

# Polarization and Policy Design\*

Giovanni Andreottola<sup>†</sup>      Christopher Li<sup>‡</sup>

April 2021

## Abstract

Political polarization has been rising in the U.S. and elsewhere, and there is concern that it will lead to increasingly divisive policymaking. We investigate this issue in a distributive context and show that the effect of polarization on policymaking is nuanced. As voters become more polarized, politicians may design policies to be more equitable. This is the case even if politicians are partisan-minded. Political turnover and the gradual resolution of policy uncertainty are key for the results. The implications of alternative electoral systems are also examined. Contrary to conventional wisdom, proportional systems may exacerbate partisan policymaking compared to majoritarian systems.

---

\*We would like to thank Tim Feddersen, David Levine, Andrea Mattozzi and Alberto Berrardo for useful comments and suggestions. We are also grateful to the audiences at New York University, Princeton University, the European University Institute, the Federico II University of Naples, the Virtual Formal Theory Seminar and the OPESS Political Economy Seminar for their attention and stimulating questions.

<sup>†</sup>CSEF and University of Naples Federico II, Naples, Italy. Email: giovanni.andreottola@gmail.com

<sup>‡</sup>Florida State University, Tallahassee, Florida. Email: christopher.m.li@hotmail.com

# 1 Introduction

Various democracies are experiencing rising political polarization (Carothers and O’Donohue 2019). In the U.S., the partisan divide is at its highest in over a century (McCarty et al., 2016). Many observers are concerned about the consequences of polarization on democracy. Some recent studies suggest that many American voters prioritize party loyalty over even fundamental moral or democratic principles (Hatemi et al. 2019; Graham and Svobik 2020). At the extreme, polarized societies are more prone to civil conflicts (Esteban and Ray 1994; Montalvo and Reynal-Querol 2005).

One important driver of the growing partisan divide is the mutual fear of mistreatment by the opposing faction (Fearon 1999; Padró i Miquel 2007). Clearly, the prospect of partisan policymaking, where the interests of one group are promoted at the expense of others, is always present in a democracy.<sup>1</sup> Nevertheless, the perception is that polarization will exacerbate such tendencies. Indeed, divisive and discriminating policies do often arise in highly polarized polities: land reforms and other kinds of forced redistribution by various populist regimes around the world are notorious examples. Some are worried that the U.S. might be drifting in the same direction, given the increasing prominence of divisive political figures and policy proposals.

Some anecdotal evidence, however, points to a more nuanced relationship between polarization and the extent of partisan policymaking. Consider the Affordable Care Act (ACA), a major healthcare reform introduced in the U.S. in 2010. The political climate surrounding the legislation was highly partisan, with Republicans, who were the minority in Congress, steadfastly refusing to support the bill or even provide any input to it. Public opinion of the legislation was similarly divided along partisan lines. Nonetheless, the Democrats made

---

<sup>1</sup>The concept of “tyranny of majority” has a long tradition in political philosophy. Relatedly, social psychologists have documented the tendency of individuals to favor those considered of the same “group” and exclude those of the “out-group” (Tajfel 1970; Chen and Li 2009).

several compromises when designing the ACA. For example, the bill incorporated a variety of ideas from previous healthcare legislation proposed by Republicans. In fact, in many respects the ACA was less radical (e.g., more free-market oriented) than the healthcare reform proposed by the first Clinton administration. Finally, as it became clear later on, the fear of many Republicans that they would stand to lose from the reform proved unjustified. As of 2016, 68 of the 70 counties most heavily subsidized under the ACA were Republican-leaning (Noam 2017).

It is also interesting to note that despite the growing antagonism between the Democratic and the Republican party, their positions on some important economic issues have converged. For example, the Republican party has recently shifted from its long-term strongly pro-trade stance towards a more protectionist stance.<sup>2</sup> The strategic calculus behind this shift is ostensibly to attract the traditionally Democratic-leaning workers in manufacturing states. The success of this strategy arguably relies on the current high level of partisanship, which allows the Republican party to retain the support of the segment of its base that favors free trade.

The discussion so far begs the question: does the polarization of the electorate necessarily foster increasingly divisive policymaking? We explore this question in a model of distributive politics and find the answer to be negative. Politicians, even partisan-minded ones, may be more willing to make compromises when voters become more polarized. More precisely, the effect of polarization on *group discrimination* – that is the extent of preferential treatment given to one group of voters relative to others – can in fact be non-monotonic.

The effect of *mass polarization* (i.e., polarization of the general public) on policymaking – which constitutes the core of this paper’s contribution – is a phenomenon that still requires a thorough scholarly understanding.<sup>3</sup> In part, this reflects a lack of consensus concerning

---

<sup>2</sup>Interestingly, Democrats have softened their protectionist stance on trade since the 1990s despite the continued loss of manufacturing to overseas.

<sup>3</sup>The existing literature focuses mostly on the implications of the polarization of the political elite (e.g.,

the definition of mass polarization. One obvious definition, the divergence of ideological positions, has mixed empirical support in the U.S. context (Fiorina et al. 2008; Levendusky 2009). What is observed is increasing party loyalty of voters. For instance, there has been a sharp rise in straight-ticket voting in the U.S. (Bartels 2000; Abramowitz and Webster 2016).<sup>4</sup> More generally, there is widespread recognition that voters are becoming increasingly rigid in their allegiance to partisan causes, leading to so-called “tribal politics.” Alternative to party loyalty, another popular interpretation of polarization is based on partisan sorting. That is, there is increasing correlation between partisan affiliation and characteristics such as ideological leaning and socioeconomic status (e.g., Fiorina and Abrams 2008; Abramowitz and Saunders 2008; Bishop 2009). A strength of our model is that it accommodates both interpretations of polarization: partisan loyalty and partisan sorting. Our results suggest that it is important to account for both dimensions when exploring the implications of mass polarization.

