Political Norms*

Giovanna M. Invernizzi† Michael M. Ting‡

September 2, 2020

Abstract

Recent political developments around the world have focused attention on the fraying of political norms, often understood as informal restraints on opportunistic behavior. This paper presents a theory of political norms that incorporates seemingly norm-breaking behavior as part of politicians' equilibrium strategies. In the model, an election determines which party holds office in each period over an infinite horizon. Each period presents the party in office with an opportunity to modify a pre-existing status quo. Parties are constrained from modifying the status quo by both norms and other institutional actors. We show how much opportunism is needed to maintain at least partial cooperation under political conditions such as high polarization or electoral imbalance. The model also examines norm-breaking as a function of institutional factors such as party strength and the separation of powers.

*We thank Benjamin Ogden and panel participants at the 2020 SPSA Annual Meeting.

†Department of Political Science, Columbia University, *Email*: gmi2105@columbia.edu

‡Department of Political Science and SIPA, Columbia University, *Email*: mmt2033@columbia.edu

1. Introduction

Political norms are essential to ensure the functioning of democratic institutions. Their importance is especially salient in high polarization settings, where opposing parties need restraints on opportunistic behavior in order to realize mutually beneficial outcomes. Historically, tragic democratic breakdowns have been preceded by the degrading of basic norms (Levitsky and Ziblatt 2018). Recent political developments and institutional changes around the world have therefore raised urgent questions about the ability of political systems to maintain norms.

Recent changes in political practices and conventions in the United States illustrate the fraying of norms. The confirmation rate of federal circuit court appointments dropped from above 90 percent in the late 1970s and early 1980s to around 50 percent in recent years. Supreme Court appointments are increasingly controversial, and votes across party lines have become less frequent.[1] Beyond judicial appointments, Newt Gingrich's term as Speaker of the House (1995-1999) was widely viewed as ushering in a new era of political confrontation.[2] This trend has been exacerbated by presidents or majority parties in Congress pushing to their legal limits their powers to, for example, fire inspectors general, nominate extreme partisans and filibuster executive or judicial branch nominees or legislation. Outside the United States, recent political developments have challenged one of the other oldest democratic systems: norm-breaking behavior among British MPs have noticeably increased after the 2016 Brexit Referendum.[3]

While the breakdown of norms has understandably received much attention, less is known about the nature of political norms in the first place. This paper develops a simple formal model of cooperative norms among political opponents. In standard fashion, we formalize cooperation as the equilibrium of a repeated game (Kreps 1990). To adapt basic repeated game logic to a political context, we add the two critical features of ideology and institutions. Norms constrain the party in power by defining when politicians can change policies, and their effect is mediated by institutions that provide "hard" constraints on politicians' ability to implement preferred policies. We therefore conceive of political norms as *informal rules* that help preserve mutually beneficial policies. Norms such as constitutional

[1] While Republicans and Democrats once voted unanimously for each others' Supreme Court nominees, recent appointments have become increasingly contested. A widely accepted view suggests that the end of civil discourse in politics began in 1987 when the Senate voted down Robert Bork's nomination to the Supreme Court — cf. Nocera, J. (2011) 'The Ugliness Started With Bork', *New York Times*, 21 October: https://www.nytimes.com/2011/10/22/opinion/nocera-the-ugliness-all-started-with-bork.html.

[2] In one popular telling, Gingrich "wanted to establish the House almost as a parallel government, challenging the president and his policy initiatives at every turn" (Mann and Ornstein 2016).

[3] Jacob Rees-Mogg — head of the European Research Group faction of Conservative MPs — initiated a process that closely resembles Gingrich's "Gang of Seven" by resisting any dilution of the purity of Brexit.

conventions can limit presidential power, but to act contrary to a constitutional convention does not violate any formal rule.

In our model, an election determines which of two parties holds office in each period over an infinite horizon. Each period generates a new status quo policy that can be located anywhere in the policy space. The party in power can modify the status quo by initiating a policy process. To capture both institutional checks and balances and the uncertainty of policy making, we assume that the outcome of the process is uncertain, but closer in expectation to the incumbent's preferred policy than the status quo. The distribution reflects the characteristics of a political system: more flexible systems have few veto players and centralize power, thereby allowing for effective policy making and a distribution of policies more favorable to the incumbent party. The temptation to initiate policy varies from period to period through the status quo, and norms can constrain parties from some welfare-reducing opportunistic behavior.

In a single period, this game does not provide any reason for cooperation: parties obviously choose to initiate policy for any status quo. This amounts to exploiting institutional prerogatives without restriction. Legal scholars have referred to this behavior as "constitutional hardball": playing by the rules, but pushing their boundaries (Tushnet 2004, Fishkin and Pozen 2018). On the other hand, the repeated game admits equilibria where parties refrain from opportunistic behavior. Our analysis focuses on a class of simple equilibria where parties abstain from initiating a policy process when realized status quos are near their bliss points. We show that an optimal equilibrium must be of this form. Cooperation in these equilibria is sustained by "trigger" strategies that threaten perpetual reversion to the static Nash equilibrium. Efficiency is a product of concave policy utility, as norms prevent parties from gaining a small advantage at great expense to the other party. In the most efficient equilibrium, parties can leave most realized status quos alone.

The first set of results show that institutional constraints affect non-monotonically cooperation, and that their effect depends on political norms. In particular, we show that cooperation is not necessarily easier to sustain under more checks and balances. This result suggests that increasing checks and balances under strict political norms — e.g. when constitutional conventions are widely adhered to — can actually hamper cooperation.

We then characterize welfare-maximizing norms. Under these, the amount of initiation leeway parties enjoy varies with institutional constraints: more checks and balances are associated with more permissive norms. But while restraint increases with institutional flexibility, more centralized institutions are not always socially beneficial. A high ability to realize one's ideal policy inherently reduces welfare because of concave preferences, and so from a constitutional perspective a moderate level of institutional flexibility is desirable. Interestingly, welfare-maximizing norms become more sustainable when ideological polarization is high, as polarization increases the need for cooperation. Yet, the payoffs associated with these norms are decreasing in polarization, as the cooperation achieved can only partially compensate for the inevitably large ideological losses that result from losing power. American parties have become increasingly polarized (McCarty, Poole, and Rosenthal, 2016). These results suggest that the observed decay of norms under polarization might be best understood as an adjustment to new, lower-payoff norms.

The critical enforcement mechanism for preserving political norms is ultimately the electorate. We next study a probabilistic voting model to understand which norms are consistent with an election between two parties, where voters choose between the distribution of policies expected from each party in each period. This allows us to microfound parties' probabilities of victory as the outcome of a voting process. Under the optimal norms for the median voter, the electorally disadvantaged party compromises more, in the sense of initiating policy less often. This allows it to win elections at a higher rate than it would in the stage game. Thus, policy concessions help to increase electoral competitiveness. We also show that electoral imbalance makes optimal norms harder to uphold in equilibrium, as the disadvantaged party is less likely to follow the norms. Yet, under electoral imbalance less checks and balances can motivate the trailing party to abide by the optimal norms, which reduce the reach the advantaged party would have under defection. Conversely, more checks and balances can decrease cooperation under the optimal norms.

This paper's main contributions is its joint consideration of informal and formal constraints on policy-making. The importance of political norms for the functioning of democratic systems has been stressed by constitutional scholars (Tushnet 2004, Whittington 2013, Fishkin and Pozen 2018, Siegel 2018), but it is unclear how norms should interact with formal institutional constraints, and how this interaction should affect their sustainability. The model also produces several predictions for the effects of features of the electoral environment, such as ideological polarization and electoral institutions. Existing models of political compromise predict how political power shapes the incentives to cooperate (Dixit, Grossman, and Gul 2000, Fox 2006, Bowen and Zahran 2012); yet, this literature is silent about how cooperation is affected by ideology and institutions. On the other hand, a large literature studies the role of institutional constraints — and particularly separation of powers — on legislation (Tsebelis 1995, Binder 1997, Krehbiel 1998, Stepan and Linz 2011).. While institutional constraints are typically introduced to limit opportunistic behavior, we show that — through their interaction with informal forms — these can backfire and reduce cooperation.

Our model relates to a growing literature that analyzes the drivers of democratic breakdown (Maeda 2010, Waldner and Lust 2018, Levitsky and Ziblatt 2018, Luo and Przeworski 2019). A recent formal literature has analyzed two main explanations for democratic backsliding. Howell, Shepsle, and Wolton (2019) and Helmke, Kroeger, and Paine (2019) focus on the impact of hard constraints — imposed by political elites — on the executive's action, and show how politicians use these same constraints to gradually tilt the playing field to their advantage. A related set of models focus on voters as restraint to the executive action, showing that politicians can succeed in subverting democracy if citizens prioritize partisanship (Nalepa, Vanberg, and Chiopris 2018, Luo and Przeworski 2019, Grillo and Prato 2019, Graham and Svolik 2020). While our paper is not about democratic backsliding *per se*, it explains one of the necessary conditions for backsliding to happen, namely the erosion of cooperative norms.

Finally, our conceptualization of political norms relates to the notion of culture as a constraint to human interactions (North 1991). In particular, a recently growing literature studies the interplay between formal institutions such as legal norms and culture, and how the two tend to evolve together (Tabellini 2008a, 2008b, Guiso, Herrera, and Morelli 2016, Gorodnichenko and Roland 2017, Bisin and Verdier 2017).

The remainder of the paper proceeds as follows. In the next section we provide historical examples of political norms, and their interaction with institutional checks and balances. Section 3 describes the basic model. In Section 4 we characterize a set of simple equilibria for the game, and analyze the set of optimal equilibria. Section 5 then endogenizes the election probabilities by introducing voters. Finally, Section 6 concludes.

2. Political Norms and Institutional Constraints

A primary purpose of constitutional systems of checks and balances is to prevent leaders from abusing power. Yet, legal rules alone are not sufficient to regulate the exercise of political power (Tushnet 2004, Whittington 2013). This idea is corroborated by cross-country evidence: as noted by Levitsky and Ziblatt (2018), Latin American constitutions were modeled on the United States system of checks and balances — e.g. adopting in most instances bicameral legislatures, supreme courts and federal systems. Nevertheless, these countries saw the rise of leaders such as Perón, Marcos, Vargas, Chaves and recently Maduro, all regularly elected under US-like constitutions, who tilted existing institutions to their advantage. Similar recent examples include Turkey, Hungary, Poland and India.

Besides institutional checks and balances, constitutional scholars have identified another crucial ingredient for securing long lasting democracies: the development of strong political norms of cooperation among political opponents. These norms are sometimes referred to as "constitutional norms" (Chafetz and Pozen 2018), "unwritten constitutional conventions" (Whittington 2013), or "norms of institutional forbearance" (Levitsky and Ziblatt 2018), and can be thought of as "avoiding actions that, while respecting the letter of the law, obviously violate its spirit" (p. 91). Examples of such informal rules include presidential two-term limits, conventions limiting the appointments of judges, filling nominally independent agencies with loyalists, and acting unilaterally through decrees and executive orders.

How should we think about the interaction between political norms and formal institutions? First, when political norms are respected, political opponents do not exploit their legal prerogatives to their advantage, even when technically possible, because of a shared understanding that such opportunistic behavior could put the very essence of democracy in danger. The respect for political norms has characterized to a different extent the history of the UK and the US, playing a prominent role in both democracies until very recent episodes of norm-breaking behavior. Second, when norms are adhered to, tightening legal constraints with the intention of limiting opportunistic behavior can have the opposite effect, as the following examples suggest.

