Charismatic Leaders and Democratic Backsliding*

James R. Hollyer[1], Marko Klašnja[2], and Rocío Titiunik[3]

[1]University of Minnesota
[2]Georgetown University
[3]Princeton University

February 2021

Preliminary

Abstract

In recent years, democratic nations have frequently turned toward charismatic elected leaders. Such leaders foster a direct connection with (portions of) the electorate, developing an electoral following distinct from that of their party. Political parties benefit from such politicians, as the chances of prevailing in any given election rise in their leaders' charisma. But, they also pay a cost – reliance on a given politician's charismatic authority for electoral success lessens the ability of the party to control the candidate when, and if, she is elected. In this paper, we model the interaction between a chief executive and her party. When parties become highly reliant on a leader's charisma, they grow less able to sanction the behavior of these leaders in office. This is particularly true in settings characterized by high levels of ideological polarization. For a variety of parameter values, this inversion of the power dynamic between parties and politicians increases the likelihood that demagogic politicians are able to enact anti-democratic policies, and everywhere it increases the likelihood that politicians who enact such policies get away with it. We seek to test the proposition that the rise of charismatic leaders is associated autocratic reversions and a decline in the quality of democracy.

*We would like to thank Stephanie Rickard, Alexandra Cirone, Arturas Rozenas, Raúl Aldaz and participants in the Global Research in Political Economy seminar, the Virtual Formal Theory Workshop, the 2019 and 2020 European Political Science Association annual conferences and the 2019 American Political Science Association conference for comments and suggestions previous versions of this paper.

Political parties structure democratic competition. Parties bundle policy choices such that citizens face coherent electoral choices, assign labels to candidates in a manner that both ensures programmatic campaign promises are credible and ensures citizens know what policies their votes are supporting, and coordinate politician behavior within and between branches in a manner that allows policies to be enacted (Aldrich, 2011; Cox and McCubbins, 2007; Keefer, 2007). Parties also play an important role in selecting politicians, and may act as gatekeepers ensuring that those nominated to run for office are committed to democratic principles and willing to exercise power responsibly (Levitsky and Ziblatt, 2018). Once in office, parties may use the threat of curtailed access to resources, to withhold renomination, or even of impeachment or confidence votes to ensure politicians act to support party preferences and uphold the party brand. To play these roles, parties must be both willing and able to exercise power over their constituent politicians.

What happens when this power dynamic is inverted? Can political parties grow sufficiently dependent on specific politicians that their ability to act as a screening or disciplining device fails? In this paper, we argue that such situations may arise, and specifically occur when parties rely on the charismatic appeal of senior politicians, rather than the draw of the party platform, to mobilize support. When politicians assume chief executive roles under such circumstances, they are able to enact policies with little check from the party apparatus – including policies that may be inimical to continued democratic rule. The rise of such leaders may be associated with democratic backsliding and, in extreme cases, reversion to autocracy.

A party is least likely to exercise control over a charismatic leader when that party is ideologically distant from its competitors – i.e., when the party is ideologically extreme. This is because the party's dependence on a charismatic leader is electoral – moving to sanction the leader, failing to re-nominate her for the next contest, or removing her from power is likely to cost the party at the ballot box. This cost is rising in the charismatic appeal of the politician. The party's electoral loss is it opponents' gain. When these opponents hold ideological views sharply divergent from that of the incumbent party, the prospect of giving ground to these rivals is anathema – and so the party will be less likely to impose any kind of sanction on recalcitrant leaders.

This may imply that ideological polarization is doubly-damaging in terms of concentrating power in the hands of charismatic leaders, who have the power to exercise anti-democratic tendencies. Hollyer, Klašnja and Titiunik (forthcoming) find that parties that generate intense in-group loyalties – such as parties that are ideologically extreme or that exist in settings of high affective polarization – are particularly likely to advance politicians based on charismatic appeal rather than prolonged and loyal service to the party. Here, we find that such parties are unlikely to act as a constraint on charismatic leaders once they take power. Parties that are ideologically distant from their competitors are both less likely to act as political gatekeepers, and less able to constrain demagogic nominees should they prove electorally successful.

Our findings also emphasize the role of personalism in democratic politics. Studies of autocratic politics have long emphasized this feature – politics in autocracies is largely characterized by dictators jockeying for power with members of the political elite (Svolik, 2012). Autocratic institutions are often classified based

on the extent to which power is concentrated in the hands of the leader (Geddes, Wright and Frantz, 2014; Gandhi and Sumner, forthcoming; Meng, 2017), and these measures are highly predictive of a variety of aspects of regime behavior. Here, we argue that a similar jockeying for power takes place within political parties in democracies, between the leader and other party elites. Democratic personalism poses a threat to the stability of continued democratic rule.

In what follows, we first elaborate on our argument informally. We then relate this argument to the broader literature on democratic consolidation and survival. When then formalize our contentions in a game theoretic model. We intend to use this theoretical model as the basis for empirical testing, making use of a novel measure of leader charisma and examining its association with both autocratic reversions and diminutions in the quality of liberal democracy.

Argument

One critical role for political parties is their ability to control, or at least to influence, the choice of nominees who compete for high political office. Parties can help to secure the stability of the political system by deliberately screening out demagogues who lack commitment to democratic ideals. Levitsky and Ziblatt (2018) emphasize the importance of this gatekeeping role, noting that, in the US, many demagogic figures – who attracted large popular followings and had eyes for high office – found their aspirations derailed by party elites who denied them major party nominations.

The decision to eschew demagogic nominees highlighted by Levitsky and Ziblatt (2018) is a strategic one. Parties that successful act as gatekeepers opt to forgo any immediate electoral gains nominating a charismatic demagogue would bring, as this candidate draws her own loyal supporters to the ballot box.[1] Rather, these parties prioritize long-term reputational concerns, maintaining the party brand and helping to ensure the democratic game continues.

This gatekeeping function highlights a more general problem faced by parties: whether to prioritize the nomination of charismatic ('electable') candidates, or rather emphasize party discipline and the development of a party brand. A party's influence over nominations and career advancement is its single most powerful tool in ensuring politicians remain on brand. Disciplined parties that prioritize the creation and maintenance of a clear reputation will tend to use nomination policies to reward politicians and would-be politicians for advancing the party's message, usually through a system of gradual promotion through the ranks based on past service. Other parties, however, may be unable or unwilling to commit to such a nomination strategy. These parties may systematically seek out the most charismatic possible nominees, particularly those that have built-in name recognition or bases of support. Such a nomination strategy does not predicate advancement on prior service to the party, indeed there is a strong incentive to nominate the most charismatic candidates for the most prominent offices possible. Political outsiders may vault

[1] Presumably, a demagogue who lacked charisma and a personal following would be an unappealing candidate both on electoral and policy lines.

the ranks of such a party. The emphasis on nominee characteristics (charisma) rather than past actions tends to undermine party discipline, since charismatic politicians and would-be politicians know they can advance regardless of their behavior and uncharismatic ones know they have little chance of advancement regardless of behavior. Hollyer, Klašnja and Titiunik (forthcoming) examine parties' strategic decision over which type of nomination scheme to advance. Parties are likely to prioritize charisma when electoral outcomes are volatile and highly unpredictable and when said parties are characterized by intense in-group loyalties (or out-group hostility).

