. Polit. Sci.* **11**(563-588).

Gentzkow, M. & Shapiro, J. M. (2006), 'Media bias and reputation', *Journal of political Economy* **114**(2), 280–316.

Gentzkow, M., Shapiro, J. M. & Stone, D. F. (2015), Media bias in the marketplace: Theory, *in* 'Handbook of media economics', Vol. 1, Elsevier, pp. 623–645.

Gentzkow, M., Shapiro, J. & Sinkinson, M. (2014), 'Competition and ideological diversity: Historical evidence from us newspapers', *American Economic Review* **104**(10).

Groseclose, T. (2001), 'A model of candidate location when one candidate has a valence advantage', *American Journal of Political Science* **45**(4), 862–886.

Gul, F. & Pesendorfer, W. (2012), Media and policy.

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

Krasa, S. & Polborn, M. (2010), 'Competition between specialized candidates', *American Political Science Review* **104**(4), 745–765.

Krasa, S. & Polborn, M. K. (2012), 'Political competition between differentiated candidates', *Games and Economic Behavior* **76**(1), 249–271.

Levy, G. & Razin, R. (2015), 'Does polarisation of opinions lead to polarisation of platforms?: the case of correlation neglect', *Quarterly Journal of Political Science* **10**(3), 321–355.

Lizzeri, A. & Persico, N. (2001), 'The provision of public goods under alternative electoral incentives', *American Economic Review* **91**(1).

Martin, G. J. & Yurukoglu, A. (2017), 'Bias in cable news: Persuasion and polarization', *American Economic Review* **107**(9), 2565–99.

Martinelli, C. (2001), 'Elections with privately informed parties and voters', *Public Choice* **108**(1-2), 147–167.

Matějka, F. & Tabellini, G. (2017), 'Electoral competition with rationally inattentive voters'.

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

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

Mullainathan, S. & Shleifer, A. (2005), 'The market for news', *American Economic Review* **95**(4).

Nimark, K. & Pitschner, S. (2017), News media and delegated information choice, Technical report, mimeo.

Perego, J. & Yuksel, S. (2018), 'Media competition and social disagreement'.

Polborn, M. K. & Snyder Jr, J. M. (2017), 'Party polarization in legislatures with office-motivated candidates', *The Quarterly Journal of Economics* **132**(3), 1509–1550.

Roemer, J. E. (1994), 'A theory of policy differentiation in single issue electoral politics', *Social Choice and Welfare* **11**(4), 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**.

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