Among the institutional features that characterize the US system, the filibuster has received a high level of scrutiny for its institutional impacts. The introduction of the filibuster can be seen as an increase in checks and balances, which can possibly have the unintended consequence of limiting cooperative behavior. Indeed, Wawro and Schickler (2004) provide empirical evidence in favor of the effectiveness of cooperative norms — and against legislative obstruction — before 1917, when the filibuster was formally introduced to the Senate, which contrasts with the decay of such norms that followed its introduction.

While the US has a written, mostly judicially enforced constitution, political norms and constitutional conventions have played a major role in the UK (and, more broadly, the Commonwealth). However, recent institutional changes that have shifted the country away from pure majoritarianism are consistent with our institutional mechanism, through which increasing checks and balances can make cooperation harder to sustain. Besides the introduction of written constraints from the European Union, one of the biggest institutional changes in the UK was the Fixed-term Parliament Act of 2011 — which created more gridlock by removing the Prime Minister's ability to tie a motion of confidence to important votes. Indeed, while the Maastricht treaty was ratified in 1992 because of the attached vote of confidence, the deal on Brexit proposed by Prime Minister May in 2019 was rejected.

This increase in formal checks and balances likely negatively affected the use of constitutional conventions, as the attempt of the Prime Minister Boris Johnson to suspend the parliament in September 2019 exemplifies. While technically within the rules of British democracy — which allow the prime minister to "prorogue" Parliament — Johnson exploited those rules for political advantage, excluding Parliament from much of a Brexit debate that he was likely to lose.[4] While institutional constraints are often introduced to prevent unwanted outcomes, our model and recent UK parliamentary experience both suggest that reducing institutional flexibility can backfire by increasing opportunistic behavior.

3. Model

We analyze an infinitely repeated game of cooperation in a legislature between two parties (L and R) who may alternate in power. Parties differ in their policy preferences: we denote by z_l and z_r party L and R's preferred policy platform respectively. Each party evaluates the future according to a common discount factor $\delta \in (0, 1)$.

In each period t, party L is in power with exogenous probability p_l, party R with probability $1 - p_l$, and a status quo q_t is drawn from a uniform distribution with support $\mathcal{Z} \equiv [z_l, z_r]$, where $(z_l < 0 < z_r)$.[5] Each period results in the adoption of a policy $x_t \in \mathcal{Z}$. Party i's realized payoff from x_t is defined as

$$u_i(x_t, z_i) = \alpha - (x_t - z_i)^2, \tag{1}$$

where $\alpha > 0$ ensures that the realized payoff is nonnegative.

Policies in each period are determined as follows. The party in power can either modify the status quo by *initiating* a policy process or letting the status quo stand. Initiation results in a lottery over policy outcomes. Suppose party L is in power (when not specified otherwise, symmetry holds for party R): for any value of the status quo, party L can initiate a policy process which results in a policy outcome x_t such that $d(x_t, z_l) \leq d(q_t, z_l)$ — that is, party L weakly prefers x_t to the status quo. The fact that the resulting outcome differs from L's bliss point reflects all the uncertainty inherent to the policy making process and the compromise needed to achieve final decisions. Beside uncertainty, the party in power is further constrained in the policy choice by institutional checks and balances. Formally, when L initiates a policy process at time t, the resulting policy x_t is the realization of a random variable X_t, where

$$X_t \sim \mathcal{U}[q_t - \phi, q_t].$$

[4] Cf. Fisher, M. (2019) 'U.K. Democracy, Tested by Brexit and Boris Johnson, Holds for Now', *New York Times*, 5 September: https://www.nytimes.com/2019/09/05/world/europe/boris-johnson-british-parliament.html.

[5] Section 5 endogenizes the probability of victory as the outcome of a voting process.

Clearly, a party's myopically optimal choice is to initiate policy.

The parameter $\phi \in [0, z_r - z_l]$ is a measure of institutional flexibility, or the constraints that may affect a party's ability to realize its ideal policy. As flexibility increases, the average policy outcome resulting from the policy process gets closer to z_l. On the other hand, a system characterized by heavy institutional constraints (low ϕ) results in policy gridlocks reflecting the majority's inability to change the status quo.

For each party i, strategies in the repeated game are mappings $s_{t,i} : H_t \to \{initiate, sq\}$, where H_t is the set of histories of policy decisions up to period t. We focus on a class of equilibria in which parties are further constrained in their policy choice by norms of cooperation. A norm is defined by a pair of thresholds (h_l, h_r) in the policy space $\mathcal{Z}$ that limit parties' opportunistic behavior by preserving the status quo. That is, norms define an interval of the policy space where the party in power does not change q_t. When party L is in power, following a cooperative norm implies that the implemented policy is the status quo when $q_t \leq h_l$, and that L can only initiate a policy process when $q_t > h_l$. When $h_l = z_r$, political norms are always binding and L never initiates a policy process. Oppositely, when $h_l = z_l$ party L can always initiate a policy process to change the status quo when in power. Party L (R) *defects*, or breaks the norm, by initiating a policy process for $q_t \leq h_l$ ($q_t \geq h_r$). For simplicity, we focus on simple trigger strategies: breaking the norm results in both parties initiating policy in every subsequent period. Figure 1 displays the policy realization x_t as a function of the status quo when L is in power and follows the norms of cooperation.

The analysis that follows considers the parameter space where institutional flexibility is low enough to ensure interior realizations of the policy process. The corner case is more cumbersome and does not provide any additional insight; we provide a characterization in the Appendix. Formally:

Assumption 1. *$h_l > z_l + \phi$ and $h_r < z_r - \phi$.*

4. Results

We begin with the stage game. Given a status quo q_t, when party L initiates a policy process, the outcome is the realization of a random variable uniformly distributed with support $[q_t - \phi, q_t]$. The expected utility of party i from L's policy initiation is:

$$U_{i,l}(z_i, q_t, \phi) \quad = \quad \int_{q_t - \phi}^{q_t} u_i(x, z_i) \frac{1}{\phi} \, \mathrm{d}x$$

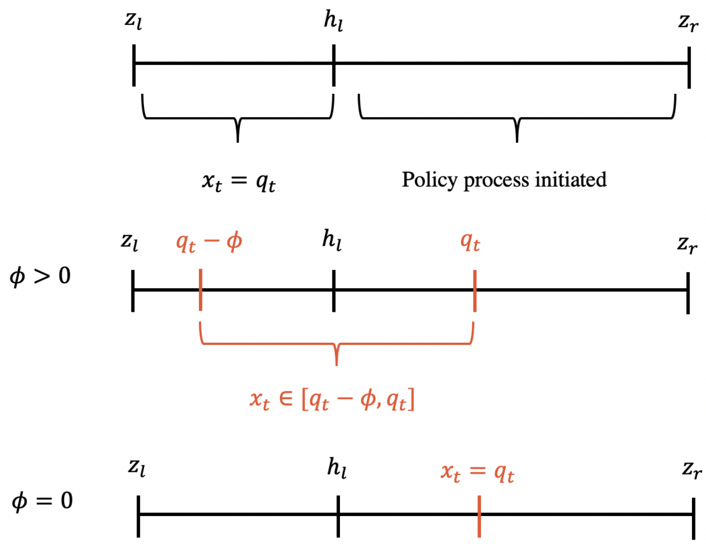

Figure 1 – Norms of cooperation for party L when in power. The parameter h_l determines how bindings norms are (upper figure): when $q_t \leq h_l$, $x_t = q_t$. When $q_t > h_l$, L initiates a policy process that results in an implemented policy distributed uniformly on $[q_t - \phi, q_t]$. When $\phi = 0$, the the party in power cannot change the status quo.

$$= \alpha - (q_t - z_i)^2 + \phi(q_t - z_i) - \frac{\phi^2}{3}. \tag{2}$$

Similarly, when R is in power and initiates a policy process the expected utility of party i is

$$\begin{aligned} U_{i,r}(z_i, q_t, \phi) &= \int_{q_t}^{q_t+\phi} u_i(x, z_i)\frac{1}{\phi}\,\mathrm{d}x \\ &= \alpha - (q_t - z_i)^2 - \phi(q_t - z_i) - \frac{\phi^2}{3}. \end{aligned} \tag{3}$$

These expressions imply that a party in power benefits from initiating policy for any $\phi > 0$. Thus in a one-shot game, or the "punishment" phase of the repeated game, parties always initiate. Lemma 1 summarizes this result without proof.

Lemma 1. Stage Game Equilibrium. *The unique Nash equilibrium of the stage game is* (*initiate*, *initiate*).

In the infinitely repeated game, the pair (h_l, h_r) defines a region of the policy space where parties "cooperate" by refraining from initiating a policy process. Formally, party L chooses *sq* if $q_t \leq h_l$, and *initiate* otherwise. Likewise, party R chooses *sq* if $q_t \geq h_r$, and *initiate* otherwise. Obviously a trivial equilibrium is $h_l = z_l$ and $h_r = z_r$, which corresponds to infinite repetition of the stage game equilibrium.

Apart from its analytical simplicity, our focus on this class of equilibria is justified by our first result. Proposition 1 shows that for any trigger strategy equilibrium in which each party i chooses *initiate* with a given probability, the optimal trigger equilibrium for both parties has initiation for the set of status quos that are most distant from their ideal points. Thus the equilibria identified by any pair $(h_l,\ h_r)$ are Pareto optimal for fixed initiation probabilities. The result is a consequence of the concavity of the parties' utility functions. The proof of this and all subsequent results can be found in the Appendix.

Proposition 1. Optimal Equilibrium Form. *For any possible equilibrium probability of initiation, the optimal equilibrium is for each party i to initiate if and only if q_t is sufficiently distant from z_i.*

We now derive payoffs for an arbitrary equilibrium defined by a pair $(h_l,\ h_r)$. By the uniform distribution of q_t, the probability of q_t falling below some threshold h is $\Pr\{q_t \leq h\} = (h - z_l)/(z_r - z_l)$. The equilibrium strategies then imply that the stage game expected utility of party i when parties L and R are in power are respectively:

$$\begin{aligned} U_i(h_l) &= \frac{h_l - z_l}{z_r - z_l}\int_{z_l}^{h_l} u_i(q_t, z_i)\frac{1}{h_l - z_l}\,\mathrm{d}q_t + \frac{z_r - h_l}{z_r - z_l}\int_{h_l}^{z_r} U_{i,l}(z_i, q_t, \phi)\frac{1}{z_r - h_l}\,\mathrm{d}q_t \\ &= \alpha - \left(z_i - \frac{z_l + z_r}{2}\right)^2 - \frac{(z_r - z_l)^2}{12} - \frac{\phi(z_r - h_l)[(2\phi + 6z_i) - 3(h_l + z_r)]}{6(z_r - z_l)} \qquad (4) \\ U_i(h_r) &= \frac{h_r - z_l}{z_r - z_l}\int_{z_l}^{h_l} U_{i,r}(z_i, q_t, \phi)\frac{1}{h_r - z_l}\,\mathrm{d}q_t + \frac{z_r - h_r}{z_r - z_l}\int_{h_r}^{z_r} u_i(q_t, z_i)\frac{1}{z_r - h_r}\,\mathrm{d}q_t \\ &= \alpha - \left(z_i - \frac{z_l + z_r}{2}\right)^2 - \frac{(z_r - z_l)^2}{12} + \frac{\phi(z_l - h_r)[(2\phi - 6z_i) + 3(h_r + z_l)]}{6(z_r - z_l)}. \qquad (5) \end{aligned}$$

Prior to an election, party i's expected stage game payoff given h_l and h_r is then:

$$V_i(h_l, h_r) = p_l U_i(h_l) + (1 - p_l) U_i(h_r). \qquad (6)$$

Using this notation, the expected payoff from a single period of the stage game is $V_i(z_l, z_r)$. This value will be used to calculate payoffs after deviations from cooperative play. It is clear that a party would always initiate once in the punishment phase.