While we see no reason to assume an association between individuals' charismatic appeal and tendencies toward demagoguery or preferences for anti-democratic policies, party strategies may induce a correlation between charisma and *caudillismo* among those who are actually nominated for high office. When parties prioritize their long-term reputation, and choose to advance nominees based on past service advancing the party-line, they are able to commit to act as gatekeepers and are likely to screen out would-be demagogues. When parties prioritize charisma and short-run electoral appeal, they are unable to so-commit, forgoing their gatekeeping function. As a result, charismatic nominees are subject to systematically less screening than less-charismatic nominees, and so are more likely to prove to be demagogues who take anti-democratic actions if or when elected to office. Less charismatic candidates advanced gradually through the ranks are subject to greater screening, and so are less likely to deviate from democratic principles if or when they reach high office.

This differential screening may be sufficient to produce an association between the election of charismatic leaders to high executive office and democratic backsliding. But, we expect that an additional mechanism may magnify an such relationship. As parties become more reliant on the charisma of individual candidates for electoral success, and less focused on the party's brand as a means of attracting voters, the relative bargaining power of the leader and party is altered in favor of the former.

Consider the relationship between a chief executive and her party. The executive may have a number of personal policy priorities that are not shared by her party, and *vice versa*. The party may attempt to induce the executive to pursue its preferences through the threat of a variety of possible sanctions. The party may either remove the executive (either through a confidence vote in a parliamentary system or through impeachment in a presidential), fail to renominate her in subsequent elections, sideline supporters of the executive in upcoming electoral contests, etc.

However, openly moving against the executive carries a cost for the party. Insofar as voters are attracted to the *candidate* rather than the *party*, the party risks alienating its electoral base. Voters particularly loyal to the executive may choose to sit out subsequent elections should the party turn on its candidate, or may even choose to support the opposition. In some circumstances, splits between the party and executive have even led that executive to defect and campaign as an independent or start a new party, siphoning votes from the party to which they previously belonged. Acting against the leader, then, is likely to jeopardize a party's position of power.[2] Since voters are most likely to show this loyalty when the ex-

[2]The argument that attempts to hold a leader to account jeopardize the leader's coalition allies is common in theories of autocratic politics. See, *inter alia*, Bueno de Mesquita et al. (2003); Besley and Kudamatsu (2007); Padró i Miquel (2007). Our

ecutive wields a strong charismatic appeal, the party's costs from sanctioning an executive are rising the latter's charisma.

An emphasis on candidate charisma in nomination decisions is thus doubly damaging for party control. It directly undermines the party's ability to use career concerns to influence candidate behavior in elections. This leads parties to loose their gatekeeping role. It further undermines the bargaining position of the party vis-à-vis elected politicians once they assume office. This may weaken a check on executive power, potentially undermining one of the guardrails of democracy.

Merely because charismatic political leaders exercise bargaining power vis-à-vis their party, there is no reason to suspect that they will systematically seek to use such power to undermine democratic protections. Rather, this bargaining power implies that when the preferences of the executive and her party diverge, the executive will be able to have her way with high frequency. Charismatic leaders might use this leverage to enact conventional policies that depart from those most preferred by the bulk of her co-partisans. Or, it may be the case that the preferences of the executive and her party are more or less identical, in which case this leverage is never used. But, charismatic leaders who do seek to subvert democratic institutions – by, for instance, concentrating power in the executive, diminishing civil liberties, exercising greater control over the press, etc. – will find less resistance to seeing these preferences realized than would less charismatic leaders. Thus, one should an anticipate a correlation between charismatic leaders taking office and democratic backsliding.

These effects should be largest when the executive's party is ideologically distant from its opponents. Recall that the cost to the party of moving against the leader is electoral. Sanctioning a charismatic leader costs the party votes, increasing the likelihood that an opposing party assumes power. If this opposition is ideologically distant, the loss of power may lead to the enactment of policies that are anathema to the party. Increasing the likelihood of this outcome is therefore costly indeed – the costs of moving against the leader rise in ideological distance.

Ideological polarization among parties, therefore, is likely to interact in a complex manner with candidate charisma. Recall, Hollyer, Klašnja and Titiunik (forthcoming) argue that parties that are ideologically distant from their competitors are particularly likely to emphasize charisma in their nomination decisions. Extreme parties are particularly likely to advance charismatic candidates. Here, we argue that parties that are distant from their competitors are unlikely to be able to control such candidates should they assume office. Polarization may thus give rise to charismatic candidates, even as it renders any such candidates who secure election less constrained. Polarization among parties may thus play a role in democratic backsliding that parallels existing findings on the relationship between polarization *among voters* and backsliding (Svolik, 2017).

argument is that a similar logic applies within democratic political parties.

Existing Literature

This paper builds on a literature on democratic consolidation and survival far too vast to adequately survey here. Much of this literature focuses on structural features of the polity which tend to stabilize democratic rule. Most importantly, the probability of autocratic reversion is monotonically falling in *per capita* income, such that no high-income democracy has ever experienced such a revision (Przeworski and Limongi, 1997; Przeworski et al., 2000). Income – and past instances of instability – are the most robust predictors of autocratic reversion across extensive specification searches (Gassebner, Lamla and Vreeland, 2013). We omit structural considerations from our theoretical account, not because we think they are unimportant, but because they are not directly implicated in the mechanisms we seek to document. In light of existing findings, however, all our empirical results will control for these structural features.

Work that focuses on the strategic interactions that sustain or undermine continued democratic rule can be divided into two veins: one that emphasizes the role of the populace as the force that sustains or undermines the democratic equilibrium; and another that emphasizes the importance of institutional actors, particularly political parties.

The former branch of the literature stems from Weingast (1997), who argues that the threat of coordinated popular insurrection ensures that it is a best response for politicians to refrain from infringing on democratic freedoms. Recent work has examined how coordination problems among citizens in maintaining this threat may be resolved through the electoral process (Fearon, 2011; Little, LaGatta and Tucker, 2015; Tucker, 2007), and how low expectations of politicians can lead to a self-reinforcing cycle in which the citizenry will no longer be willing to play this role (Svolik, 2013). Others emphasize features that ensure citizens do not use the threat of mobilization to undermine politicians who are complying with the democratic rules of the game. Wright (2008) notes that high levels of initial political competition helps to ensure portions of the electorate don't feel excluded from the democratic order, such that they mount subsequent challenges to democratic rule. Hollyer, Rosendorff and Vreeland (2018, Part II) argue that greater levels of public information promotes both the functioning of elections and perceptions of their legitimacy – as it becomes more difficult to castigate other voters as ill-informed – promoting democratic survival.

In our theoretical account, we largely elide the strategic considerations of the electorate, focusing instead on interactions at the level of political parties. However, our theoretical account does share important similarities with one piece that falls in this branch of the literature, namely Svolik (2017). Svolik (2017) argues that increased ideological polarization within the electorate can drive even voters who hold pro-democracy attitudes to support autocratic candidates. As polarization increases, partisan voters may be willing to support demagogic co-partisans – knowing that these candidates are demagogues – because the policy costs of supporting candidates who share their commitment to democracy but not their partisan loyalties are simply too high. In our theory, partisan commitments play a similar role at the level of the party politics – partisan politicians may be unwilling to sanction a co-partisan executive as ideological polarization among parties rises. Critically, polarization in our account occurs at the level of the party; whereas, in Svolik's (2017) it occurs at the level of the citizen.