The dynamic aspect of policymaking is central to our narrative, and it distinguishes our model from the classic framework for studying distributive politics (e.g., Lindbeck and Weibull 1987; Lizzeri and Persico 2001). In particular, we assume that policies may be repealed in the future because of (endogenous) political turnover. Moreover, individual voters face uncertainty about the policy’s impact, which is revealed slowly over time. As a result of these features, the incumbent party faces a fundamental trade-off when designing a policy: myopically maximizing partisan interests entails a risk of repeal if power changes hands. To achieve policy durability, the incumbent must either build an election-winning coalition to stay in power, or ensure that the policy is sufficiently popular with the opposition to be “repeal-proof.” How the incumbent resolves this variety of trade-offs depends on partisan sorting and partisan loyalty.

---

legislators) (see Barber and McCarty 2015 and Mann and Ornstein 2016 for surveys).

<sup>4</sup>Party loyalty has long been the standard measure of polarization of the political elite (see Barber and McCarty 2015).

To illustrate some of the forces at play, consider the following example. Suppose the Democratic party is in power and seeks to craft a policy that differentially impacts rural vs. urban voters, with urban voters more likely to be Democrats and rural voters more likely to be Republicans.<sup>5</sup> The Democratic party wants to give preferential treatment to urban voters, but a highly biased policy may face repeal if the Republicans win power in the future. When sorting is low, i.e., Democrats and Republicans have a roughly similar composition of urban and rural voters, it takes little compromise from the Democratic party to make the policy sufficiently appealing to Republicans so as to prevent future repeal. As the partisans become better sorted, however, the Democrats need to make larger and larger compromises to ensure the policy is repeal-proof. At some point, this becomes too costly, and the Democrats will instead turn to more partisan proposals that please their base. Thus, the extent of policy bias in favor of urban voters relative to rural voters may appear non-monotonic in the degree of partisan sorting.

A similar comparative statics holds when voters become more polarized in terms of partisan loyalty. For moderate levels of partisan loyalty, the Democratic party can keep its base together while attracting some Republican voters through compromises. This allows the Democrats to stay in power and implement a partially biased policy. However, when partisan loyalty is either low or high, it is impossible for the Democratic party to build a winning coalition, forcing it to either go with a maximally partisan proposal or make sufficiently large compromises to ensure that the policy is repeal-proof.

In addition to policy design, our model makes predictions about the effect of polarization on political turnover and the extent of partisan voting. Our framework is also sufficiently flexible to accommodate extensions in various directions. For instance, the model can be generalized to consider the performance of different electoral institutions in a polarized environment. Our

---

<sup>5</sup>Indeed, in the U.S. there is evidence of geography being a relevant dimension of sorting into parties (see Bishop 2009).

results suggest that whether a proportional system leads to more or less group discrimination compared to a majoritarian system crucially depends on the degree of partisan sorting.

## 2 Related Literature

This paper has some ties to the literature on distributive/pork-barrel politics, although our modeling approach is distinct from the canonical framework based on Downsian competition. Seminal works by Lindbeck and Weibull (1987), Cox and McCubbins (1986), and Dixit and Londregan (1995, 1996) feature office-motivated parties competing by promising particularistic benefits (“pork”) to different segments of the electorate. In equilibrium, parties target benefits to groups that are most responsive. Studies in the literature typically do not consider the implications of polarization. An exception is a recent paper by Krasa and Polborn (2014), which extends the classic model by considering voters who are heterogeneous in two dimensions, economic and cultural. In equilibrium, candidates, who are differentiated in their ability to provide public goods, diverge in their economic platforms. Moreover, when voters become more culturally polarized, the parties respond by shifting economic policies in the *same* direction (e.g., more taxation). In our model, parties may shift policies in *opposite* directions when the electorate becomes more polarized.

In our model, politicians make decisions in the shadow of political turnover, which has long been recognized as an important factor in policymaking in the political economy literature. Persson and Svensson (1989) and Alesina and Tabellini (1990) study the effects of political turnover on government debt. They show that the incumbent may purposefully borrow more than is optimal to tie the hands of future governments. Aghion and Bolton (1990) show that fiscally conservative governments may paradoxically take on inefficiently high levels of debt in order to secure electoral victory. The implications of political turnover for policymaking have also been examined in the context of privatization (Biais and Perotti 2002)

and public spending (Robinson and Torvik 2005 and Besley and Coate 1998), and climate treaties (Battaglini and Harstad 2020). In a different context, Padró i Miquel (2007) shows that ethnic polarization allows corrupt leaders to exploit the mutual fear of opposing groups to extract rents and hold on to power.

Finally, in modeling policy uncertainty and its temporal resolution, we follow Fernandez and Rodrik (1991). The aggregate outcome under the policy is known but individuals face uncertainty regarding whether they are winners or losers under the policy. Fernandez and Rodrik (1991) take the policy proposals as given and demonstrate potential reversals of public support due to the gradual resolution of uncertainty. We embed this concept in a richer setting where the incumbent can design policies subject to political constraints.