What conditions can sustain an equilibrium in which players sometimes do not initiate a policy process? To begin, parties obviously have no incentive to choose the status quo when the equilibrium allows them to initiate: initiation results in a strictly better policy as well as the maintenance of cooperative play. What remains is to check whether parties can opportunistically initiate when norms dictate staying with the status quo. Given q_t, the condition for party i to choose the status quo according to the cooperative norm instead of initiating a policy is:

$$(1-\delta)u_i(q_t, z_i) + \delta V_i(h_l, h_r) \geq (1-\delta)U_{i,i}(z_i, q_t, \phi) + \delta V_i(z_l, z_r). \tag{7}$$

The incentive compatibility condition (7) allows us to derive the conditions that sustain cooperative norms defined by (h_l, h_r). The critical compatibility condition is whether parties are willing to maintain the status quo when $q_t = h_i$. This is because each party i has the greatest incentive to initiate policy when $q_t = h_i$, as the benefit of initiation declines as the status quo becomes closer to z_i.

It will be convenient to describe the conditions for an equilibrium in terms of election probability p_l. Let p_l^i denote the value of p_l such that party i is indifferent between cooperative play and defection (in the sense of initiating out of equilibrium) at $q_t = h_i$. Since each party's expected equilibrium payoffs $V_i(h_l, h_r)$ are linear in p_l, L prefers to initiate at $q_t = h_l$ for $p_l > p_l^l$, and R prefers to initiate at $q_t = h_r$ for $p_l < p_l^r$. Intuitively, better election prospects ensure more future opportunities to initiate policy and thereby improve the long-run payoff of defection.

The expressions for p_l^l and p_l^r can be presented simply when party ideal points are symmetric ($z_r = -z_l$). Performing the substitution and solving (7) for p_l produces:

$$p_l^l = \frac{-36(\delta-1)h_l z_l - \delta(3h_r+\phi)^2 + 6z_l(4\delta\phi + 3\delta h_r - 2\phi) + 9(7\delta-4)z_l^2}{3\delta\left[3h_l^2 - 2h_l(3z_l+\phi) - 3h_r^2 + h_r(6z_l - 2\phi) + 6z_l(2z_l+\phi)\right]} \tag{8}$$

$$p_l^r = \frac{(3h_r + 3z_l + \phi)(\delta(3h_r+\phi) + (12-9\delta)z_l)}{3\delta\left[3h_r^2 + 2h_r(3z_l+\phi) - 3h_l^2 + h_l(2\phi - 6z_l) + 6z_l(2z_l+\phi)\right]}. \tag{9}$$

For given values of δ, h_i, z_l, and ϕ, a cooperative equilibrium therefore exists if and only if $p_l \in [p_l^r, p_l^l]$.

Proposition 2 uses expressions (8) and (9) to provide comparative statics on the bounds p_l^r and p_l^l. Part (i) confirms the standard repeated game logic that the range of election probabilities that can support any given (h_l, h_r) pair expands as δ increases. Next, part (ii) addresses asymmetric norms. An equilibrium that gives one side greater ability to initiate policy is supported by an electoral advantage. Intuitively, a higher probability of election raises both an incumbent's temptation to initiate and the opposition's incentive to continue cooperation.

Proposition 2. Comparative Statics on Cooperative Equilibria. *Let $z_r = -z_l$. Then:*

(i) p_l^l (p_l^r) is increasing (decreasing) in δ.

(ii) p_l^l and p_l^r are decreasing in h_l, h_r.

(iii) When $h_r = -h_l$, p_l^l (p_l^r) is increasing (decreasing) in ϕ for $d(z_l, h_l) > (\leq)\ d(z_l, \hat{h}_l)$, where

$$\hat{h}_l = \frac{2\delta\phi - \sqrt{\delta^2\phi^2 + 36(\delta-1)^2 z_l^2 + 24\delta(\delta-1)z_l\phi} + (9\delta - 6)z_l}{3\delta}. \tag{10}$$

Perhaps more surprisingly, part (iii) shows that the range of cooperation can both contract and expand with flexibility. To see how this happens, observe that ϕ affects parties' incentives to cooperate via two channels. On one hand, "permissive norms" — i.e., when $d(z_l, h_l) < d(z_l, \hat{h}_l)$ — can make cooperation harder to sustain because of the increased payoff from deviation. When this is the case, the range of cooperation can shrink as flexibility increases. On the other, higher flexibility also implies a worse continuation value from defection because of concave preferences and the stream of future policy initiations by the opposition. This second effect prevails when political norms are "stringent" enough — dictating that parties usually leave the status quo alone.

Figure 2 presents examples of the interval $[p_l^r, p_l^l]$ of election probabilities that support norms (h_l, h_r) as a function of ϕ. As Proposition 2(i) shows, the interval $[p_l^r, p_l^l]$ always expands as δ increases. In the left panel, norms are symmetric and each party's initiation threshold h_i is distant from its ideal policy, which means that in equilibrium both parties leave the status quo alone for most values of q_t. This meets the condition of Proposition 2(iii) and as a result p_l^l increases in ϕ and p_l^r decreases in ϕ.[6]

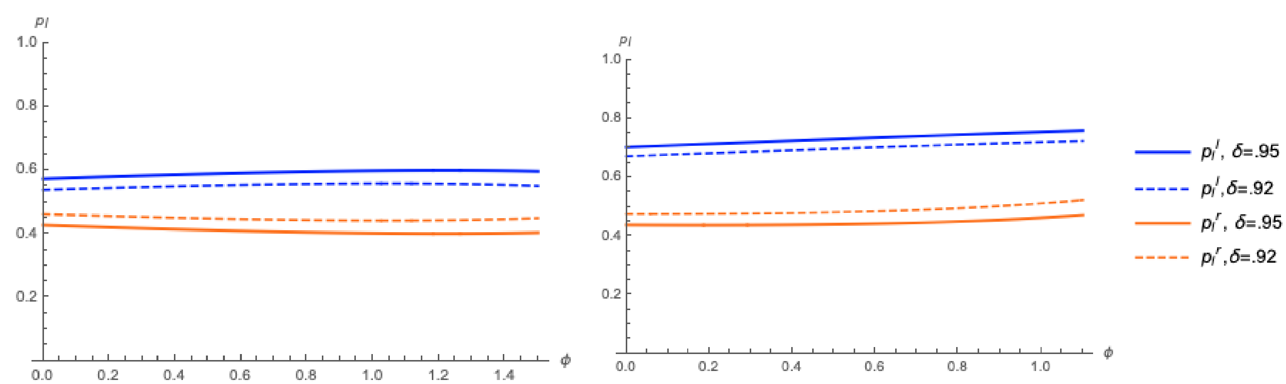

Figure 2 – Values of p_l supporting equilibria as a function of ϕ, with $z_l = -1$, $z_r = 1$. Norms thresholds are set to $h_l = .5$ and $h_r = -.5$ (left panel) and $h_l = .1$ and $h_r = -.4$ (right panel). The region between the dashed lines shows the values of p_l supporting equilibria when $\delta = .92$. As δ increases, so does the cooperation region.

In the right panel in Figure 2, norms are asymmetric. Party R is disadvantaged under this norm

[6] Moreover, as norms restrain cooperation even more, approaching the opposite party's bliss point, cooperation only becomes possible for ϕ (and δ) high enough.

because it shows greater restraint and initiates less often than party L. In this case, both p_l^l and p_l^r are increasing in ϕ: party L is willing to cooperate for a larger set of p_l values because of the increased benefits from cooperation (initiation possible for more realizations of the status quo). By contrast, party R is less willing to cooperate because of the initiation power of its rival. Hence, an equilibrium that advantages party L can only be sustained when L enjoys an electoral advantage.

Optimal Equilibria

What are the political norms that maximize parties' total welfare? Can they be sustained in equilibrium? The efficient norms – which may not be sustainable in equilibrium — can be obtained by solving the following problem:

$$\max_{h_l,h_r} \quad p_l\big[U_l(h_l)+U_r(h_l)\big]+(1-p_l)\big[U_l(h_r)+U_r(h_r)\big]. \tag{11}$$

Our next result reports properties of the optimal norms.

Lemma 2. Optimal Norms. *(i) For a fixed ϕ, the welfare-maximizing norm is defined by the following thresholds:*

$$h_l^* = \frac{z_l+z_r}{2}+\frac{\phi}{3}, \tag{12}$$
$$h_r^* = \frac{z_l+z_r}{2}-\frac{\phi}{3}. \tag{13}$$

(ii) The welfare-maximizing amount of flexibility equals half of the policy space, i.e.

$$\phi^* = \frac{z_r-z_l}{2}. \tag{14}$$

Lemma 2 shows that the first-best solution requires significant constraint by parties. The result is due to the concavity of utility functions: restraining from initiating a policy process when the status quo is further away implies a lower expected cost suffered as a consequence of the other party's policy initiation. As institutional checks and balances increase, the welfare maximizing norm gives more leeway to parties for initiation: as $\phi \to 0$, both (h_l^*, h_r^*) converge to the midpoint the policy space. Conversely, an increase in institutional flexibility reduces the initiation power of parties that maximize total welfare.[7]

[7] Notice that if we relaxed Assumption 1 and allow for the possibility that $h_l < z_l+\phi$ or $h_r > z_r-\phi$, then the optimal norm $(\tilde{h}_l, \tilde{h}_r)$ would be more stringent than the interior case:

$$\tilde{h}_l = \frac{z_l+3z_r}{4}$$

Part (ii) of Lemma 2 derives the welfare-maximizing level of institutional flexibility under the optimal norm, which parties might conceivably choose if they could write the constitutional rules of the game. This is derived simply by maximizing total welfare (refeq: problem welfare max) with respect to the parameter ϕ. Intuitively, more checks and balances ($\phi < \phi^*$) makes cooperation harder to sustain and the expected value of defection more appealing; on the other hand, too much flexibility harms parties in a punishment phase because of increased losses from initiation by the opposition. As a consequence, intermediate levels of flexibility are welfare improving.

We now ask whether these thresholds are sustainable in equilibrium – i.e., whether the incentive compatibility condition (7) is satisfied at $q_t = h_i^*$:

$$(1-\delta)u_i(h_i^*, z_i) + \delta V_i(h_l^*, h_r^*) \geq (1-\delta)U_{i,i}(z_i, q_t, \phi) + \delta V_i(z_l, z_r). \tag{15}$$

Solving for p_l for both parties produces the following threshold levels of p_l^{l*}, p_l^{r*} that support an equilibrium:

$$p_l^{l*} = \frac{3(7\delta - 4)(z_r - z_l) - 4\delta\phi}{4\delta(3z_r - 3z_l - \phi)}, \tag{16}$$

$$p_l^{r*} = -\frac{3(3\delta - 4)(z_r - z_l)}{4\delta(3z_r - 3z_l - \phi)}. \tag{17}$$

Analogously with expressions (8) and (9), for given values of δ and z_i, a welfare-maximizing cooperative equilibrium exists if and only if $p_l \in [p_l^{r*}, p_l^{l*}]$.