A second strand in the literature on democratic consolidation and survival, to which this paper belongs, more prominently emphasizes the role of political parties. Typically, this literature focuses on the willingness of opposition parties to accept defeat at the polls, rather than turning to extra-constitutional methods to ensure their preferred policies prevail. The decision to comply with the democratic order is thus one of accepting defeat today, and the chance to compete again tomorrow; versus the chance to seek the overthrow of democracy and the imposition of one's preferred policies by force. This trade off will be dictated by the odds of success in democratic elections relative to those of extra-constitutional attempts to seize power, and the distance between the policies desired by the parties or the groups they represent (often proxied by economic inequality).[3]

Our paper departs from this literature in its focus on interactions within a given party, between that party's wider political membership and a co-partisan chief executive, rather than between two or more parties. In this, our paper shares features of Levitsky and Ziblatt (2018), who note that the willingness of any given party to act as a gatekeeper screening out demagogic politicians is critical to democratic survival. We borrow this insight, but further discuss the ability of a party to constrain its leader in the event that it has failed in its gatekeeping role.

In focusing on relations within a given party, wherein actions by party members to sanction a chief executive may jeopardize those co-partisan's grip on power, our paper adapts many key assumptions of the literature on autocratic politics. Models of autocratic accountability often share these central assumptions.[4] The dependence of the party on its leader for its continued survival in power is often referred to in this literature as the regime's degree of personalism, where leaders attempt to amass more personalist authority over their time in office (Svolik, 2012). Contrastingly, regimes that are less dependent on their leader are institutionalized (Meng, 2017). An empirical literature attempts to classify autocratic regimes by their level of personalism (Gandhi and Sumner, forthcoming; Geddes, Wright and Frantz, 2014), which, in turn, predicts regime stability and aspects of regime behavior.

Our argument is that personalism is a feature of democratic, as well as autocratic, politics. Specifically, democratic parties that rely heavily on the charisma of their leading politicians – most notably their nominees for chief executive office – are highly personalized. As in autocracies, personalism implies that the party will be unable to act as a check on its leader, should she come to office.

Model

Primitives

Consider an interaction between two actors, an incumbent political party I and a leader drawn from that party L. There also exists an opposition party O, but, in this model, O is non-strategic. Our focus is

[3] See, *inter alia*, Acemoglu and Robinson (2006); Boix (2003); Boix and Stokes (2003); Przeworski (1991, 2005); Przeworski, Rivero and Xi (2013); Scartascini and Tommasi (2012); Wantchekon (2004).

[4] Bueno de Mesquita et al. (2003); Besley and Kudamatsu (2007); Padró i Miquel (2007) are prominent examples.

on the ability of the incumbent party to discipline an incumbent leader, hence we focus on the strategic considerations involving these two actors. We index actors $j \in \{L, I, O\}$.

The interaction will take place over two periods $t \in \{1, 2\}$. In each period, L will make policy choices along two dimensions. The first dimension is a standard unidimensional policy space, for instance the left-right ideological space that typifies many developed democracies. Denote L's choice of policy in this dimension as $x_t \in \mathbb{R}$.

L will make an additional policy choice in each period of play along an orthogonal policy dimension which we use to capture the degree of 'authoritarianism.' To capture this notion, let the policy be indexed over the non-negative real line $a_t \in \mathbb{R}_+$, where higher values of a_t correspond to more authoritarian policies. High values of a_t may thus indicate the concentration of power in the hands of the executive, the placing of limits on freedom of the press or assembly, or – at an extreme – staging an *autogolpe*.

After L makes her choice of policy in the first period of play, I makes a decision to retain or remove the leader $r \in \{0, 1\}$, where $r = 1$ denotes a decision to retain. This might represent a decision to call a vote of confidence, stage an impeachment, or simply to fail to renominate the incumbent leader in subsequent electoral contests. For analytic simplicity, we restrict our attention to this extreme form of sanctioning. However, the insights presented here should extend to other forms of party-sanctioning, so long as these are costly both to the leader and to the future political prospects of the incumbent party. If $r = 0$, the incumbent party selects a new candidate for the leadership position.

A contest for political power then takes place between the incumbent party I and the opposition O, where I prevails with probability $\rho(r, \nu) \in [0, 1]$. This contest for power may most readily be thought of as an election, but may extend to an extra-constitutional contest for power if one considers high values of a_t as equivalent to an *autogolpe*. Critically, $\rho(r, \nu)$ is a function both of I's prior decision regarding the fate of the incumbent leader ($r \in \{0, 1\}$) and of $\nu \in \mathbb{R}$, which represents the charismatic appeal of that leader. Higher values of ν are equivalent to higher levels of charisma. We assume that: (1) $\rho(1, \nu) > \rho(0, \nu), \forall\, \nu$, incumbent-party infighting harms the party's political future; (2) $\rho(\cdot, \nu)$ is increasing in ν, the more charismatic its leader the better I's political chances; and (3) $\rho(r, \nu)$ is subject to increasing differences in ν – i.e., the value $\rho(1, \nu) - \rho(0, \nu)$ is increasing in ν. This final assumption posits that, in moving against a charismatic leader, a party jeopardizes its relationship with it core supporters, harming its future political chances.

If O prevails in this contest for power, it chooses a new leader from among its ranks. If I prevails, its chosen leader remains in power. The sitting L then chooses x_2, a_2.

We assume that all actors have additively separable preferences over each policy dimension. Let each actor have an ideal point over x defined as $\hat{x}_J$ and an ideal point over a defined as $\hat{a}_j$. Utilities over the policy dimension are given by the strictly quasi-concave function $g(x; \hat{x}_j)$, and those over the authoritarianism dimension are given by a similarly strictly quasi-concave function $h(a; \hat{a}_j)$. Without loss of generality, we assume $\hat{x}_I > \hat{x}_O$ – and it will be useful to define $D \equiv \hat{x}_I - \hat{x}_O$. D will be our measure of ideological polarization.

Since we are interested in the ability of political parties to constrain the authoritarian impulses of political leaders, we assume that $\hat{a}_I = \hat{a}_O = 0$, i.e., that both political parties are committed to democratic norms. Of course, were this not to hold, the leader would have greater leeway to enact authoritarian policies. However, the conclusion that the election of anti-democratic parties can lead to anti-democratic outcomes is not particularly strategically rich. Our question of interest is whether, and when, authoritarian leaders can defy co-partisans who are largely committed to democracy.

The utility of each party $j \in \{I, O\}$ in a given period $t \in \{1, 2\}$ is thus given by:

$$u_{j,t} = g(x_t; \hat{x}_j) + h(a_t; 0).$$

All actors share a common discount factor $\delta \in (0, 1)$.

Leader utilities are similar to those of the parties. However, we assume that leaders may vary in their authoritarian tendencies. To capture this, let $\hat{a}_L$ be a random variable that assumes values $\{\underline{\alpha}, \bar{\alpha}\}$ where $\bar{\alpha} > \underline{\alpha} \geq 0$. Let the probability that L has a taste for authoritarian policies be defined as $Pr(\hat{a}_L = \bar{\alpha}) \equiv q$. The realization of $\hat{a}_L$ is known to L, but unobserved by any other actor.

For simplicity, we assume that leader ideal points over the policy space x are identical to those of their nominating party.[5]

The leader derives utility from her policies in both x- and a-space whenever she is in power. We assume she derives a utility normalized to zero if removed from office. Hence, L's utility is given by:

$$u_{L,t} = \begin{cases} g(x_t; \hat{x}_L) + h(a_t; \hat{a}_L) \text{ if in power} \\ 0 \text{ otherwise.} \end{cases}$$

The order of play is as follows:

1. The type of the sitting leader is drawn from the Bernoulli distribution $Pr(\hat{a}_L = \bar{\alpha}) = q, Pr(\hat{a}_L = \underline{\alpha}) = 1 - q$. L sets the policy $\{x_1, a_1\}$.