### 3 The Model

There are two parties,  $\mathcal{D}$  (Democratic) and  $\mathcal{R}$  (Republican). There is a unit mass of voters who are heterogeneous along two dimensions: party affiliation, denoted by  $\pi \in \{D, R\}$  and targetable traits, denoted by  $\tau \in \{\delta, \rho\}$ . We shall identify voters by the tuple of characteristics,  $(\pi, \tau)$ . As the notation suggests,  $(D, \cdot)$  voters are Democratic partisans and  $(R, \cdot)$  Republican partisans. Targetable traits are factors that a policy can explicitly (and legally) account for, such as socioeconomic or demographic attributes. The marginal distribution of partisan affiliation and of targetable traits is assumed to be uniform, i.e.,  $Pr(D) = Pr(\delta) = \frac{1}{2}$ . There is, however, positive correlation between the two characteristics, with  $Pr(\delta|D) = Pr(\rho|R) = q > \frac{1}{2}$ . The parameter  $q$  measures *partisan sorting*, with higher  $q$  reflecting smaller overlap in the “demographics” of the two partisan camps. This, as argued in the introduction, has been identified in the empirical literature as one measure of polarization. The game proceeds over two periods, indexed by  $t \in \{1, 2\}$ . The Democratic party is assumed

to be in power at  $t = 1$ . The Democratic incumbent (whom we will often refer to as “the Democrat” for simplicity) has an opportunity to adopt a reform. The reform will make some voters “winners” and others “losers.” The aggregate share of winners,  $w > 0$ , is exogenously given and commonly known, but there is uncertainty about the outcomes at the individual level. Such uncertainty resolves inter-temporally. Specifically, as in Fernandez and Rodrik (1991), voters only know the probability of being a winner under the reform at  $t = 1$ ; they observe their status as winners or losers at  $t = 2$ .

The incumbent can design the reform to “statistically” discriminate voters based on their targetable traits.<sup>6</sup> Formally, let  $w_\tau$  denote the probability that a generic  $(\cdot, \tau)$  voter is a winner. The incumbent can adopt any reform  $(w_\delta, w_\rho) \in [0, 1] \times [0, 1]$  subject to the “budget” constraint:

$$\frac{w_\delta + w_\rho}{2} = w. \tag{1}$$

Note that given the constraint, it is without loss of generality to specify the reform by  $w_\delta$  or  $w_\rho$  alone. Recall that we are interested in whether polarization leads to more divisive policies. Here, divisiveness is embodied in the disparity of treatment based on targetable traits,  $|w_\delta - w_\rho|$ .<sup>7</sup> We refer to this measure as *group discrimination*.

The incumbent may choose to forgo the opportunity to reform, in which case voters obtain a status quo payoff of 0 and the game ends. If the incumbent adopts a reform, then an election between the two parties takes place at the end of  $t = 1$ . The winner is decided by majority rule, with a fair-coin toss to break ties.<sup>8</sup> At  $t = 2$ , the ruling party decides whether or not to repeal the reform. This occurs after the identity of winners and losers is revealed but

---

<sup>6</sup>Realistically, policies cannot perfectly discriminate voters based on partisan affiliation, due to either legal or informational restrictions.

<sup>7</sup>Alternatively, one may consider divisiveness as the difference in the expected payoffs in equilibrium. We show in Appendix A.3 that this does not alter our qualitative insights.

<sup>8</sup>This assumption is not crucial to our results; all that is required is that the incumbent loses with some probability when the election is tied.

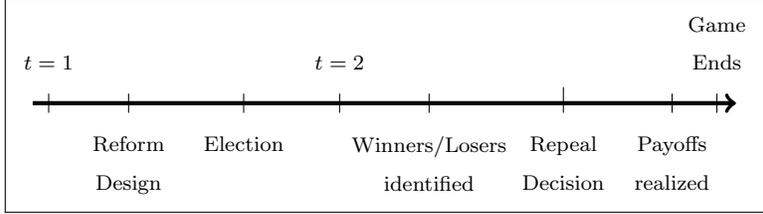


Figure 1: Game timing

before voters obtain their payoffs.<sup>9</sup> If the reform is repealed, then all voters get the status quo payoff of 0. Otherwise, a voter obtains a payoff of 1 if he is a winner and  $-1$  if he is a loser. Figure 1 summarizes the timing of the game.

Voters observe  $(w_\delta, w_\rho)$  and are forward looking. They vote instrumentally based on the anticipated decisions of the two parties at  $t = 2$ , but with a bias toward the party they are affiliated with. Formally, let  $u_{\pi,\tau}^j$  be the expected payoff for  $(\pi, \tau)$  voters if party  $j \in \{\mathcal{D}, \mathcal{R}\}$  is in office at  $t = 2$ .<sup>10</sup> Then,  $(D, \tau)$  voters vote for  $\mathcal{D}$  if and only if

$$u_{D,\tau}^{\mathcal{D}} - u_{D,\tau}^{\mathcal{R}} \leq v, \quad (2)$$

where  $v > 0$ . An analogous decision criterion applies for  $(R, \tau)$  voters. Note that since the reform may discriminate voters based on targetable traits, voters with the same partisan affiliation but different targetable traits may vote differently. The parameter  $v$  measures the extent of party loyalty, which is another dimension of polarization we shall consider along with partisan sorting.<sup>11</sup>

The parties' objective is to maximize the expected utility of their median party member.<sup>12</sup>

This is analogous to “policy motivation” in a spatial model and can be microfounded by

<sup>9</sup>This assumption is not crucial for our results. Voters may obtain a payoff prior to the repeal decision, so long as this is smaller in magnitude than the payoffs accrued afterwards, which can be thought to represent the net present value of payoffs from the reform in the long run.

<sup>10</sup>Specifically,  $u_{\pi,\tau}^j$  is zero if  $j$  is expected to repeal the reform, and it is  $2w_\tau - 1$  otherwise.

<sup>11</sup>We are agnostic about the source of partisan loyalty. It may arise due to ideological alignment on some unmodeled policy dimensions, or to psychological forces associated with group identity.

<sup>12</sup>Given the specification of payoffs, this is also equivalent to maximizing aggregate welfare of all partisans.

considering intra-party politics, such as party leaders or platforms being selected through primary elections. Moreover, allowing for “direct” office motivation would not alter our core findings (see Section A.4 of the Appendix). Given the gradual resolution of policy uncertainty, the identity of the median partisan differs in the two periods. At  $t = 1$ , voters’ expected utility is a function of their targetable traits. Given partisan sorting, the median Democratic partisan is a  $(D, \delta)$  voter. Thus, the Democrat’s objective in period 1 is to maximize the payoff of  $(D, \delta)$  voters.<sup>13</sup> In period 2, voters observe their status as winners or losers. It follows that to maximize the payoff of its median partisan, the ruling party will repeal the reform if and only if a majority of its members are losers.