The thresholds (16) and (17) depend on $z_r - z_l$, which conveniently expresses the level of ideological polarization. This allows us to characterize the effects of polarization on norms. As the next result shows, increasing polarization makes optimal norms *easier* to uphold, in the sense of expanding the set of election probabilities $[p_l^{r*}, p_l^{l*}]$ that support an equilibrium. However, the optimal norms produce lower payoffs in equilibrium.

Proposition 3. Polarization. *(i) p_l^{r*} is decreasing and p_l^{l*} is increasing in $z_r - z_l$.*

(ii) Welfare under (h_l^*, h_r^*) is decreasing in $z_r - z_l$.

Proposition 3 conveys a simple intuition about the nature of political norms. The consequences of opportunistic policy initiation become progressively worse as polarization increases. This raises the

---

$$\tilde{h}_r = \frac{3z_l + z_r}{4}.$$

Intuitively, in the corner case norms are less permissive because initiation is more harmful to parties due to the concavity assumption.

demand for constraining norms, and thus the broader electoral conditions that support optimal norms. However even under optimal norms, polarization increases the probability of policy initiation and reduces a party's payoffs when it is out of power. While our analysis cannot capture changing norms within an equilibrium, the result is suggestive of how norms evolve. The threat of lower aggregate payoffs increases the scope for adopting the most mutually beneficial norms, but such norms can only partially offset the losses.

Figure 3 illustrates optimal norms as a function of electoral prospects. When $z_r = -z_l$ the equilibrium is symmetric, in the sense that both parties initiate a policy process for a range of the same length, given that $h_l^* = \phi/3 = -h_r^*$. The first panel plots (h_l^*, h_r^*) subject to the incentive compatibility condition (15) for both parties, when $z_l = -z_r$ and $\phi = 0.3$. For these parameters, the unconstrained welfare maximizing thresholds are $h_l^* = 0.1 = -h_r^*$, which correspond to the values parallel to the x axis when $p_l \in \left[p_l^{r*}, p_l^{l*}\right]$. The simulation shows that, for values of the discount factor high enough and for $p_l \in \left[p_l^{r*}, p_l^{l*}\right]$, the unconstrained welfare maximizing norms are sustained in equilibrium by both parties.

How does the efficient equilibrium change when parties' bliss point are not symmetric around zero? In this case, $h_l^* \neq h_r^*$ and the welfare maximizing norms (h_l^*, h_r^*) shift towards the more extreme party's preferred policy. The second panel in Figure 3 plots (h_l^*, h_r^*) subject to the incentive compatibility condition (15) for both parties, when $\phi = 0.3$ and $z_l > -z_r$. Without changing the probability of being in power, more extreme policy preferences simply tilt the optimal norms towards them.

5. Political Norms as Electoral Equilibria

Thus far we have treated the probability of victory as exogenous, which simplifies the analysis. We now introduce voters, thus microfounding the probability of victory as the outcome of a voting process. This allows us to derive electoral conditions under which cooperative norms are sustained in equilibrium, and how optimal norms change as a function of the electorate's preferences.

The game is unchanged from that of Section 3, with the exception of voters that determine election results. We continue to focus on the interior case: flexibility ϕ is low enough to constrain policies to within $(z_l,\, z_r)$ in equilibrium. There exists a continuum of voters, indexed by v, who vote for L or R: voter v's ideal point is denoted by $z_v \in \mathcal{Z}$. Each voter's utility function takes the same form as that of the parties (1). For an implemented policy $x_t \in \mathcal{Z}$, voter v thus receives $u_v(x_t, z_v) = \alpha - (x_t - z_v)^2$. We denote by m the median voter.

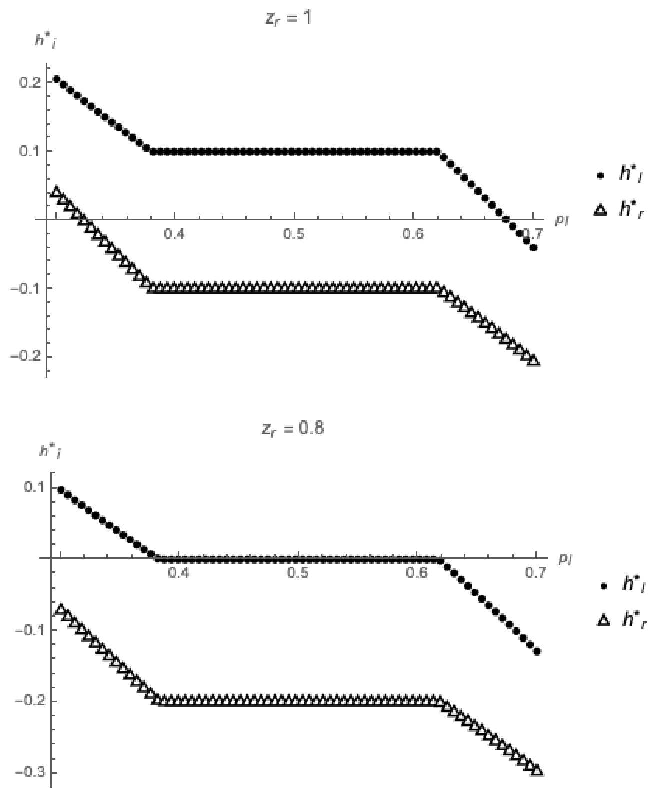

Figure 3 – Welfare maximizing norms (h_l^*, h_r^*) sustainable in equilibrium for $p_l \in \left[p_l^r, p_l^l\right]$, $\phi = 0.3$ and $\delta = 0.9$. Parties' bliss points set at $z_l = -1$ and $z_r = 1$ in the first panel, and at $z_l = -1$ and $z_r = 0.8$ in the second panel.

Before the election, an exogenous shock ξ that favors party R affects all voters equally, where ξ is uniformly distributed in $[-\frac{1}{2\psi}, \frac{1}{2\psi}]$. We assume that ψ is small enough to guarantee interior election probabilities for both parties. Since no voter is ever pivotal, we adopt the standard assumption that voters vote sincerely in each election; thus, the median voter is decisive in each election. The analysis proceeds largely as before, with the exception that election probabilities in the cooperative and punishment phases are now determined by the median voter's assessment of the parties' expected policy choices in each. Clearly, the stage game equilibrium in which parties always initiate remains the optimal strategy in the punishment phase, as well as an equilibrium in the electoral game.

To see when cooperation under non-trivial norms is possible, we start with the voters' decisions, which determines the probability of each party winning the election. Given this winning probability, we derive the expected payoff of each party as a function of norms, and parties' decision to initiate or cooperate. Voter utility is affected by political norms and institutional constraints, as these affect the distribution of policies. Formally, voter v's stage game expected utility when parties L and R are in power is respectively:

$$
\begin{aligned}
U_v(h_l) &= \frac{h_l - z_l}{z_r - z_l} \int_{z_l}^{h_l} u_v(q_t, z_v) \frac{1}{h_l - z_l} \, \mathrm{d}q_t + \frac{z_r - h_l}{z_r - z_l} \int_{h_l}^{z_r} U_{v,l}(z_v, q_t, \phi) \frac{1}{z_r - h_l} \, \mathrm{d}q_t \\
U_v(h_r) &= \frac{h_r - z_l}{z_r - z_l} \int_{z_l}^{h_l} U_{v,r}(z_v, q_t, \phi) \frac{1}{h_r - z_l} \, \mathrm{d}q_t + \frac{z_r - h_r}{z_r - z_l} \int_{h_r}^{z_r} u_v(q_t, z_v) \frac{1}{z_r - h_r} \, \mathrm{d}q_t
\end{aligned}
$$

Analogously to equations (2) and (3), $U_{v,l}(z_v, q_t, \phi)$ and $U_{v,r}(z_v, q_t, \phi)$ are v's expected utilities from initiating under parties L and R respectively, given status quo q_t. The expressions for $U_v(h_l)$ and $U_v(h_r)$ are exactly those in equations (4) and (5), with the voter's ideal point z_m substituted in for the party i ideal point z_i.

The probability that L wins the election is the probability that the median voter prefers L. Under the assumed uniform distribution and norms (h_l, h_r), this evaluates to:

$$
\begin{aligned}
p_l(h_l, h_r, z_m) &= \frac{1}{2} + \psi[U_v(h_l) - U_v(h_r)] \\
&= \frac{1}{2} + \frac{\psi\phi}{2} \left[z_l + z_r - 2z_m + \frac{3(h_l - h_r)(h_l + h_r - 2z_m) - 2\phi(h_l + h_r - z_l - z_r)}{3(z_l - z_r)} \right]. \quad (18)
\end{aligned}
$$

Similarly, the probability of election in a punishment phase is:

$$
p_l(z_l, z_r, z_m) = \frac{1}{2} + \psi\phi \left(1 + \frac{\phi}{2z_l - 2z_r} \right) (z_l + z_r - 2z_m). \quad (19)
$$

Hence, the election probability depends on the policy distributions implied by the norms, and these will be different if $h_i \neq z_i$.

Each party i's expected stage game payoff can then be written as:

$$
V_i(h_l, h_r) = p_l(h_l, h_r, z_m) U_i(h_l) + (1 - p_l(h_l, h_r, z_m)) U_i(h_r). \quad (20)
$$

The critical incentive compatibility condition to maintain the status quo is evaluated at $q_t = h_i$, where party i has the greatest incentive to initiate policy — i.e. when condition (7) is satisfied:

$$
(1 - \delta) u_i(q_t, z_i) + \delta V_i(h_l, h_r) \geq (1 - \delta) U_{i,i}(z_i, q_t, \phi) + \delta V_i(z_l, z_r),
$$

where norms affect the parties' expected utilities directly, through $U_i(h_l)$, and indirectly, through the electoral decision which alters their winning probabilities.

In what follows we describe the conditions for an equilibrium in terms of the location of the median voter bliss point, z_m. Let z_m^i denote the value of z_m such that party i is indifferent between cooperating (following the norms) and initiating out of equilibrium at $q_t = h_i$. We first show that the difference between the expected payoff from cooperation and defection is strictly increasing (decreasing) in z_m for party L (R).

Remark 1. *Let Δ_i be the difference between the expected payoff from cooperation and defection for party i, i.e.:*

$$\Delta_i = (1-\delta)\left(u_i(q_t, z_i) - U_{i,i}(z_i, q_t, \phi)\right) + \delta\left(V_i(h_l, h_r) - V_i(z_l, z_r)\right).$$

For all the parameter values the difference Δ_l is increasing in z_m and Δ_r is decreasing in z_m.

As the median voter location moves away from a party-preferred policy, the higher expected loss resulting from initiation by the opposition makes cooperation more appealing. It follows from Remark 1 that L prefers to initiate at $q_t = h_l$ for $z_m < z_m^l$, and R prefers to initiate at $q_t = h_r$ for $z_m > z_m^r$. For given values of δ, h_i, z_i, ϕ and ψ, a cooperative equilibrium therefore exists if and only if $z_m \in [z_m^l, z_m^r]$.