2. The incumbent party I observes L's choice $\{x_1, a_1\}$ and determines whether or not to retain the leader $r \in \{0, 1\}$. If $r = 0$, L is replaced by a new leader with $\hat{x}_L = \hat{x}_I$, who's type is drawn from an identical Bernoulli distribution to that above.

3. An election (or other struggle for power) takes place. I prevails with probability $\rho(r, \nu)$. If I looses (with probability $1 - \rho(r, \nu)$), a new leader is selected with $\hat{x}_L = \hat{x}_O$, who's type is drawn from an identical Bernoulli distribution to that above.

4. The sitting L sets polity $\{x_2, a_2\}$.

5. All payoffs are realized and the game ends.

[5] Our results would be substantively unchanged if this assumption were weakened, so long as leaders are closer to their nominating party than to their opposition.

Equilibrium

We consider perfect Bayesian equilibria (PBE) to this game. Such an equilibrium will consist of a strategy profile, and a set of posterior beliefs for I, such that each player is adopting a best response, given her beliefs; and beliefs are weakly consistent with player strategies (updated according to Bayes' rule, wherever possible). A strategy for L will consist of (1) in the second period of play, a mapping from ideal points into policies, $\{a_2, x_2\} : \{\underline{\alpha}, \bar{\alpha}\} \times \{\hat{x}_I, \hat{x}_O\} \rightarrow \mathbb{R}_+ \times \mathbb{R}$; and (2) in the first period of play, a mapping from the typespace, and the value of the parameter ν, into policy $\{a_1, x_1\} : \{\underline{\alpha}, \bar{\alpha}\} \times \mathbb{R} \rightarrow \mathbb{R}_+ \times \mathbb{R}$.[6] A strategy for I is a choice of $r \in \{0, 1\}$, wihcih is a mapping from the first period levels of authoritarianism and the parameter ν, $r : \mathbb{R}_+ \times \mathbb{R} \rightarrow \{0, 1\}$. Incumbent party beliefs, $Pr(\hat{a}_L = \bar{\alpha}|a_1)$ are updated according to Bayes' rule, wherever possible. We additional restrict our attention to PBE that satisfy the intuitive criterion refinement of Cho and Kreps (1987), which rules out implausible pooling equilibria to this game.

Before we can begin characterizing equilibria that satisfy the criteria described above, a preliminary definition and lemma are required. All proofs are presented in the appendix.

Lemma 1. *Consider the following inequality:* $(1-q)[h(\bar{\alpha};0)-h(\underline{\alpha};0)] \geq [1-\frac{\rho(0,\nu)}{\rho(1,\nu)}][g(\hat{x}_O;\hat{x}_I)-g(\hat{x}_I;\hat{x}_I)]$. *If this inequality holds for the parameter pair* $\{\nu', D'\}$*, then it will also hold for any* $\{\nu', D''\}$ *where* $D'' > D'$*. If the inequality holds for* $\{\nu', D'\}$*, then it will also hold for any* $\{\nu'', D'\}$ *where* $\nu'' > \nu'$*.*

Definition 1. *For a given value of* D*, define the set* $V_{accounabiltiy}$ *as containing all values of* ν *such that for* $\{\nu \in V_{accountability}, D\}$ *the inequality* $(1-q)[h(\bar{\alpha};0) - h(\underline{\alpha};0)] \geq [1 - \frac{\rho(0,\nu)}{\rho(1,\nu)}][g(\hat{x}_O;\hat{x}_I) - g(\hat{x}_I;\hat{x}_I)]$ *is satisfied. If* $V_{accountability}$ *is non-empty, define* $\bar{\nu}$ *as the minimal element of this set. If* $V_{accountability}$ *is empty, define* $\bar{\nu}$ *as equal to* ∞*.*

The relevance of $\bar{\nu}$ is as follows: this is the maximal value of charisma for which the party can ever threaten to hold the leader accountable. For any $\nu \geq \bar{\nu}$, the political costs to the party of removing the leader are too high to ever be credibly threatened. The party stands to alienate a broad swath of its supporters from removing such a charismatic leader, implying a sharply reduced likelihood of success in future political contests with the opposition. Given this, I has a dominant strategy of retaining the incumbent leader, i.e., setting $r = 1$.

This now allows us to characterize one PBE to the model:

Proposition 1. *If* $\nu \geq \bar{\nu}$*, the following strategy profile and set of posterior beliefs constitutes a unique PBE:*

- *For* L*:* $x_t = \hat{x}_L$,, $a_t = \begin{cases} \bar{\alpha} \text{ if } \hat{a}_L = \bar{\alpha} \\ \underline{\alpha} \text{ if } \hat{a}_L = \underline{\alpha} \end{cases}$ *for* $t \in \{1, 2\}$.
- *For* I*:* $r = 1$.
- *Posterior beliefs are given by:* $Pr(\hat{a}_L = \bar{\alpha}|a_1) = \begin{cases} 1 \text{ if } a_1 > \underline{\alpha} \\ 0 \text{ otherwise.} \end{cases}$

[6] Note, that in the first round of play, $\hat{x}_L = \hat{x}_I$ by construction, so there is no variation in this term.

In this equilibrium, the party ceases to act as a guardrail of democracy. Given its reliance on the charisma of its leader, it will never sanction her regardless of any authoritarian turn in her policies. L recognizes this, and so sets policies in line with her preferences. If she is vested in democratic norms, policies will remain minimally authoritarian $a_t = \underline{\alpha}$. But, if she is an autocratic type, then she is free to pursue these preferences $a_t = \bar{\alpha}$.

We can further state that this equilibrium will prevail for a wider range of realization of ν as ideological polarization rises. Leader charisma plays a greater role in undermining the guardrails of democracy when polarization is high than when it is low. This is because the cost the incumbent suffers from removing a charismatic leader comes in the form of diminished future political success – and, thus, an increased likelihood of success for the opposition. If the opposition's preferred policies along the policy dimension x are very distant from those of the incumbent party – D is large – this cost becomes more burdensome.

Lemma 2. *$\bar{\nu}$ is falling in D.*

It remains for us to characterize the equilibrium when $\nu < \bar{\nu}$ – i.e., when the guardrails of democracy stand firm, and the incumbent party can credibly commit to remove authoritarian leaders. This does not guarantee that the party can induce the incumbent to adopt democratic policies in the first period of play, but it does ensure that any attempt by the leader to deviate from such policies will be punished. Whether this is threat of punishment is sufficient to induce compliance depends on whether an autocratic leader is better off from making concessions to the party in the first period, in the hopes of subsequent reelection, or instead prefers to act on her authoritarian preferences today in spite of party sanctioning.