## 4 Analysis

In this section, we characterize the reform arising in the subgame perfect equilibrium and then derive the comparative statics of group discrimination with regard to the two dimensions of polarization. The section concludes with a discussion of the robustness of the results and other findings of interest.

### 4.1 Three Types of Reform

Suppose the Democrat decides to adopt the reform in period 1: how would the reform be structured? Recall the Democrat’s objective is to maximize the expected utility of  $(D, \delta)$  voters. One obvious strategy for the Democrat is to set  $w_\delta$  as high as possible, which we call the *No-Compromise* reform. However, by myopically maximizing  $w_\delta$ , the Democrat may alienate  $(\cdot, \rho)$  voters, who constitute the majority of the Republican base. This results in the

---

<sup>13</sup>To be precise, the Democrat seeks to maximize  $\mathbb{E}[u_{D,\delta}^j]$  where the expectation is taken with respect to the identity of the incumbent at  $t = 2$ . One may consider  $v$  as part of the instrumental utility and the results would go through.

repeal of the reform if the Republican comes into power in period 2. Therefore, the Democrat may consider making the reform less “divisive” so as to ensure durability.

Intuitively, the Democrat can make the reform durable by (i) making it sufficiently palatable to the Republican base so as to prevent a repeal, or (ii) winning the election. We call the reform most favorable to  $(D, \delta)$  voters while satisfying the first condition the *Repeal-Proof* reform, and the most favorable reform satisfying the second condition the *Election-Winning* reform. Now, if  $w$  is too low, then the reform cannot be made durable no matter its specification. To make the incumbent’s problem non-trivial,  $w$  must be at least  $\frac{1}{2}$ , that is, the reform has to be efficient from the utilitarian perspective.<sup>14</sup> Henceforth, we assume that  $w > \frac{1}{2}$ , unless otherwise stated. An immediate implication of this assumption is that the incumbent strictly prefers reform over the status quo, and consequently we can focus on the design of the reform. Lemma 1 formalizes the observations so far.

**Lemma 1.** *Given that  $w > \frac{1}{2}$ , the incumbent chooses reform over status quo, and the optimal reform takes one of three forms:*

- **No-Compromise:**  $w_\delta = 1$
- **Repeal-Proof:**  $w_\delta = \min \left\{ \frac{wq-1/4}{q-1/2}, 1 \right\}$
- **Election-Winning:**  $w_\delta = \min \left\{ 2w - \frac{1-v}{2}, 1 \right\}$ , but it is only applicable for  $v \in [\underline{v}, 1]$ , with  $\underline{v} = \frac{2(1-q)(2w-1)}{2q-1}$  (the interval is considered empty if  $\underline{v} > 1$ ).

It is worth noting why the Election-Winning reform only exists for intermediate levels of partisan loyalty (i.e.,  $v$ ). In order to win the election, the Democrat needs the support of some Republicans while maintaining the support of its base. The group of Republicans that is easiest to sway are  $(R, \delta)$  voters, who stand to benefit as a byproduct of the Democrat’s partisan objectives. At the same time, the Democrat cannot bias the reform too much lest

---

<sup>14</sup>When  $w < \frac{1}{2}$ , the incumbent will simply choose the No-Compromise reform.

$(D, \rho)$  voters defect. If  $v$  is too high, it becomes impossible to sway  $(R, \delta)$  voters. Conversely, if  $v$  is very small, then to prevent  $(D, \rho)$  voters from defecting, the Democrat must lower  $w_\delta$  (and correspondingly raise  $w_\rho$ ) to such a degree that the reform becomes repeal-proof. But then  $(R, \delta)$  voters have no reason to vote for the Democratic party, as there is no longer any differentiation between the two parties on instrumental grounds.

We now discuss several observations useful for the upcoming analysis of comparative statics. The first one concerns how the specification of the three types of reform varies with partisan loyalty and sorting.

**Observation 1.** *The following holds:*

- *Under the No-Compromise reform,  $w_\delta$  is independent of  $v$  and  $q$ .*
- *Under the Repeal-Proof reform,  $w_\delta$  is independent of  $v$  and decreasing in  $q$ .*
- *Under the Election-Winning reform,  $w_\delta$  is independent of  $q$  and increasing in  $v$ .*

By definition, the No-Compromise reform is independent of both partisan sorting and partisan loyalty. The Repeal-Proof reform is pinned down by the condition that exactly half of Republican partisans are winners. The compromise required to achieve this is increasing (i.e.,  $w_\delta$  decreasing) in the extent of sorting. Finally, concerning the Election-Winning reform, the “binding constraint” for the Democrat is to prevent the defection of  $(D, \rho)$  voters. As partisan loyalty increases,  $(D, \rho)$  voters are less willing to defect, allowing the Democrat to bias the policy more in favor of  $(D, \delta)$  voters (i.e.,  $w_\delta$  increasing).

The second observation concerns the electoral impact of the different types of reforms.

**Observation 2.** *The Repeal-Proof and the No-Compromise reforms, whenever distinct from the Election-Winning reform, result in a tied election.*

Under the No-Compromise reform, the Republican party would repeal the reform in period 2, since a majority of its base are losers. As a result, voting is divided along targetable traits if partisan loyalty is not too extreme:  $(\cdot, \delta)$  voters are sure winners and thus vote for the Democratic party, whereas  $(\cdot, \rho)$  voters are likely to be losers and will vote for the Republican party. If partisan loyalty is sufficiently high, however, voters split along partisan lines. In either case, given the symmetric distribution of voters, the election is tied.<sup>15</sup> The Repeal-Proof reform also leads to a tied election, though for a different reason. As mentioned above, by making the reform repeal-proof, there is no differentiation between the parties on instrumental grounds, hence the voting decision is dictated by party affiliation.