How do the bounds z_m^l, z_m^r change with the parameters of the model? As standard repeated game intuition would suggest, the interval $[z_m^l, z_m^r]$ increases with δ. Figure 4 plots this interval for symmetric norms (h_l, h_r) as a function of flexibility ϕ. There are two cases, one in which norms allow initiation for a wide range of status quo realizations ($h_l = -h_r = 0.3$), and one in which norms are more stringent ($h_l = -h_r = 0.5$). In both cases, z_m^l (z_m^r) increases (decreases) in ϕ: as institutional flexibility increases both parties cooperate over a smaller range of z_m values. Conversely, more checks and balances facilitate cooperation. Fixing ϕ, the range of z_m values for which both parties cooperate *decreases* with the strictness of norms. As norms allow initiation for less status quo realizations, the expected value of cooperation decreases while that of defection remains the same (given ϕ). As a result, parties are less willing to cooperate. The magnitude of the change in the cooperation interval depends on ϕ: at high values parties are slightly more willing to cooperate under permissive norms than under strict norms, but at low values they become substantially more willing to cooperate.

Finally, Figure 5 plots the interval $[z_m^l, z_m^r]$ when norms are asymmetric. In particular, party L is advantaged in this example, initiating more often ($h_l = 0.1$, $h_r = -0.4$). As the values of z_m sustaining cooperation show, an equilibrium that advantages party L can only be sustained when L enjoys an electoral advantage.

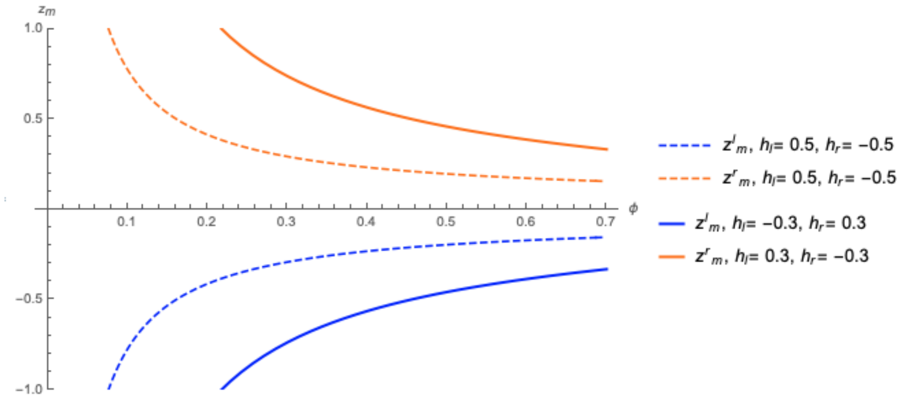

Figure 4 – Symmetric Norms. Values of z_m supporting equilibria as a function of ϕ, with $z_l = -1$, $z_r = 1$, $\delta = 0.95$ and $\psi = .4$. Norms thresholds are set to $h_l = -0.3$ and $h_r = 0.3$ (thick lines) and $h_l = 0.5$ and $h_r = -0.5$ (dashed lines). The region between the thick (dashed) lines shows the values of z_m supporting equilibria when $h_l = -0.3$ and $h_r = 0.3$ ($h_l = 0.5$ and $h_r = -0.5$).

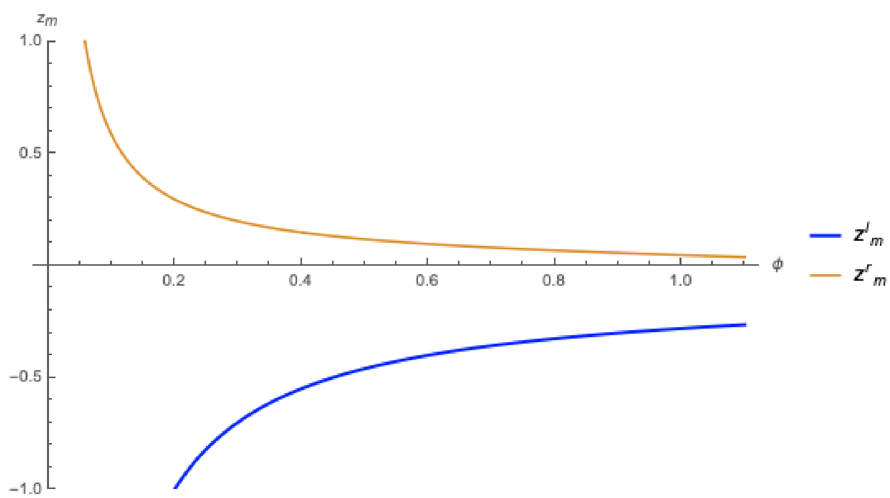

Figure 5 – Asymmetric Norms. Values of z_m supporting equilibria as a function of ϕ, with $z_l = -1$, $z_r = 1$, $\delta = 0.95$ and $\psi = .4$. Norms thresholds are set to $h_l = 0.1$ and $h_r = -0.4$. The region between the lines shows the values of z_m supporting equilibria when $h_l = 0.1$ and $h_r = -0.4$.

Optimal Norms

The next result derives the optimal norms from the median voter's perspective. As with the analogous Lemma 2, we ignore for the moment whether parties are willing to follow them in equilibrium. A significant difference between the voter's welfare and the parties' welfare analyzed in the previous section is that the former is not necessarily concave; in some cases, the optimal norms are located at a corner. However when electoral uncertainty is sufficiently high and institutional flexibility is sufficiently low, the objective is concave and a simple interior solution is obtainable. These bounds are not in practice too constraining. The subsequent numerical examples illustrate optimal interior equilibria where concavity is maintained.[8]

Lemma 3. *For ϕ, ψ sufficiently small, the voter welfare maximizing norm is defined by the following thresholds:*

$$h_l^* = z_m + \frac{\phi}{3} \tag{21}$$

$$h_r^* = z_m - \frac{\phi}{3}. \tag{22}$$

The optimal norms (h_l^*, h_l^*) do not depend on the parties' ideologies or election probabilities: the voter simply wants both parties to treat status quo points symmetrically with respect to her own ideal point. Note also that under the assumed conditions they also coincide with the optimal norms that maximize parties' total welfare in the baseline model when the median voter lies exactly in the middle of the policy space.

As expressions (21) and (22) in Lemma 3 clearly show, the optimal norms require symmetric behavior from parties when their ideologies are symmetric around z_m. When party ideologies are asymmetric, the norms tend to benefit the advantaged party, by letting it initiate more often. Proposition 4 demonstrates how norms affect election probabilities. The first part conveys the main intuition, which is that voter-optimal norms place more constraints on expected winners. When $z_m > (z_l + z_r)/2$, the median voter favors party R, and is more likely to elect it in the stage game. Under (h_l^*, h_r^*), party L remains disadvantaged but not by as much as under the stage game. By initiating infrequently, it keeps the median voter closer to indifferent between the two parties. Thus, norms can indirectly maintain electoral competitiveness.

The second part of the proposition shows that the size of this moderating effect is increasing in ϕ. In other words, optimal norms allow high institutional flexibility to offset electoral imbalances. To see why, observe that when $\phi = 0$, both parties essentially do nothing, as initiation and keeping the status quo achieve the same result. In this case, both parties are elected with probability $1/2$. As flexibility increases, the advantaged party can use initiation to secure a large electoral advantage, but by roughly equalizing behavior, norms reduce this advantage. Note, however, that this reduction is relative to the advantaged party's stage game probability of election: the disadvantaged party's probability of election might still decrease in ϕ.

[8] Unfortunately the complexity of the objective makes a characterization of the optimal level of institutional flexibility difficult.

The third part of the result establishes the consequence of increasing polarization. To do this, it is first necessary to define a working notion of polarization. We say that polarization increases if z_l decreases and z_r increases by the same amount, thus holding $z_l + z_r$ constant. This ensures that any increase in polarization does not change the identity of the advantaged party, and allows us to focus exclusively on the level of divergence $z_r - z_l$ between party platforms. Because of concave utility, polarization inevitably reduces the median voter's assessment of the unfavored party relative to the favored party. It follows that the favored party's probability of election increases, even as (by Proposition 4(i)) the optimal norms mute the extent of the increase.

Proposition 4. Election Probabilities. *Under the optimal norms,*

(i) The advantaged party is elected with lower probability than in the stage game: $p_l(h_l^*, h_r^*, z_m) >$ $(<)$ $p_l(z_l, z_r, z_m)$ *for* $z_m > (<)$ $(z_l + z_r)/2$*.*

(ii) Institutional flexibility reduces electoral imbalances: $|p_l(h_l^*, h_r^*, z_m) - p_l(z_l, z_r, z_m)|$ *is increasing in* ϕ*.*

(iii) As polarization increases, the probability of election of the advantaged party increases.

Finally, we ask how institutional checks and balances affect the sustainability of optimal norms in equilibrium. Because of the complexity of the expressions, we resort to numerical results. Figure 6 plots the region for which both parties are willing to cooperate under the optimal norms (h_l^*, h_r^*), as a function of ϕ (horizontal axis) and z_m (vertical axis). Parties' bliss points are set to $z_l = -1$, $z_r = 1$. The orange region represents all the points such that R prefers cooperation to defection, while the blue one is the corresponding region for L. The area where the two colored regions overlap represents all values of z_m and ϕ such that cooperation is incentive compatible for both parties under the optimal norms (h_l^*, h_r^*).

The first plot of Figure 6 shows that when the median voter is located in the middle of the policy space cooperation is sustainable: the incentive compatibility condition does not bind for either party for any value of ϕ. Intuitively, this region expands for greater value of parties' discount factor, as the second plot of Figure 6 indicates.

Figure 6 also shows that electoral imbalance can make optimal norms harder to uphold. As z_m increases, party R becomes relatively more advantaged. As a consequence, the expected value from cooperation of party R increases, while that of party L is reduced — making L's incentive compatibility constraint more binding. In other words, as the median voter moves away from the center, the electorally advantaged party benefits more, and is more likely to abide by, optimal norms, while the disadvantaged party is less likely to cooperate. This happens because optimal norms give an initiation advantage to R, which implies that cooperation becomes harder to sustain from the electorally disadvantaged party L, which deviates from the optimal norms for high values of z_m.

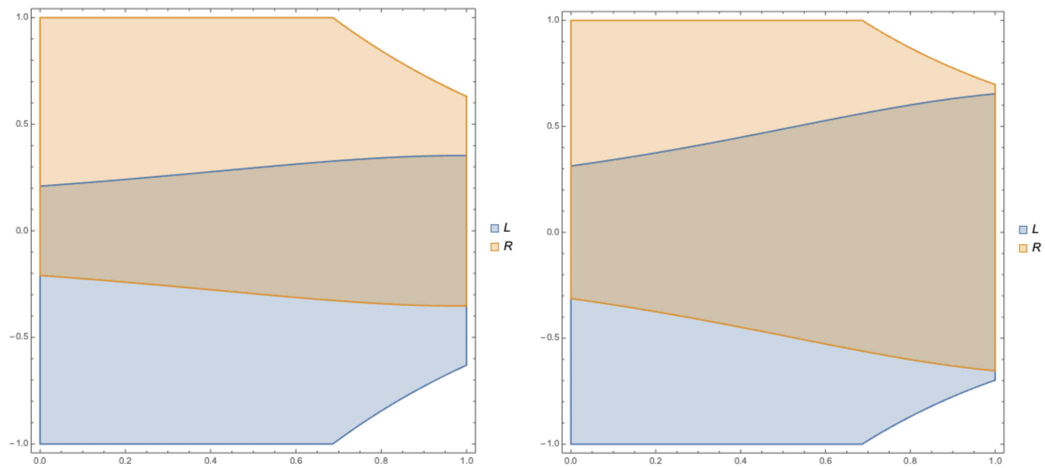

Figure 6 – Optimal Norms. Values of ϕ (horizontal axis) and z_m (vertical axis) supporting cooperation under the optimal norms (h_l^*, h_r^*). In the orange region, cooperation is incentive compatible for party R. In the blue region, cooperation is incentive compatible for party L. Parties bliss points are set to $z_l = -1$, $z_r = 1$, and $\psi = .3$. The discount factor is set at $\delta = 0.9$ in the first plot, and $\delta = 0.95$ in the second plot.