Proposition 2. *If $\nu < \bar{\nu}$, the following strategy profile and posterior beliefs constitute a PBE that uniquely satisfies an intuitive criterion refinement:*

- *For L: $x_t = \hat{x}_L$ for $t \in \{1, 2\}$. $a_2 = \hat{a}_L$. $a_1 = \begin{cases} \bar{\alpha} \text{ if } \hat{a}_L = \bar{\alpha} \text{ and } h(\bar{\alpha}; \bar{\alpha}) - h(\underline{\alpha}; \bar{\alpha}) \geq \delta\rho(1, \nu)[g(\hat{x}_I; \hat{x}_I) + h(\bar{\alpha}; \bar{\alpha})] \\ \underline{\alpha} \text{ otherwise.} \end{cases}$*
- *For I: $r(a_1) = \begin{cases} 1 \text{ if } a_1 \leq \underline{\alpha} \\ 0 \text{ otherwise.} \end{cases}$*
- *Posterior beliefs are given by $Pr(\hat{a}_L = \bar{\alpha}|a_1) = \begin{cases} 1 \text{ if } a_1 > \underline{\alpha} \\ q \text{ if } a_1 = \underline{\alpha} \text{ and } h(\bar{\alpha}; \bar{\alpha}) - h(\underline{\alpha}; \bar{\alpha}) < \delta\rho(1, \nu)[g(\hat{x}_I; \hat{x}_I) + h(\bar{\alpha}; \bar{\alpha})] \\ 0 \text{ otherwise.} \end{cases}$*

Given that $\nu < \bar{\nu}$, the incumbent party is free to remove leaders they believe to be authoritarian and retain those they believe not to be. Given this, it is always a best response for democratic types of L to follow their innate policy preferences, setting $a_t = \underline{\alpha}$.[7] The question then is whether autocratic types wish to pool with democrats, by setting $a_1 = \underline{\alpha}$, ensuring that the party cannot discern type and so will never act to remove. Or if, instead, autocratic types of L wish to satisfy their primitive preference by setting $a_1 = \bar{\alpha}$,

[7] Formally speaking, one could construct alternative PBE in which democratic types are induced to pool with autocratic types by setting a_1 equal to some other value. However, any such pooling PBE is equilibrium dominated by that described in Proposition 2. Hence, this PBE uniquely satisfies the intuitive criterion refinement.

ensuring their certain removal by the party. The answer to this question is dictated by the expression $h(\bar{\alpha};\bar{\alpha}) - h(\underline{\alpha};\bar{\alpha}) \gtrless \delta\rho(1,\nu)[g(\hat{x}_I;\hat{x}_I) + h(\bar{\alpha};\bar{\alpha})]$. If the left-hand side of this expression is greater than or equal tothe right, autocratic types prefer to separate and set $a_1 = \bar{\alpha}$. If not, they prefer to pool and set $a_1 = \underline{\alpha}$, and survive in power for another day.

Comparative Statics

We can now turn to the comparative static predictions generated by our model, which will also form the basis of our empirical investigation.

Before we proceed, it will be useful to define another threshold in ν.

Lemma 3. *If the expression $h(\bar{\alpha};\bar{\alpha}) - h(\underline{\alpha};\bar{\alpha}) \geq \delta\rho(1,\nu)[g(\hat{x}_I;\hat{x}_I) + h(\bar{\alpha};\bar{\alpha})]$ holds for ν', then it also holds for any $\nu'' < \nu'$.*

Definition 2. *Define the set $V_{compatibility}$ as comprised of all values ν such that the expression $h(\bar{\alpha};\bar{\alpha}) - h(\underline{\alpha};\bar{\alpha}) \geq \delta\rho(1,\nu)[g(\hat{x}_I;\hat{x}_I) + h(\bar{\alpha};\bar{\alpha})]$ holds. If $V_{compatibility}$ is non-empty, let $\hat{\nu}$ be the maximal element in this set. If $V_{compatibility}$ is empty, let $\hat{\nu} = -\infty$.*

Thus, if $\nu < \hat{\nu},\ \bar{\nu}$, the separating equilibrium defined in Proposition 2 holds. Autocratic types satisfy their preference for accruing power in the first period of play, but face immediate sanctioning by their party. If, on the other hand, $\bar{\nu} > \nu > \hat{\nu}$, the pooling equilibrium defined by this proposition holds.

Proposition 3. *If $\hat{\nu} < \bar{\nu}$, the* ex ante *probability the incumbent adopts authoritarian policies ($Pr(a_1 = \bar{\alpha})$) is non-monotonic in ν. Specifically,* $Pr(a_1 = \bar{\alpha}) = \begin{cases} q \text{ if } \nu \leq \hat{\nu} \\ 0 \text{ if } \nu \in (\hat{\nu}, \bar{\nu}) \\ q \text{ if } \nu \geq \bar{\alpha} \end{cases}$

If, contrastingly, $\hat{\nu} \not< \bar{\nu}$, the ex ante *probability the incumbent adopts authoritarian policies ($Pr(a_1 = \alpha)$) is fixed and equal to q for all values of ν.*

The incumbent party faces little cost from removing highly uncharismatic leaders, implying that such leaders are subject to party sanctioning for violating democratic norms. But, such candidates also face relative poor prospects of reelection, even if they escape party sanctioning. Because of this, there is little reason for the leader to attempt to hide her true preferences – if an authoritarian type, she will adopt policies consistent with her type in the first period and accept the risk of party sanction. As she grows more charismatic, the leader's chances in subsequent elections improve. So too does the cost to the party from sanctioning the leader. If $\hat{\nu} < \bar{\nu}$, the first effect initially dominates the second. For $\nu \in (\hat{\nu}, \bar{\nu})$, party can commit to sanction the leader for pursuing authoritarian policies, and the leader values the chance to contest future elections sufficiently to comply with the party's demands. For $\nu > \bar{\nu}$, however, the latter effect dominates, and the party can no longer credibly sanction the leader. If $\hat{\nu} \not< \bar{\nu}$, then the space in which (1) the party can credibly threaten the leader, and (2) the leader is willing to comply with the party's demands is empty.

Remark 1. *In any instance in which $a_1 > \underline{\alpha}$ in equilibrium, the incumbent is sanctioned (removed) by the party if $\nu < \bar{\nu}$ and is unpunished otherwise.*

Remark 1 follows from the definitions of Propositions 2 and 1. The combination of this remark and Proposition 3 leads us to our empirical conclusion that charismatic leaders pose a greater threat of democratic backsliding than party insiders. Proposition 3 establishes that authoritarian policies are most likely to be pursued by leaders at the extreme ends of the charisma spectrum. Highly uncharismatic leaders may be prone to adopting authoritarian policies. But, Remark 1 clarifies that they will be removed from office for doing so, limiting the damage they can inflict. Whereas, those at the high end of the charisma spectrum are also likely to adopt authoritarian policies. Remark 1 clarifies that these charismatic leaders are unlikely to face punishment for their actions, making the threat of democratic backsliding more acute.

Proposition 4. *An increase in ideological polarization (D) weakly increases the* ex ante *probability the incumbent adopts authoritarian policies ($Pr(a_1 = \bar{\alpha})$).*

As noted above, increasing levels of ideological polarization paralyze the party. Any weakening of the future political prospects of the party are more costly as polarization rises, so attempts to sanction the leader are more risky. In terms of the model, $\bar{\nu}$ declines as D rises. Ideological polarization, however, doesn't affect the value of $\hat{\nu}$ – the decision of an autocratic leader to comply or defect from party demands are not impacted by this term. Hence, the space in which the party is able to induce compliance by the leader $\nu \in (\hat{\nu}, \bar{\nu})$ is shrinking in this term. Empirically, we anticipate that ruling parties that are ideologically distant from their competitors will be associated with democratic backsliding.