**Observation 3.** *The No-Compromise reform leads to greater group discrimination than the Election-Winning reform, which in turn is more discriminatory than the Repeal-Proof reform.*

Note that  $w_\delta > w$  for all three types of reform, which means that  $w_\delta$  is equivalent to  $|w_\delta - w_\rho|$  in measuring group discrimination. By design, the No-Compromise reform induces maximal group discrimination. To see why the Election-Winning reform is more discriminatory than the Repeal-Proof reform, recall that the Democrat can build an election-winning coalition only when the Republican party is expected to repeal the reform at  $t = 2$ . This means that under the Election-Winning reform, a majority of Republicans are losers. On the other hand, a reform is repeal-proof precisely when a majority of Republicans are winners.

## 4.2 Polarization and Group Discrimination

In this section we present the main result of the paper: the comparative statics of group discrimination with respect to polarization. To this end, Proposition 1 characterizes equilibrium reform design as a function of the parameters.

---

<sup>15</sup>Having a symmetric electorate is not crucial to our results, see discussion in Section 4.3.

**Proposition 1.** *If  $q \leq \frac{1}{4(1-w)} \equiv \underline{q}$ , then the incumbent chooses the No-Compromise reform, which coincides with the Repeal-Proof reform.*

*If  $\underline{q} < q \leq \frac{1}{6-8w} \equiv \bar{q}$ , then there is a threshold  $\underline{v} \equiv \frac{2(1-q)(2w-1)}{2q-1} < 1$ , such that the incumbent chooses the Repeal-Proof reform for  $v \leq \underline{v}$  and  $v \geq 1$  and the Election-Winning reform for  $v \in (\underline{v}, 1)$ .*

*If  $q > \bar{q}$ , then there is a threshold  $\tilde{v} \equiv \frac{5}{2} - 4w$ , with  $\underline{v} < \tilde{v} < 1$ , such that the incumbent chooses the No-Compromise reform for  $v \leq \tilde{v}$  and  $v \geq 1$  and the Election-Winning reform for  $v \in (\tilde{v}, 1)$ .*

When partisan sorting is sufficiently low (i.e.,  $q \leq \underline{q}$ ), the Democrat does not face a trade-off in designing the reform: the No-Compromise reform benefits a sufficient proportion of Republican partisans so as to be repeal-proof. When  $q > \underline{q}$ , the Democrat faces a trade-off between maximizing partisan interests and policy durability.

Whether the Democrat opts for policy durability depends on how much compromise this requires. For moderate levels of partisan sorting (i.e.,  $\underline{q} < q \leq \bar{q}$ ), a slight moderation from the No-Compromise reform ensures repeal-proofness, and this is optimal for the Democrat. Moreover, the Election-Winning reform, when feasible, is preferred to the Repeal-Proof reform since both are durable but the former is more favorable to the median Democratic partisan (see Observation 3). Given that the Election-Winning reform is feasible for intermediate levels of partisan loyalty, in equilibrium the Democrat chooses the Election-Winning reform when partisan loyalty is moderate and the Repeal-Proof reform otherwise.

For high levels of partisan sorting (i.e.,  $q > \bar{q}$ ),<sup>16</sup> the Repeal-Proof reform becomes too “costly” for the Democrat and is therefore dominated by the No-Compromise reform. This leaves the Election-Winning reform as the only viable option for policy durability. Recall that for the Election-Winning reform, the Democrat compromises just enough so that  $(D, \rho)$  voters

---

<sup>16</sup>Notice that the threshold  $\bar{q}$  may not exist if  $w$  is large.

do not defect. As partisan loyalty increases, this constraint becomes less demanding. In fact, for  $v$  sufficiently close to 1, the Democrat need not make any compromise at all. Thus, the Democrat chooses the Election-Winning reform over the No-Compromise reform for  $v \in (\tilde{v}, 1)$ .

**Polarization and group discrimination** Proposition 1 together with the observations in Section 4.1 allows us to derive the comparative statics of group discrimination with respect to the two dimensions of polarization: partisan loyalty and partisan sorting. First, Corollary 1 shows that group discrimination is non-monotonic in partisan loyalty when voters are at least moderately sorted (see Figures 2a and 2b).

**Corollary 1** (Partisan Loyalty). *For  $q < \underline{q}$ , group discrimination is maximal and constant in  $v$ . For  $\underline{q} < q \leq \bar{q}$ , group discrimination is inverse-U-shaped in  $v$ . For  $q > \bar{q}$ , group discrimination is U-shaped in  $v$ .*

As discussed above, when the Election-Winning reform is feasible, an increase in partisan loyalty allows the incumbent to bias the reform more in favor of  $(\cdot, \delta)$  voters, because  $(D, \rho)$  voters are less likely to defect. For extreme values of  $v$ , however, the Election-Winning reform is not feasible and the Democrat must choose between the No-Compromise and the Repeal-Proof reform. For low sorting, it does not take much compromise to make the reform repeal-proof, and therefore it is worthwhile for the Democrat to do so. The “cost” of making the reform repeal-proof is, however, increasing in sorting. For sufficiently high levels of sorting, making the reform repeal-proof is no longer worthwhile for the Democrat, who will instead revert to the No-Compromise reform. This is the intuition from the urban vs. rural example in the introduction.