How does electoral imbalance interact with checks and balances? Both plots of Figure 6 show that an increase in institutional flexibility makes the disadvantaged party more likely to cooperate for certain values of z_m. Suppose that the median voter is located at $z_m = 0.5$ in the second plot. While for low values of institutional flexibility party L is not willing to cooperate, for higher values of ϕ cooperation becomes incentive compatible. Intuitively, higher flexibility implies a worse continuation value from defection because of concave preferences and the stream of future policy initiations by the advantaged opposing party. This expected loss is mitigated by the optimal norms, which reduce the reach of the advantaged party. Thus, an increase in institutional flexibility can make optimal norms *easier* to uphold. Conversely, more checks and balances can decrease cooperation under the optimal norms.

6. Conclusion

In any political system, policy outcomes are the result of formal institutions and informal norms. Within the context set by commonly-studied features such legislative organization, executive authority, judicial review, and bureaucratic delegation, politicians still face the temptation to push legal and constitutional limits. This paper is an initial attempt to capture the confluence of norms and institutions. We formalize the notion of political norms of cooperation among different actors who have the power to initiate policy change and are constrained by existing institutions. Norms are defined as equilibrium constraints on parties' opportunistic behavior which help preserving the status quo. This framework allows us to assess the political and institutional conditions under which norms can produce beneficial outcomes over time.

Our baseline model produces three main findings. First, institutions that provide more policy-making leeway to the party in power do not necessarily hinder the preservation of the status quo. In fact, the model shows that the opposite can be true: political cooperation can be easier to sustain in political systems with less checks and balances. When institutional flexibility increases, so does the "temptation" payoff from policy initiation. Cooperation then becomes more sustainable because of the higher expected loss from the stream of future policies initiated by the opposition. A higher probability of election raises both an incumbent's temptation to initiate and the opposition's incentive to continue cooperation; taken together, these opposing incentive could result in asymmetric norms. Finally, by increasing the stakes of a breakdown of norms, ideological polarization actually makes norms more sustainable, while also reducing each party's payoffs.

With small modifications, we also adapt model to an electoral setting, where election probabilities are determined endogenously by voter preferences and expected politician strategies. One of the main findings is that the voter's optimal norms require policy concessions from the disadvantaged party, which in turn results in the advantaged party winning less often than it would in a world with no cooperation. We also show that less checks and balances can make the disadvantaged party more willing to cooperate under optimal norms, which implies that voters can be better off under a more flexible institutional system.

While our model begins to define the popular notion of political norms, one of its most serious limitations is that norms do not break down on the equilibrium path. In light of current events, the fraying of norms and the establishment of new ones is perhaps the one of the most promising opportunities for future research.

7. Appendix

Proof of Proposition 1. Without loss of generality, we focus on payoffs when party L is in power. For a given equilibrium, let $\tilde{V}_i$ be its expected value and let $\Theta_l \subseteq \mathcal{Z}$ denote the set of status quos for which party L initiates. Also, let θ_l be the measure of Θ_l, and $\underline{\theta}_l = \inf \Theta_l$. We show that if Θ_l is sustainable in an equilibrium and there exists some $[q', q''] \subseteq \mathcal{Z} \setminus \Theta_l$ where $q' > \underline{\theta}_l$, then initiating when $q \in [z_r - \theta_l, z_r]$ is sustainable as an equilibrium as well. We begin by assuming that $\Theta_l \cap [z_l, z_l + \phi] = \emptyset$, i.e. initiating always results in an interior policy outcome.

The first step is to show that when L is in power, the relative value of initiation increases in q. For a party with ideal point z_i, define i's net utility gain from initiation given status quo q as:

$$\begin{aligned} \hat{u}_i(q) &= \int_{q-\phi}^{q} u_i(x, z_i) \frac{1}{\phi} \, \mathrm{d}x - u_i(q, z_i) \\ &= \phi(q - z_i) - \frac{\phi^2}{3}. \end{aligned}$$

Thus initiating at $q+\epsilon$ for any $\epsilon > 0$ produces a net benefit of $\phi(q+\epsilon-z_i)-\phi^2/3-\big(\phi(q - z_i) - \phi^2/3\big) = \epsilon\phi$ over initiating at q for both parties. Since $q' > \underline{\theta}_l$, there exists an interval $[\hat{q}', \hat{q}''] \in \Theta_l$ such that $\hat{q}'' < q'$ and $\hat{q}'' - \hat{q}' \leq q'' - q'$. The resulting set of status quos $[\hat{q}' + q'' - \hat{q}'', q''] \cup \Theta_l \setminus [\hat{q}', \hat{q}'']$ then produces a net benefit of:

$$\int_{[\hat{q}'+q''-\hat{q}'',q'']\cup\Theta_l\setminus[\hat{q}',\hat{q}'']} \frac{\hat{u}_i(q)}{z_r - z_l} \, \mathrm{d}q - \int_{\Theta_l} \frac{\hat{u}_i(q)}{z_r - z_l} \, \mathrm{d}q = \frac{q'' - \hat{q}''}{z_r - z_l} (\hat{q}'' - \hat{q}')\phi$$

for both parties, relative to Θ_l. This benefit is clearly positive, and thus the implied expected value for each party i is strictly higher than $\tilde{V}_i$. Iterating over all possible intervals $[q', q'']$, the set of status quos of measure θ_l that maximize expected payoffs is $[z_r - \theta_l, z_r]$. Denote these expected payoffs under this candidate equilibrium $\hat{V}_i$ for each party i, where clearly $\hat{V}_i > \tilde{V}_i$.

Next, we verify that changing L's strategy to initiating for $q \in [z_r - \theta_l, z_r]$ (holding R's fixed) remains an equilibrium. First consider the decision of party L. It is clear that for any such q, L will initiate as specified. For $q < z_r - \theta_l$, it is sufficient to verify that she is willing to choose sq at $q = z_r - \theta_l$. The condition for maintaining the status quo is:

$$(1 - \delta)u_l(z_r - \theta_l, z_l) + \delta\hat{V}_l > (1 - \delta)U_{l,l}(z_l, z_r - \theta_l, \phi) + \delta V_l(z_l, z_r). \tag{23}$$

Recall that in the original equilibrium party L facing $q = \overline{\theta}_l \equiv \sup \Theta_l \setminus \Theta_l$ maintained the status quo;

i.e.,

$$(1-\delta)u_i(\overline{\theta}_l, z_l) + \delta\tilde{V}_l > (1-\delta)U_{l,l}(z_l, \overline{\theta}_l, \phi) + \delta V_l(z_l, z_r). \tag{24}$$

Since $\hat{V}_l > \tilde{V}_l$, condition (23) holds if:

$$\begin{aligned} u_l(z_r-\theta, z_l) - u_l(\overline{\theta}_l, z_l) &> U_{l,l}(z_l, z_r-\theta_l, \phi) - U_{l,l}(z_l, \overline{\theta}_l, \phi) \\ U_{l,l}(z_l, \overline{\theta}_l, \phi) - u_l(\overline{\theta}_l, z_l) &> U_{l,l}(z_l, z_r-\theta, \phi) - u_l(z_r-\theta, y_L). \end{aligned}$$

The last statement holds because $\overline{\theta}_l > z_r - \theta_l$, and as established earlier, the increasing net gain from initiation is increasing in q.

For party R, the incentive to initiate as specified in the original equilibrium is obvious. The incentive to maintain the status quo as required is strictly stronger in the candidate equilibrium, since $\hat{V}_r > \tilde{V}_r$.

For the corner cases where $\Theta_l \cap [z_l, z_l+\phi] \neq \emptyset$, the argument is almost identical; observe that the relative gain from initiation is even lower for low values of q such that $q - \phi < z_l$. □

Proof of Proposition 2. The result is obtained simply by differentiating the expressions for p_l^l and p_l^r (equations (8) and (9)). We prove the first two claims of the proposition for general z_i and h_i, and assume symmetry for the last claim. We present the results for p_l^l; the results for p_l^r are derived analogously and are omitted.

(i) For δ we have:

$$\frac{\partial p_l^l}{\partial \delta} = \frac{2(z_l - z_r)(3z_l + \phi - 3h_l)}{\delta^2\left[3h_l^2 - 2h_l(3z_l+\phi) - 3h_r^2 + 6h_r z_l - 2h_r\phi + 3z_l^2 - 6z_l z_r + 5z_l\phi + 3z_r^2 - z_r\phi\right]}.$$

The numerator of this expression is positive because $\phi < h_l - z_l$ by assumption, and it is straightforward to verify that the denominator is non-negative.

(ii) For h_r we have:

$$\frac{\partial p_l^l}{\partial h_r} = \frac{2(3h_l - 3z_l - \phi)(3h_r - 3z_l + \phi)\left[\delta\phi - 3\delta h_l - 3(\delta-2)(z_l - z_r)\right]}{3\delta\left[-3h_l^2 + 2h_l(3z_l+\phi) + 3h_r^2 - 6h_r z_l + 2h_r\phi - 3z_l^2 + 6z_l z_r - 5z_l\phi - 3z_r^2 + z_r\phi\right]^2}$$

The denominator of this expression is clearly positive, and it is straightforward to verify the numerator is non-positive.

For h_l the expression is lengthier and hence omitted, but it is straightforward to verify that the derivative is is non-positive for any parameters such that $p_l^l > 0$.

(iii) We now show that, when $h_r = -h_l$, p_l^l (p_l^r) is increasing (decreasing) in ϕ for $d(z_l, h_l) > d(z_l, \hat{h}_l)$.

Differentiating the expressions for p_l^l (equation (8)) with respect to ϕ, we obtain:

$$\frac{\partial p_l^l}{\partial \phi} = -\frac{\delta\phi^2 + 3\delta h_l^2 - 2h_l(2\delta\phi + (9\delta - 6)z_l) + 3(5\delta - 4)z_l^2 + 4\delta z_l \phi}{18\delta z_l(-2h_l + 2z_l + \phi)^2}, \tag{25}$$

the sign of which depends on h_l. Solving $\partial p_l^l/\partial\phi = 0$ for ϕ yields the following stationary point:

$$\phi^* = 2h_l - 2z_l - \frac{\sqrt{\delta(h_l - z_l)(\delta h_l + (11\delta - 12)z_l)}}{\delta},$$

which is a local maximum as the second-order derivative evaluated at $\phi = \phi^*$ is negative. It is left to show the value of $\hat{h}_l$ s.t. $\partial p_l^l/\partial\phi > 0$ for $d(z_i, h_i) > d(z_l, \hat{h}_l)$.