Empirics

In Progress

In what follows, we subject our prediction that charismatic chief executives are associated with democratic backsliding to empirical scrutiny. We will specifically examine two forms of backsliding: (1) autocratic reversions (change in regime-type), and (2) diminution in the quality of democracy, which will also capture declines in protections for civil liberties. We assess the association of this term with a bespoke measure of leader charisma. To construct this measure, we apply text analytic methods to a corpus of academic works referencing all democratic leaders holding high executive office between XXXX and XXXX. We elaborate on our definitions and measurement strategy below.

Data Description

Outcome Measures

Our empirical analyses make use of two alternative outcome variables, corresponding to differing definitions of democracy and democratic backsliding. The first of these is an indicator variable $democracy \in$

$\{0, 1\}$ as coded by Cheibub, Gandhi and Vreeland (2010), following the minimalist definition of democracy advanced by Alvarez et al. (1996). A country is coded as democratic if executive and legislative positions are filled via elections, fought between competing parties, where the outcome of these elections is *ex ante* uncertain.[8] Since our interest is in democratic backsliding, we estimate the probability of a transition from a one to a zero according to this measure. Our sample will consist of all democratic-spell years between 1950 and 2008, defined as one or more years of continuous democracy in a given country, and we estimate the hazard that the democracy-spell comes to an end. The outcome of interest is thus a recoded indicator which takes the value one if and only if the indicator $democracy$ changes from a one to a zero in a given country at a given time.

Transitions in political regime-type are extreme, and rare, political events. Autocratic reversions require the suspension of elections or of political parties. Our theory, however, allows for the possibility of less drastic 'anti-democratic' actions. Rather than fully suspending democratic institutions, charismatic leaders may instead seek to diminish the role of checks and balances, weaken civil liberties, or extend the control of the government over the press. Recent actions by governments in Hungary and Poland, which stop short of the suspension of democracy, have increased academic interest in such behaviors. We refer to such backsliding as a reduction in the quality of democracy.

To measure such actions, we turn to the Liberal Democracy Index drawn from the Varieties of Democracy (V-Dem) project, version 9 (Coppedge et al., 2019). The V-Dem indexes adopt a more comprehensive definition of democracy than do Cheibub, Gandhi and Vreeland (2010), incorporating civil liberties, the rule of law, electoral participation and the strength of civil society, etc. – building off of the theoretical framework of Dahl (1971). We particularly focus on the Liberal Democracy Index, which measures the extent of (1) protection of civil liberties, (2) the rule of law, and (3) checks and balances. The aim of the index is to construct a measure of constraints on the executive – which, in our theory, are precisely the types of constraints a charismatic leader might seek to obviate. This index is a continuous measure, scaled to fall between zero and one, with higher values denoting greater checks on the executive.

We are also considering using the executive constraints component of the Polity2 index (Marshall and Jaggers, N.d.) to capture this outcome of theoretical interest.

Explanatory Terms

[In Progress]

Results

[TBD]

[8] Formally, this last criterion is satisfied if there has been at least one change in the party in power under the existing constitutional order. Provided such a change in power, years under the same constitutional order are retroactively coded as democratic.

Conclusion [?]

Charismatic authority plays an important role in democratic politics. When political parties advance candidates who possess such authority, they may increase their chances for short-term electoral success – both directly and as a result of coattail effects. But, such short-term gains may come at a cost for the party. Prioritizing charisma may mean that parties become less focused on their gatekeeping role in nominating candidates for office (Levitsky and Ziblatt, 2018). Beyond this effect, should a charismatic candidate succeed in attaining office, her party will find itself in a weak position to constrain her behavior once in power.

In this paper, we argue that the accumulation of these two effects has important implications for democratic survival. Reliance on candidates' charismatic authority increases the risks of democratic backsliding. When parties cease to exercise a gatekeeping role, the odds that a politician with anti-democratic preferences takes executive power rises. When the executive's party finds itself in a weak bargaining position vis-à-vis its leader, one check on executive power is weakened.

This paper contributes to a recent resurgence in work on democratic backsliding and (conversely) democratic consolidation. Its argument that parties' strategies of nomination and promotion contribute to democratic (in)stability is, to our knowledge, novel. This also offers the possibility of a relatively objective means of characterizing and coding parties that may pose a threat to democratic rule.

We also highlight this paper's contribution in linking personalism to studies of democratic politics. Personalism – the dependence of a winning coalition on its leader for continued access to positions of privilege – has long been a feature of theoretical focus in autocratic politics. But, just as this feature shapes the bargaining between an autocrat and his winning coalition, it may shape the bargaining between a democratic chief executive and her party. In a democratic context, personalism constitutes a threat to the stability of the political order.

References

Acemoglu, Daron and James A. Robinson. 2006. *Economic Origins of Dictatorship and Democracy*. Cambridge University Press.

Aldrich, John H. 2011. *Why Parties?: A second look*. University of Chicago Press.

Alvarez, Mike, Jose Antonio Cheibub, Fernando Limongi and Adam Przeworski. 1996. "Classifying Political Regimes." *Studies in Comparative International Development* 31:3–36.

Besley, Timothy and Masayuki Kudamatsu. 2007. "Making Autocracy Work." Development Economics Discussion Paper Series 48, http://eprints.lse.ac.uk/3764/1/Making_Autocracy_Work.pdf.

Boix, Carles. 2003. *Democracy and Redistribution*. Cambridge University Press.

Boix, Carles and Susan C. Stokes. 2003. "Endogenous Democratization." *World Politics* 55(4):517–549.

Bueno de Mesquita, Bruce, Alastair Smith, Randolph M. Siverson and James D. Morrow. 2003. *The Logic of Political Survival*. Cambridge, MA: The MIT Press.

Cheibub, José Antonio, Jennifer Gandhi and James Raymond Vreeland. 2010. "Democracy and Dictatorship Revisited." *Public Choice* 143:67–101.

Cho, In-Koo and David M. Kreps. 1987. "Signaling Games and Stable Equilibria." *The Quarterly Journal of Economics* 102(2):179–222.

Coppedge, Michael, John Gerring, Carl Henrik Knutsen, Staffan I. Lindberg, Jan Teorell, David Altman, Michael Bernhard, M. Steven Fish, Adam Glynn, Allen Hicken, Anna Lührmann, Kyle L. Marquardt, Kelly McMann, Pamela Paxton, Daniel Pemstein, Brigitte Seim, Rachel Sigman, Svend-Erik Skaaning, Jeffrey Staton, Agnes Cornell, Lisa Gastaldi, Haakon Gjerløw, Valeriya Mechkova, Johannes von Römer, Aksel Sundtröm, Eitan Tzelgov, Luca Uberti, Yi ting Wang, Tore Wig and Daniel Ziblatt. 2019. "V-Dem Codebook v9.".

Cox, Gary W. and Matthew D. McCubbins. 2007. *Legislative Leviathan: Party Government in the House*. Cambridge University Press.

Dahl, Robert A. 1971. *Polyarchy: Participation and Opposition*. Yale University Press.

Fearon, James D. 2011. "Self-Enforcing Democracy." *Quaterly Journal of Economics* 126(4):1661–1708.

Gandhi, Jennifer and Jane Lawrence Sumner. forthcoming. "Measuring the Consolidation of Power in Non-Democracies." *Journal of Politics* .

Gassebner, Martin, Michael J. Lamla and James Raymond Vreeland. 2013. "Extreme Bounds of Democracy." *Journal of Conflict Resolution* 57(2):171–197.

Geddes, Barbara, Joseph Wright and Erica Frantz. 2014. "Autocratic Breakdown and Regime Transitions: A New Data Set." *Perspectives on Politics* 12(2):313–331.