Next we consider the comparative statics of group discrimination with respect to partisan sorting.<sup>17</sup>

---

<sup>17</sup>In Section A.1 of the Appendix we present an alternative version of equilibrium characterization that

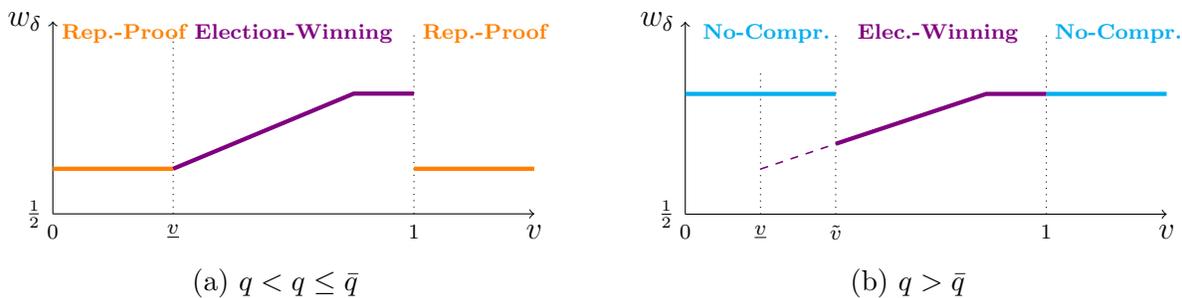


Figure 2: Group Discrimination as a Function of Partisan Loyalty

**Corollary 2** (Partisan Sorting). *For  $v \leq \tilde{v}$  or  $v \geq 1$ , group discrimination is U-shaped in  $q$ . For  $v \in (\tilde{v}, 1)$ , group discrimination is decreasing in  $q$ .*

Figures 3a and 3b are an illustration of the corollary. Recall that for high or low partisan loyalty, the Election-Winning reform is not feasible and thus the choice is between the Repeal-Proof and the No-Compromise reform. It follows then from the discussion of Corollary 1 that group discrimination would be U-shaped in partisan sorting. Now, for intermediate levels of partisan loyalty, the incumbent chooses the No-Compromise reform for low levels of sorting, because it is in fact repeal-proof, and the incumbent opts for the Election-Winning reform otherwise.<sup>18</sup> Consequently, group discrimination is (weakly) decreasing in partisan sorting.

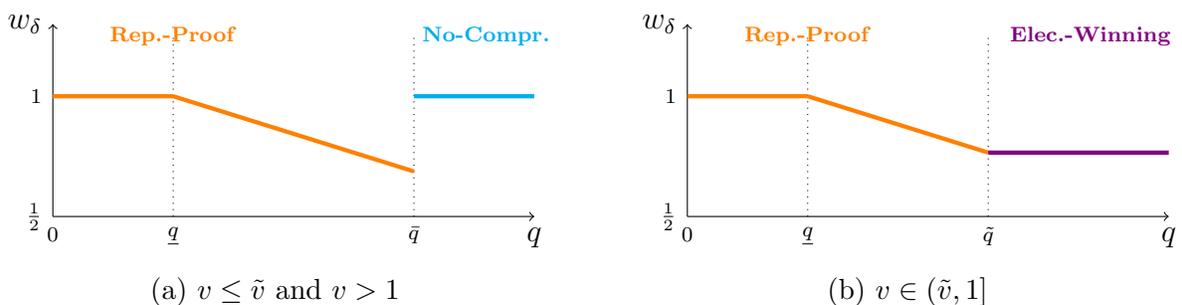


Figure 3: Group Discrimination as a Function of Partisan Sorting

better highlights the effect of partisan sorting on equilibrium reform design.

<sup>18</sup>Note that for  $v$  close to 1, the Election-Winning reform coincides with the No-Compromise reform.

### 4.3 Discussion

The results in the previous section suggest that an increase in polarization does not necessarily lead to more divisive policymaking. Analogous to the idea that competition protects consumers from predatory firms, political competition and the prospect of turnover temper politicians' partisan inclinations. Our analysis also highlights the importance of accounting for the different dimensions of mass polarization when conducting empirical investigations of its consequences.

In Section 3 we have already addressed the (in)significance of various modeling assumptions. Here we further comment on the robustness of our results. First, the baseline model features an electorate that is symmetric in both partisan affiliation and targetable traits. This is mostly to simplify exposition and is not crucial to our qualitative insights provided that one incorporates some electoral uncertainty (more on this below). Similarly, our main insights would not qualitatively change if we allowed for asymmetries in the payoffs for winners/losers; the same can be said if the extent of partisan loyalty were to vary with targetable traits. Finally, in Sections 5 we show that the main findings are robust to electoral uncertainty.

One important aspect of our model is the inter-temporal resolution of policy uncertainty. For one, the viability of the Election-Winning reform leverages the fact that voters face uncertainty about their winner/loser status at the time of the election.<sup>19</sup> Moreover, the resolution of uncertainty at  $t = 2$  makes the Repeal-Proof reform sensitive to partisan sorting, as the latter determines the proportion of winners within each partisan camp.

Another implication of the gradual resolution of uncertainty is that the identity of the median party member differs across two periods. This can cause changes in a party's attitude toward a policy, echoing the insight of Fernandez and Rodrik (1991). For example, according to our

---

<sup>19</sup>If the identity were known at the time of the election, voters would either split their votes along partisan lines (when  $v$  is high) or winner/loser status (when  $v$  is low).

model, the Republican party would oppose the Repeal-Proof reform initially, but then lack the incentive to repeal it in period 2. This may help explain why political oppositions who are adversarial to certain policy initiatives – and may in fact win office on the promise of undoing such policies – sometimes lack the political wherewithal to go through with their promises once in office. One recent example of this is the above-mentioned ACA: Republicans swept into House majority in 2010, and later into the White House, partly on the promise to repeal the ACA. Yet, once in power, their various attempts to repeal the ACA failed to garner enough support, even among their own members.