We can obtain it by finding $\hat{h}_l$ s.t. $\partial p_l^l/\partial\phi|_{\hat{h}_l} = 0$, that is:

$$\hat{h}_l = \frac{2\delta\phi - \sqrt{\delta^2\phi^2 + 36(\delta-1)^2 z_l^2 + 24\delta(\delta-1)z_l\phi} + (9\delta - 6)z_l}{3\delta}.$$

Finally, notice that $\partial^2 p_l^l/\partial^2\phi|_{\hat{h}_l} > 0$, which completes the proof. $\square$

Proof of Lemma 2. (i) The efficient frontier of the game can be obtained by solving the following problem

$$\max_{h_l, h_r} \quad p_l\big[U_l(h_l) + U_r(h_l)\big] + (1 - p_l)\big[U_l(h_r) + U_r(h_r)\big]. \tag{26}$$

The first-order necessary conditions with respect to h_l, h_r are:

$$(h_l) \quad p_l \left(\frac{\partial}{\partial h_l}\left[\frac{h_l - z_l}{z_r - z_l}\int_{z_l}^{h_l} u_l(q, z_l)\frac{1}{h_l - z_l}\,\mathrm{d}q + \frac{z_r - h_l}{z_r - z_l}\int_{h_l}^{z_r} U_{l,L}(z_l, q, \phi)\frac{1}{z_r - h_l}\,\mathrm{d}q\right] \right. \tag{27}$$

$$\left. + \frac{\partial}{\partial h_l}\left[\frac{h_l - z_l}{z_r - z_l}\int_{z_l}^{h_l} u_r(q, z_r)\frac{1}{h_l - z_l}\,\mathrm{d}q + \frac{z_r - h_l}{z_r - z_l}\int_{h_l}^{z_r} U_{r,L}(z_r, q, \phi)\frac{1}{z_r - h_l}\,\mathrm{d}q\right]\right) = 0$$

$$(h_r) \quad (1 - p_l) \left(\frac{\partial}{\partial h_r}\left[\frac{h_r - z_l}{z_r - z_l}\int_{z_l}^{h_r} U_{l,R}(z_l, q, \phi)\frac{1}{h_r - z_l}\,\mathrm{d}q + \frac{z_r - h_r}{z_r - z_l}\int_{h_r}^{z_r} u_l(q, z_l)\frac{1}{z_r - h_r}\,\mathrm{d}q\right] \right. \tag{28}$$

$$\left. + \frac{\partial}{\partial h_r}\left[\frac{h_r - z_l}{z_r - z_l}\int_{z_l}^{h_r} U_{r,R}(z_r, q, \phi)\frac{1}{h_r - z_l}\,\mathrm{d}q + \frac{z_r - h_r}{z_r - z_l}\int_{h_r}^{z_r} u_r(q, z_r)\frac{1}{z_r - h_r}\,\mathrm{d}q\right]\right) = 0$$

which identify the following stationary points

$$
\begin{aligned}
h_l &= \frac{3z_l + 3z_r + 2\phi}{6} \\
h_r &= \frac{3z_l + 3z_r - 2\phi}{6}.
\end{aligned}
$$

To see whether the identified solution is a maximum, we inspect the Hessian of the system

$$
\mathbf{H} \equiv \begin{bmatrix} \frac{2p_l\phi}{z_r - z_l} & 0 \\ 0 & -\frac{2(p_l-1)\phi}{3(z_l - z_r)} \end{bmatrix}.
$$

Let $\Delta_k = \det \mathbf{H_k}$, the kth *principal minor* of $\mathbf{H}$. Evaluating Δ_1 at the stationary point identified by the first-order necessary condition (h_l) yields $2p_l\phi/(z_r - z_l)$, hence $\Delta_1 < 0$, as $z_l < 0 < z_r$. The other condition that needs to be satisfied is $\Delta_2 > 0$. Since

$$
\Delta_2 = -\frac{2(p_l - 1)\phi}{3(z_l - z_r)} \frac{2p_l\phi}{z_r - z_l}
$$

is always positive, the stationary points (h_l, h_r) identify a local maximum of the welfare maximization problem in Lemma 2.

We now analyse the corner case, that is $h_l < z_l + \phi$ and $h_r > z_r - \phi$. First, define the expected utility of party i from L and R's policy initiation as

$$
\begin{aligned}
\tilde{U}_{i,l}(z_i, q, \phi) &= \int_{z_l}^{q} u_i(x, z_i) \frac{1}{q - z_l}\, \mathrm{d}x, \\
\tilde{U}_{i,R}(z_i, q, \phi) &= \int_{q}^{z_r} u_i(x, z_i) \frac{1}{z_r - q}\, \mathrm{d}x,
\end{aligned}
$$

and the stage game expected utility of party i when parties L and R are in power respectively as

$$
\begin{aligned}
\tilde{U}_i(h_l) &= \frac{h_l - z_l}{z_r - z_l} \int_{z_l}^{h_l} u_i(q, z_i) \frac{1}{h_l - z_l}\, \mathrm{d}q + \frac{z_l + \phi - h_l}{z_r - z_l} \int_{h_l}^{z_l+\phi} \tilde{U}_{i,L} \frac{1}{z_l + \phi - h_l}\, \mathrm{d}q + \\
&\quad \frac{z_r - (z_l + \phi)}{z_r - z_l} \int_{z_l+\phi}^{z_r} \frac{U_{i,L}}{z_r - (z_l + \phi)}\, \mathrm{d}q \\
\tilde{U}_i(h_r) &= \frac{z_r - \phi - z_l}{z_r - z_l} \int_{z_l}^{z_r-\phi} U_{i,R} \frac{1}{z_r - \phi - z_l}\, \mathrm{d}q + \frac{h_r - (z_r - \phi)}{z_r - z_l} \int_{h_r}^{z_r} \frac{\tilde{U}_{i,R}}{h_r - (z_r - \phi)}\, \mathrm{d}q +
\end{aligned}
$$

$$\frac{z_r - h_r}{z_r - z_l} \int_{h_r}^{z_r} u_i(q, z_i) \frac{1}{z_r - h_r} \,\mathrm{d}q.$$

Similarly to the interior case, the first-order necessary conditions with respect to h_l, h_r are:

$$\begin{aligned}
&(h_l) \quad p_l \Big[\frac{\partial}{\partial h_l} \tilde{U}_l(h_l) + \frac{\partial}{\partial h_l} \tilde{U}_r(h_l)\Big] = 0 \\
&(h_r) \quad (1 - p_l) \Big[\frac{\partial}{\partial h_r} \tilde{U}_l(h_r) + \frac{\partial}{\partial h_r} \tilde{U}_r(h_r)\Big] = 0,
\end{aligned}$$

which identify the following stationary points

$$\begin{aligned}
\tilde{h}_l &= \frac{z_l + 3z_r}{4} \\
\tilde{h}_r &= \frac{3z_l + z_r}{4}.
\end{aligned}$$

The Hessian of the system is

$$\mathbf{H} \equiv \begin{bmatrix} \frac{p_l(8h_l - 5z_l - 3z_r)}{3(z_l - z_r)} & 0 \\ 0 & \frac{(p_l - 1)(8h_r - 3z_l - 5z_r)}{3(z_l - z_r)} \end{bmatrix}.$$

Let $\Delta_k = \det \mathbf{H_k}$, the kth *principal minor* of $\mathbf{H}$. Evaluating Δ_1 at the stationary point identified by the first-order necessary condition (h_l) yields $-p_l < 0$. The other condition that needs to be satisfied is $\Delta_2 > 0$. Since $\Delta_2 = p_l(1 - p_l)$, we have that the stationary points (h_l, h_r) identify a local maximum of the welfare maximization problem in the corner case.

(ii) The welfare maximizing ϕ solves

$$\max_{\phi} \quad p_l \big[U_l(h_l) + U_r(h_l)\big] + (1 - p_l) \big[U_l(h_r) + U_r(h_r)\big].$$

and the first-order necessary conditions with respect to ϕ is:

$$\frac{1}{6} \Big(-3h_l^2 p_l + 4h_l p_l \phi + 3h_r^2 (p_l - 1) + 4h_r (p_l - 1)\phi - 4\phi + 3\Big) = 0. \tag{29}$$

Substituting the values of the efficient h_l^*, h_r^* derived in Lemma 2 into Equation 29 yields two roots, $\phi^{'} = (z_r - z_l)/2$ and $\phi^{''} = 3(z_r - z_l)/2$.

The second-order condition evaluated at h_l^*, h_r^* is

$$-\frac{2(z_l - z_r + \phi)}{3(z_l - z_r)}, \tag{30}$$

which is clearly positive for $\phi^{''} = 3(z_r - z_l)/2$ and negative for $\phi^{'} = (z_r - z_l)/2$, the welfare maximizing solution.

In the corner case, the first-order necessary conditions with respect to ϕ yields two roots, $\tilde{\phi}^{'} = 0$ and $\tilde{\phi}^{''} = z_r - z_l$. The second-order condition evaluated at $\tilde{\phi}^{'}$ is negative, while at $\tilde{\phi}$ is positive. Hence, $\tilde{\phi}^{''}$ is the welfare maximizing solution in the corner case. □

Proof of Proposition 3. (i) Let $z = z_r - z_l$ denote polarization. Differentiating (16) and (17) with respect to z produces:

$$\frac{\partial p_l^{l*}}{\partial z} = \frac{3(4-3\delta)\phi}{4\delta(\phi - 3z)^2}.$$
$$\frac{\partial p_l^{r*}}{\partial z} = -\frac{3(4-3\delta)\phi}{4\delta(\phi - 3z)^2}.$$

Clearly $\frac{\partial p_l^{l*}}{\partial z} > 0$ and $\frac{\partial p_l^{r*}}{\partial z} < 0$.