Hollyer, James R., B. Peter Rosendorff and James Raymond Vreeland. 2018. *Information, Democracy, and Autocracy: Economic Transparency and Political (In)Stability*. Cambridge University Press.

Hollyer, James R., Marko Klašnja and Rocío Titiunik. forthcoming. "Parties as Disciplinarians: Charisma and Commitment Problems in Programmatic Campaigning." *American Journal of Political Science* .

Keefer, Philip. 2007. "Clientelism, Credibility, and the Policy Choices of Young Democracies." *American Journal of Political Science* 51(4):804–821.

Levitsky, Steven and Daniel Ziblatt. 2018. *How Democracies Die*. Crown.

Little, Andrew T., Tom LaGatta and Joshua A. Tucker. 2015. "Elections, Protest and Alternation of Power." *Journal of Politics* 77(4):1142–1156.

Marshall, Monty G. and Keith Jaggers. N.d. "Polity IV Project: Political Regime Characteristics and Transitions, 1800-2009.".

Meng, Anne. 2017. "When do Autocrats Share Power? A Theory of Party Institutionalization and Leader Strength.".

Padró i Miquel, Gerard. 2007. "The Control of Politicians in Divided Societies: The Politics of Fear." *Review of Economic Studies* 74:1259–1274.

Przeworski, Adam. 1991. *Democracy and the Market*. Cambridge University Press.

Przeworski, Adam. 2005. "Democracy as an Equilibrium." *Public Choice* 123:253–273.

Przeworski, Adam and Fernando Limongi. 1997. "Modernization: Theories and Facts." *World Politics* 49.2:155–183.

Przeworski, Adam, Gonzalo Rivero and Tianyang Xi. 2013. "Elections Under the Shadow of Force." Working Paper, Presented at the Workshop on Elections and Political Order, The Carter Center.

Przeworski, Adam, Michael E. Alvarez, José Antonio Cheibub and Fernando Limongi. 2000. *Democracy and Development: Political Institutions and Well-Being in the World, 1950-1990*. Cambridge University Press.

Scartascini, Carlos and Mariano Tommasi. 2012. "The Making of Policy: Institutionalized or Not?" *American Journal of Political Science* 56(4):787–801.

Svolik, Milan. 2017. "When Polarization Trumps Civic Virtue: Partisan Conflict and the Subversion of Democracy by Incumbents.".

Svolik, Milan W. 2012. *The Politics of Authoritarian Rule*. Cambridge University Press.

Svolik, Milan W. 2013. "Learning to Love Democracy: Electoral Accountability and the Success of Democracy." *American Journal of Political Science* 57(3):685–702.

Tucker, Joshua A. 2007. "Enough! Electoral Fraud, Collective Action Problems, and Post-Communist Colored Revolutions." *Perspectives on Politics* 5(3):535–551.

Wantchekon, Leonard. 2004. "The Paradox of "Warlord" Democracy: A Theoretical Investigation." *The American Political Science Review* 98.1:17–33.

Weingast, Barry R. 1997. "The Political Foundations of Democracy and the Rule of Law." *The American Political Science Review* 91(2):245–263.

Wright, Joseph. 2008. “Political Competition and Democratic Stability in New Democracies.” *British Journal of Political Science* 38(2):221–245.

A Proofs of Theoretical Propositions

Proof of Lemma 1

We are considering the expression

$$(1-q)[h(\bar{\alpha};0)-h(\underline{\alpha};0)] \geq [1-\frac{\rho(0,\nu)}{\rho(1,\nu)}][g(\hat{x}_O;\hat{x}_I)-g(\hat{x}_I;\hat{x}_I)]$$

Given that $h(\cdot,0)$ is a strictly quasi-concave function with a maximum at zero, and given that $\bar{\alpha} > \underline{\alpha} \geq 0$, the left-hand side of this expression is a strictly negative constant.

Similarly, given that $g(\cdot,\hat{x}_I)$ is a strictly quasi-concave function with a maximum at $\hat{x}_I$, and $\hat{x}_O < \hat{x}_I$, $g(\hat{x}_O;\hat{x}_I) - g(\hat{x}_I;\hat{x}_I) < 0$. Moreover, this expression is growing more negative in D, implying that the entire right-hand side of the initial inequality is falling in D. Hence, if the initial inequality holds for a given parameter pair $\{\nu', D'\}$, it must also hold for $\{\nu', D''\}$ where $D'' > D'$.

Finally, since $\rho(1,\nu) > \rho(0,\nu)\ \forall\ \nu$, the expression $1-\frac{\rho(0,\nu)}{\rho(1,\nu)} \in [0,1]$. Moreover, since $\rho(\cdot,\cdot)$ is subject to increasing differences in ν, this expression is rising in ν. Since this expression is rising in ν, and it multiplies $g(\hat{x}_O;\hat{x}_I) - g(\hat{x}_I;\hat{x}_I) < 0$ on the right-hand side of the inequality, the right hand side is falling in ν. Hence, if the initial expression holds for $\{\nu', D'\}$, it must also hold for $\{\nu'', D'\}$ where $\nu'' > \nu'$. □

Proof of Proposition 1

Proceed via backward induction. In the final period of play, the sitting L will set $x_2 = \hat{x}_L$ and $a_2 = \hat{a}_L$. This is a dominant strategy.

Given this, in the previous stage, I prefers to retain the current leader (set $r=1$) iff:

$$\rho(1,\nu)\{g(\hat{x}_I;\hat{x}_I) + Pr(\hat{a}_L=\bar{\alpha})h(\bar{\alpha};0) + [1-Pr(\hat{a}_L=\bar{\alpha})]h(\underline{\alpha};0)\} + [1-\rho(1,\nu)]\{g(\hat{x}_O;\hat{x}_I) + qh(\bar{\alpha};0) + (1-q)h(\underline{\alpha};0)\} \geq$$
$$\rho(0,\nu)g(\hat{x}_I;\hat{x}_I) + [1-\rho(0,\nu)]g(\hat{x}_O;\hat{x}_I) + qh(\bar{\alpha};0) + (1-q)h(\underline{\alpha};0)$$
$$\Rightarrow [Pr(\hat{a}_L=\bar{\alpha}) - q][h(\bar{\alpha};0) - h(\underline{\alpha};0)] \geq [1-\frac{\rho(0,\nu)}{\rho(1,\nu)}][g(\hat{x}_O;\hat{x}_I) - g(\hat{x}_I;\hat{x}_I)]$$

Since the proposition states $\nu > \bar{\nu}$, by definition we have $(1-q)[h(\bar{\alpha};0)-h(\underline{\alpha};0)] \geq [1-\frac{\rho(0,\nu)}{\rho(1,\nu)}][g(\hat{x}_O;\hat{x}_I)-g(\hat{x}_I;\hat{x}_I)]$. So, I has a dominant strategy of setting $r=1$.