#### 4.4 Other Comparative Statics

In addition to group discrimination, the model yields a variety of other comparative statics, many of which are readily testable. Corollary 3 below describes the relationship between incumbency advantage, i.e. reelection probability, and the two measures of polarization.

**Corollary 3.** *Incumbency advantage (i.e., reelection probability) is:*

- *Increasing (weakly) in  $q$ .*
- *Constant in  $v$  if  $q \leq \underline{q}$*
- *Inverse-U-shaped in  $v$  if  $q > \underline{q}$*

Another interesting observation, formalized in Corollary 4, is that the share of voters who vote according to party affiliation, or *partisan voting*, need not be monotonically increasing in partisan loyalty. The reason is that voting also depends on the policy proposed by the incumbent, which is endogenous. Generally, voters are split along partisan lines under the Repeal-Proof reform. Under the No-Compromise reform, voters are split along partisan lines when partisan loyalty is high and along targetable traits otherwise. Under the Election-Winning reform, Democratic partisans stay loyal while some Republican partisans defect.

**Corollary 4.** *If  $\underline{q} < q \leq \bar{q}$ , the proportion of voters who vote according to their party affiliation is U-shaped in  $v$ . If  $q > \bar{q}$ , then it is inverse-U-shaped in  $v$ .*

It is also interesting to consider how group discrimination changes with the efficiency of the reform, measured by  $w$ . One might intuitively think that as the “pie” becomes larger (i.e.,  $w$  increases), politicians can afford to give more to the other side in exchange for policy durability. Whereas this is true at low values of  $w$ , eventually leading to a switch from the No-Compromise to the Repeal-Proof or Election-Winning policy, further increasing  $w$  when the reform is already durable causes an increase in group discrimination. Thus, a better reform from a utilitarian perspective may be detrimental from an equity stand point. Finally, when  $w$  is high enough, the No-Compromise policy becomes repeal-proof and group discrimination starts decreasing again in  $w$ , as  $w_\delta$  is bounded above by 1.

**Corollary 5.** *For low and high values of  $w$ , group discrimination is decreasing in  $w$ . For intermediate values of  $w$ , group discrimination is increasing in  $w$ .*

## 5 Alternative Electoral System

In the baseline model, the electoral system can be thought of as majoritarian, or winner-take-all. Another electoral institution common in democracies is the proportional system, in which power is split among candidates based on the vote share. In this section, we consider a slight modification of the baseline model to allow for the proportional system. Specifically, we assume that the probability that a party is in power at  $t = 2$  is equal to its vote share. Note that this assumption can be more generally be interpreted as embodying some electoral uncertainty.

The full equilibrium characterization is given in Proposition A2 in the Appendix. There are two takeaways. First, the relationship between group discrimination and polarization remains

qualitatively unchanged (i.e., Corollary 1 and 2 go through subject to some changes to the thresholds). Second, in comparison to the majoritarian system, the proportional system may reduce or exacerbate group discrimination, depending on partisan sorting. This is illustrated in Figures 4a and 4b.

Intuitively, the proportional system changes the Democrat’s trade-offs insofar as it reduces the appeal of the Election-Winning reform, which no longer guarantees reelection.<sup>20</sup> For  $\underline{q} < q \leq \bar{q}$ , the Democrat chooses the Repeal-Proof reform over the more discriminatory Election-Winning reform for a broader range of values of  $v$  than in the baseline. On the other hand, when  $q > \bar{q}$ , there exists a range of values of  $v$  where the No-Compromise reform is chosen in the proportional system but the Election-Winning reform is chosen in the majoritarian system. In other words, the proportional system leads to more partisan policymaking. This is in contrast to the general wisdom of the existing theoretical literature. For example, Persson and Tabellini (1999) and Lizzeri and Persico (2001) argue that proportional systems lead to greater provision of public goods and/or less targeted transfers.<sup>21</sup> As far as we are aware, the observation of how polarization may lead to different policy outcomes under majoritarian vs. proportional systems is novel to the discussion of institutional design. We believe that the comparative performance of different institutions in a polarized environment is an interesting topic for further study.

## 6 Conclusion

Virtually all policies result in winners and losers, and often policymaking is a balancing act between serving partisan interests and building a broader political coalition. In this paper we explore how politicians address this trade-off when facing a polarized electorate.

---

<sup>20</sup>However, it remains significant in that it maximizes reelection odds.

<sup>21</sup>The relationship between electoral system and inequality is also addressed in Genicot et al. (2021), who find that majoritarian systems can reduce inequality in polities with multiple districts.

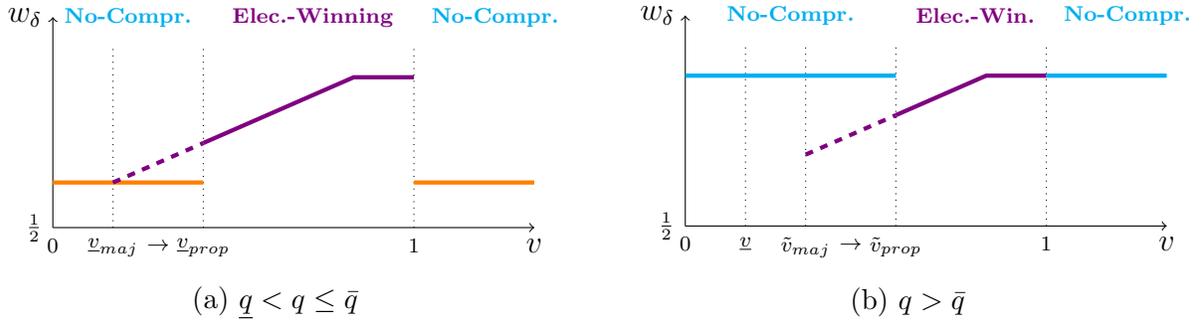


Figure 4: Outcomes under Proportional versus Majoritarian

We show that, somewhat surprisingly, an increase in polarization need not result in more divisive policymaking as the conventional wisdom about “tribal politics” suggests. Our model is parsimonious, yet it delivers a variety of testable findings, for example on how political turnover and partisan voting vary with mass polarization. In an extension of the model, we examine the impact of different electoral systems. Other interesting extensions are possible subjects for future research.