(ii) Welfare can be expressed by substituting (4) and (5) into expression (11), obtaining:

$$2\alpha - \frac{2(z_r - z_l)^2}{3} + \frac{3\phi\left(h_l^2 p_l + h_r^2(1-p_l)\right) - \phi(z_l+z_r)(3h_l p_l + 3h_r(1-p_l) - 2p_l\phi) - 2\phi^2(h_l p_l - h_r(1-p_l)) + z_l\phi(3z_r - 2\phi)}{3(z_l - z_r)}.$$

Rewriting this by substituting in the optimal values of h_l and h_r from Lemma 2 produces:

$$\frac{12(6a - \phi^2)(z_r - z_l) + 9\phi(z_r - z_l)^2 - 24(z_r - z_l)^3 + 4\phi^3}{36(z_r - z_l)}.$$

The derivative of this expression with respect to $z_r - z_l$ is:

$$\frac{\phi}{4} - \frac{\phi^3}{9(z_r - z_l)^2} - \frac{4(z_r - z_l)}{3},$$

which is easily shown to be negative for $\phi > 0$ and $z_r - z_l > 0$. □

Proof of Remark 1. Let $z_r = -z_l$. Differentiating Δ_l and Δ_r with respect to z_m produces

$$\frac{\partial \Delta_l}{\partial z_m} = \frac{\delta\psi\phi^2 \left(2\phi\left(h_l^2 + 2z_l(h_l+h_r) - h_r^2 - 24z_l^2\right) + 12z_l(h_r-h_l)(h_r-2z_l) - 3(h_l-h_r)^2(h_l+h_r) - 72z_l^3 - 6z_l\phi^2\right)}{24z_l^2},$$

$$\frac{\partial \Delta_r}{\partial z_m} = \frac{\delta\psi\phi^2 \left(2\phi\left(h_l^2 + 2h_l z_l - (h_r-6z_l)(h_r+4z_l)\right) + 12z_l(h_r-h_l)(2z_l+h_l) - 3(h_l-h_r)^2(h_l+h_r) + 72z_l^3 + 6z_l\phi^2\right)}{-24z_l^2}$$

which are positive and negative respectively. The same is true for general z_r; expressions are omitted for their length. $\square$

Proof of Lemma 3. We focus on the interior case. The optimal norm h_l^* solves

$$\max_{h_l,h_r} \; p_l(h_l,h_r)U_m(h_l) + (1-p_l(h_l,h_r))U_m(h_r)$$

The first-order necessary condition with respect to h_l, h_r are:

$$\begin{aligned}
(h_l) \quad & \frac{\partial p_l(h_l,h_r)}{\partial h_l}\Bigg[\frac{h_l-z_l}{z_r-z_l}\int_{z_l}^{h_l} u_m(q_t,z_m)\frac{1}{h_l-z_l}\,\mathrm{d}q_t + \frac{z_r-h_l}{z_r-z_l}\int_{h_l}^{z_r} U_{m,l}(z_v,q_t,\phi)\frac{1}{z_r-h_l}\,\mathrm{d}q_t \\
& - \frac{h_r-z_l}{z_r-z_l}\int_{z_l}^{h_r} U_{l,R}(z_l,q,\phi)\frac{1}{h_r-z_l}\,\mathrm{d}q - \frac{z_r-h_r}{z_r-z_l}\int_{h_r}^{z_r} u_l(q,z_l)\frac{1}{z_r-h_r}\,\mathrm{d}q\Bigg] \\
& + p_l(h_l,h_r)\left(\frac{\partial}{\partial h_l}\left[\frac{h_l-z_l}{z_r-z_l}\int_{z_l}^{h_l} u_l(q,z_l)\frac{1}{h_l-z_l}\,\mathrm{d}q + \frac{z_r-h_l}{z_r-z_l}\int_{h_l}^{z_r} U_{l,L}(z_l,q,\phi)\frac{1}{z_r-h_l}\,\mathrm{d}q\right]\right) = 0, \\
(h_r) \quad & \frac{\partial p_l(h_l,h_r)}{\partial h_r}\Bigg[\frac{h_l-z_l}{z_r-z_l}\int_{z_l}^{h_l} u_m(q_t,z_m)\frac{1}{h_l-z_l}\,\mathrm{d}q_t + \frac{z_r-h_l}{z_r-z_l}\int_{h_l}^{z_r} U_{m,l}(z_v,q_t,\phi)\frac{1}{z_r-h_l}\,\mathrm{d}q_t \\
& - \frac{h_r-z_l}{z_r-z_l}\int_{z_l}^{h_r} U_{l,R}(z_l,q,\phi)\frac{1}{h_r-z_l}\,\mathrm{d}q - \frac{z_r-h_r}{z_r-z_l}\int_{h_r}^{z_r} u_l(q,z_l)\frac{1}{z_r-h_r}\,\mathrm{d}q\Bigg] \\
& + (1-p_l(h_l,h_r))\left(\frac{\partial}{\partial h_r}\left[\frac{h_r-z_l}{z_r-z_l}\int_{z_l}^{h_r} U_{l,R}(z_l,q,\phi)\frac{1}{h_r-z_l}\,\mathrm{d}q + \frac{z_r-h_r}{z_r-z_l}\int_{h_r}^{z_r} u_l(q,z_l)\frac{1}{z_r-h_r}\right.\right) = 0.
\end{aligned}$$

There are three possible solutions for each of h_l and h_r. We derive conditions under which the objective is concave, thereby ensuring a unique solution. The Hessian matrix is quite lengthy and is therefore omitted. After some manipulation the quadratic form simplifies to:

$$\frac{\psi\phi^2\left[6\left(h_l^2-h_r^2\right)(3(z_l-z_r)(z_l+z_r-2z_m)+2\phi(z_l+z_r))-27\left(h_l^2-h_r^2\right)^2+\left((3h_l-z_m)^2-(3h_r-z_m)^2-2\phi(h_l+h_r)\right)^2\right]}{18(z_l-z_r)^2}+\frac{\phi(h_l^2+h_r^2)}{2(z_l-z_r)}$$

This expression is clearly negative when either ϕ or ψ are sufficiently small. Under these conditions, the objective is concave and the unique values of h_l and h_r satisfying the first-order conditions are then $h_l^* = z_m + \phi/3$ and $h_r^* = z_m - \phi/3$. □

Proof of Proposition 4. The stage game election probability for party L is simply the punishment phase probability, or $p_l(z_l, z_r, z_m)$. Substituting the expressions for h_l^* and h_r^* (21) and (22) into the expression for $p_l(h_l, h_r, z_m)$ (18) produces:

$$p_l(h_l^*, h_r^*, z_m) = \frac{1}{2} + \psi\phi\left(\frac{z_l+z_r}{2} - z_m - \frac{\phi(z_l+z_r-2z_m)}{3(z_r-z_l)}\right) \tag{31}$$

Taking the difference between probabilities of victory $p_l(h_l^*, h_r^*, z_m)-p_l(z_l, z_r, z_m)$ (given by expressions (18) and (19)) produces:

$$\frac{\psi\phi(2z_m-z_l-z_r)(3z_l-3z_r+\phi)}{6(z_l-z_r)}. \tag{32}$$

(i) This difference (32) is zero when $z_m = (z_l + z_r)/2$. The derivative of (32) with respect to z_m is $\psi\phi(3z_l - 3z_r + \phi)/(3z_l - 6z_r)$, which is clearly positive.

(ii) The derivative of (32) with respect to ϕ is $\psi(2z_m - z_l - z_r)(3z_l - 3z_r + 2\phi)/(6z_l - 6z_r)$. This is positive (negative) when $z_m > (<)\ (z_l + z_r)/2$.

(iii) The expression for $p_l(h_l^*, h_r^*, z_m)$ is given by expression (31). By definition, $z_r + z_l$ is constant under increasing polarization, and thus it is clear that $p_l(h_l^*, h_r^*, z_m)$ is increasing (resp., decreasing) in polarization if $z_m < (>)\ (z_r + z_l)/2$. □

8. References

Binder, Sarah A. 1997. *Minority Rights, Majority Rule: Partisanship and the Development of Congress.* Cambridge: Cambridge University Press.

Bisin, Alberto, and Thierry Verdier. 2017. "On the Joint Evolution of Culture and Institutions." Unpublished manuscript, New York University.

Bowen, T. Renee, and Zaki Zahran. 2012. "On Dynamic Compromise." *Games and Economic Behavior* 76(2): 391-419.

Chafetz, Josh, and David E. Pozen. 2018. "How Constitutional Norms Break Down." *UCLA Law Review* 65: 1430-1459.

Dixit, Avinash, Gene M. Grossman, and Faruk Gul. 2000. "The Dynamics of Political Compromise." *Journal of Political Economy* 108(3): 531-568.

Fishkin, Joseph, and David E. Pozen. 2018. "Asymmetric Constitutional Hardball." *Columbia Law Review* 118(3): 915-982.

Fox, Justin. 2006. "Legislative Cooperation Among Impatient Legislators." *Journal of Theoretical Politics* 18(1): 68-97.

Gorodnichenko, Yuriy, and Gerard Roland. 2017. "Culture, Institutions, and the Wealth of Nations." *Review of Economics and Statistics* 99(3): 402-416.

Graham, Matthew H., and Milan W. Svolik. 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.

Grillo, Edoardo, and Carlo Prato. 2019. "Opportunistic Authoritarians, Reference-Dependent Preferences, and Democratic Backsliding." Unpublished manuscript, Columbia University.

Guiso, Luigi, Helios Herrera, and Massimo Morelli. 2016. "Cultural Differences and Institutional Integration." *Journal of International Economics* 99: 97-113.

Helmke, Gretchen, Mary Kroeger, and Jack Paine. 2019. "Exploiting Asymmetries: A Theory of Democratic Constitutional Hardball." Unpublished manuscript, University of Rochester.

Howell, William G., Kenneth Shepsle, and Stephane Wolton. 2019. "Executive Absolutism: A Model." Unpublished manuscript, University of Chicago.

Krehbiel, Keith. 1998. *Pivotal Politics: A Theory of US Lawmaking.* Chicago: University of Chicago Press.

Kreps, David. 1990. "Corporate Culture and Economic Theory." In J. Alt and K. A. Shepsle (eds.) *Perspectives on Positive Political Economy.* Cambridge: Cambridge University Press, pp. 90-143.

Levitsky, Steven, and Daniel Ziblatt. 2018. *How Democracies Die.* New York: Broadway Books.

Luo, Zhaotian, and Adam Przeworski. 2019. "Subversion by Stealth: Dynamics of Democratic Backsliding." Unpublished manuscript, New York University.

Maeda, Ko. 2010. "Two modes of Democratic Breakdown: A Competing Risks Analysis of Democratic Durability." *Journal of Politics* 72(4): 1129-1143.

Mann, Thomas E., and Norman J. Ornstein. 2016. *It's Even Worse Than It Looks: How the American Constitutional System Collided with the New Politics of Extremism.* New York: Basic Books.

McCarty, Nolan, Keith T. Poole, and Howard Rosenthal. 2016. *Polarized America: The Dance of Ideology and Unequal Riches.* Boston: MIT Press.

Nalepa, Monika, Georg Vanberg, and Caterina Chiopris. 2018. "Authoritarian Backsliding." Unpublished manuscript, University of Chicago.

North, Douglass C. 1991. "Institutions." *Journal of Economic Perspectives* 5(1): 97-112.

Przeworski, Adam. 2019. *Crises of Democracy.* Cambridge: Cambridge University Press.

Siegel, Neil S. 2018. "Political Norms, Constitutional Conventions, and President Donald Trump." *Indiana Law Journal* 93(1): 177-205.

Stepan, Alfred, and Juan J. Linz. 2011. "Comparative Perspectives on Inequality and the Quality of Democracy in the United States." *Perspectives on Politics* 9(4): 841-856.

Tabellini, Guido. 2008a. "Institutions and Culture." *Journal of the European Economic Association* 6(2-3): 255-294.

Tabellini, Guido. 2008b. "The Scope of Cooperation: Values and Incentives." *Quarterly Journal of Economics* 123(3): 905-950.

Tsebelis, George. 1995. "Decision Making in Political Systems: Veto Players in Presidentialism, Parliamentarism, Multicameralism and Multipartyism." *British Journal of Political Science* 25(3): 289-325.

Tushnet, Mark. 2004. "Constitutional Hardball." *John Marshall Law Review* 37: 523-553.

Waldner, David, and Ellen Lust. 2018. "Unwelcome Change: Coming to Terms with Democratic Backsliding." *Annual Review of Political Science* 21: 93-113.

Wawro, Gregory J., and Eric Schickler. 2004. "Where's the Pivot? Obstruction and Lawmaking in the Pre-Cloture Senate." *American Journal of Political Science* 48(4): 758-774.

Whittington, Keith E. 2013. "The Status of Unwritten Constitutional Conventions in the United States." *University of Illinois Law Review* 2013(5): 1847-1870.