## References

- Abramowitz, A. and K. Saunders (2008). Is polarization a myth? *The Journal of Politics* 70(2), 542–555.
- Abramowitz, A. and S. Webster (2016). The rise of negative partisanship and the nationalization of us elections in the 21st century. *Electoral Studies* 41, 12–22.
- Aghion, P. and P. Bolton (1990). Government domestic debt and the risk of default: a political-economic model of the strategic role of debt. *Public debt management: theory and history* 315.
- Alesina, A. and G. Tabellini (1990). A positive theory of fiscal deficits and government debt. *The Review of Economic Studies* 57(3), 403–414.
- Barber, M. and N. McCarty (2015). Causes and consequences of polarization. *Political negotiation: A handbook* 37, 39–43.
- Bartels, L. M. (2000). Partisanship and voting behavior, 1952-1996. *American Journal of Political Science* 44(1), 35–50.
- Battaglini, M. and B. Harstad (2020). The political economy of weak treaties. *Journal of Political Economy* 128(2), 544–590.
- Besley, T. and S. Coate (1998). Sources of inefficiency in a representative democracy: a dynamic analysis. *American Economic Review* 88(1), 139–156.
- Biais, B. and E. Perotti (2002). Machiavellian privatization. *American Economic Review* 92(1), 240–258.
- Bishop, B. (2009). *The big sort: Why the clustering of like-minded America is tearing us apart*. Houghton Mifflin Harcourt.

- Carothers, T. and A. O'Donohue (2019). *Democracies divided: The global challenge of political polarization*. Brookings Institution Press.
- Chen, Y. and S. X. Li (2009). Group identity and social preferences. *American Economic Review* 99(1), 431–57.
- Cox, G. W. and M. D. McCubbins (1986). Electoral politics as a redistributive game. *The Journal of Politics* 48(2), 370–389.
- Dixit, A. and J. Londregan (1995). Redistributive politics and economic efficiency. *American Political Science Review* 89(4), 856–866.
- Dixit, A. and J. Londregan (1996). The determinants of success of special interests in redistributive politics. *The Journal of Politics* 58(4), 1132–1155.
- Esteban, J.-M. and D. Ray (1994). On the measurement of polarization. *Econometrica* 62, 819–851.
- Fearon, J. D. (1999). Why ethnic politics and ‘pork’ tend to go together. In *an SSRC-MacArthur sponsored conference on “Ethnic Politics and Democratic Stability,” University of Chicago*. Citeseer.
- Fernandez, R. and D. Rodrik (1991). Resistance to reform: Status quo bias in the presence of individual-specific uncertainty. *The American Economic Review* 81(5), 1146–1155.
- Fiorina, M. and S. Abrams (2008). Political polarization in the american public. *Annual Review of Political Science* 11, 563–588.
- Fiorina, M. P., S. A. Abrams, and J. C. Pope (2008). Polarization in the american public: Misconceptions and misreadings. *The Journal of Politics* 70(2), 556–560.

- Genicot, G., L. Bouton, and M. Castanheira (2021). Electoral systems and inequalities in government interventions. Technical report, Forthcoming, Journal of the European Economic Association.
- Graham, M. H. and M. W. Svobik (2020). Democracy in america? partisanship, polarization, and the robustness of support for democracy in the united states. *American Political Science Review* 114(2), 392–409.
- Hatemi, P. K., C. Crabtree, and K. B. Smith (2019). Ideology justifies morality: Political beliefs predict moral foundations. *American Journal of Political Science* 63(4), 788–806.
- Krasa, S. and M. Polborn (2014). Social ideology and taxes in a differentiated candidates framework. *American Economic Review* 104(1), 308–22.
- Levendusky, M. (2009). *The partisan sort: How liberals became Democrats and conservatives became Republicans*. University of Chicago Press.
- Lindbeck, A. and J. W. Weibull (1987). Balanced-budget redistribution as the outcome of political competition. *Public choice* 52(3), 273–297.
- Lizzeri, A. and N. Persico (2001). The provision of public goods under alternative electoral incentives. *American Economic Review* 91(1), 225–239.
- Mann, T. E. and N. J. Ornstein (2016). *It's even worse than it looks: How the American constitutional system collided with the new politics of extremism*. Basic Books.
- McCarty, N., K. T. Poole, and H. Rosenthal (2016). *Polarized America: The dance of ideology and unequal riches*. MIT Press.
- Montalvo, J. G. and M. Reynal-Querol (2005). Ethnic polarization, potential conflict, and civil wars. *American Economic Review* 95(3), 796–816.

- Noam, L. (2017). Trump voters would be among the biggest losers in republicans' obamacare replacement plan. *Los Angeles Times*.
- Padró i Miquel, G. (2007). The control of politicians in divided societies: The politics of fear. *The Review of Economic Studies* 74(4), 1259–1274.
- Persson, T. and L. E. Svensson (1989). Why a stubborn conservative would run a deficit: Policy with time-inconsistent preferences. *The Quarterly Journal of Economics* 104(2), 325–345.
- Persson, T. and G. Tabellini (1999). The size and scope of government: Comparative politics with rational politicians. *European Economic Review* 43(4-6), 699–735.
- Robinson, J. A. and R. Torvik (2005). White elephants. *Journal of Public Economics* 89(2-3), 197–210.
- Tajfel, H. (1970). Experiments in intergroup discrimination. *Scientific American* 223(5), 96–103.