Given this, the unique best response for L in the first period of play is to set $x_1 = \hat{x}_L$ and $a_1 = \hat{a}_L$. $\hat{a}_L \in \{\bar{\alpha}, \underline{\alpha}\}$, which gives us the equilibrium postulated. Since the best responses are unique everywhere, and since all information sets are hit in equilibrium, the PBE is unique. □

Proof of Lemma 2

This follows from the proof of Lemma 1. Since $g(\hat{x}_O;\hat{x}_I) - g(\hat{x}_I;\hat{x}_I)$ is growing more negative in D; the value of $1-\frac{\rho(0,\nu)}{\rho(1,\nu)}$ necessary to satisfy the inequality specified in that lemma is falling in D. Since $1-\frac{\rho(0,\nu)}{\rho(1,\nu)}$

is rising in ν, the requisite value of ν necessary to sustain the inequality must be falling in D. $\square$

Proof of Proposition 2

From the proof of Proposition 1, we saw that in any PBE, I has a best response of setting $r = 1$ iff:

$$[Pr(\hat{a}_L = \bar{\alpha}) - q][h(\bar{\alpha}; 0) - h(\underline{\alpha}; 0)] \geq [1 - \frac{\rho(0, \nu)}{\rho(1, \nu)}][g(\hat{x}_O; \hat{x}_I) - g(\hat{x}_I; \hat{x}_I)]$$

Given that $\nu < \bar{\nu}$ in the statement of the proposition, we know that this expression fails to hold when $Pr(\hat{a}_L = \bar{\alpha}) = 1$. So, in any separating equilibrium, wherein I can perfectly infer L's type, I's best response must be to remove autocratic types.

Contrastingly, since the right-hand side of this expression is strictly negative (as is $h(\bar{\alpha}; 0) - h(\underline{\alpha}; 0)$), in any pooling equilibrium where $Pr(\hat{a}_L = \bar{\alpha}|a_1) = q$ the expression must strictly hold. I's best response must be to retain any L who pools.

In principal, one could construct a pooling PBE in which leaders of all types pool on a given value of a_1 for any admissible value of this term. But, a pooling equilibrium in which all types of L pooling on setting $a_1 = \underline{\alpha}$ equilibrium dominates all alternatives.

Notice that $a_1 = \underline{\alpha}$ dominates all alternatives for types $\hat{a}_L = \underline{\alpha}$. If for some $a_1 = a'$, $h(a'; \bar{\alpha}) \leq h(\underline{\alpha}; \bar{\alpha})$, then $\underline{\alpha}$ dominates a for autocratic types as well. Hence, under the intuitive criterion, I must have a posterior belief that $Pr(\hat{a}_L = \bar{\alpha}|a_1 = \underline{\alpha}) = q$, which would cause this alternative pooling equilibrium to break down – the alternative is equilibrium dominated. Contrastingly, if $h(a'; \bar{\alpha}) > h(\underline{\alpha}; \bar{\alpha})$, $a_1 = \underline{\alpha}$ dominates only for democratic types. This implies I must attach posterior beliefs on seeing $Pr(\hat{a}_L = \bar{\alpha}|a_1 = \underline{\alpha}) = 0$, which also causes the equilibrium to break down. The alternative is again equilibrium dominated.

Given that the only form of pooling we need to consider involves pooling at the democratic type's ideal point, the only remaining question is whether authoritarian types wish to pool or separate. If they will separate, their utility maximizing option is naturally to set policy at their ideal point $a_1 = \bar{\alpha}$. This decision will then be dictated by the inequality:

$$h(\bar{\alpha}; \bar{\alpha}) - h(\underline{\alpha}; \bar{\alpha}) \geq \delta\rho(1, \nu)[g(\hat{x}_I; \hat{x}_I) + h(\bar{\alpha}; \bar{\alpha})].$$

If this inequality is satisfied, authoritarian types will prefer to separate. If not, they will wish to pool. Hence, we are left with a unique PBE in which:

- For L: $x_t = \hat{x}_L$ for $t \in \{1, 2\}$. $a_2 = \hat{a}_L$. $a_1 = \begin{cases} \bar{\alpha} \text{ if } \hat{a}_L = \bar{\alpha} \text{ and } h(\bar{\alpha}; \bar{\alpha}) - h(\underline{\alpha}; \bar{\alpha}) \geq \delta\rho(1, \nu)[g(\hat{x}_I; \hat{x}_I) + h(\bar{\alpha}; \bar{\alpha})] \\ \underline{\alpha} \text{ otherwise.} \end{cases}$
- For I: $r(a_1) = \begin{cases} 1 \text{ if } a_1 \leq \underline{\alpha} \\ 0 \text{ otherwise.} \end{cases}$
- Posterior beliefs are given by $Pr(\hat{a}_L = \bar{\alpha}|a_1) = \begin{cases} 1 \text{ if } a_1 > \underline{\alpha} \\ q \text{ if } a_1 = \underline{\alpha} \text{ and } h(\bar{\alpha}; \bar{\alpha}) - h(\underline{\alpha}; \bar{\alpha}) < \delta\rho(1, \nu)[g(\hat{x}_I; \hat{x}_I) + h(\bar{\alpha}; \bar{\alpha})] \\ 0 \text{ otherwise.} \end{cases}$

□

Proof of Lemma 3

We are concerned with whether the following expression holds:

$$h(\bar{\alpha};\bar{\alpha}) - h(\underline{\alpha};\bar{\alpha}) \geq \delta\rho(1,\nu)[g(\hat{x}_I;\hat{x}_I) + h(\bar{\alpha};\bar{\alpha})].$$

Given that the function $h(\cdot,\bar{\alpha})$ is strictly quasi-concave with a maximum at $\bar{\alpha}$, the left-hand side of this inequality is strictly positive.

On the right-hand side $\delta\rho(1,\nu) \in (0,1)$. Hence, if the expression holds for any value of $\nu = \nu'$, it must be the case that $g(\hat{x}_I;\hat{x}_I) + h(\bar{\alpha};\bar{\alpha}) > 0$. Moreover, by definition, $\rho(1,\nu)$ is increasing in ν. Hence, $\delta\rho(1,\nu')[g(\hat{x}_I;\hat{x}_I) + h(\bar{\alpha};\bar{\alpha})] > \delta\rho(1,\nu'')[g(\hat{x}_I;\hat{x}_I) + h(\bar{\alpha};\bar{\alpha})]$ iff $\nu'' < \nu'$. □

Proof of Proposition 3

From Proposition 1, if $\nu > \bar{\nu}$, L sets $a_1 = \bar{\alpha}$ iff $\hat{a_L} = \bar{\alpha}$, which occurs with probability q.

From Proposition 2, and from the definition of $\hat{\nu}$ if Definition 2, if $\nu < \hat{\nu}$ and $\nu < \bar{\nu}$, L sets $a_1 = \bar{\alpha}$ iff $\hat{a}_L = \bar{\alpha}$. If, contrastingly, $\nu < \bar{\nu}$ but $\nu \geq \hat{\nu}$, then all types of L pool on setting $a_1 = \underline{\alpha}$. □

Proof of Proposition 4

From Proposition 3, the determination of a_1 depends on whether $\nu \gtrless \{\hat{\nu},\bar{\nu}\}$. $\hat{\nu}$ is invariant in D. By contrast, from Lemma 2, $\bar{\nu}$ is falling in D. Notate this dependence as $\bar{\nu}(D)$.

Hence, for any two values of D, D', D'', $D'' > D'$, either $\nu \geq \bar{\nu}(D') \geq \bar{\nu}(D'')$, in which case $Pr(a_1 = \bar{\alpha})$ is constant across $D = D'$ and $D = D''$; $\bar{\nu}(D') \geq \bar{\nu}(D'') > \nu$, in which case $Pr(a_1 = \bar{\alpha})$ is constant across $D = D'$ and $D = D''$; or $\bar{\nu}(D') > \nu > \bar{\nu}(D'')$, in which case $Pr(a_1 = \bar{\alpha})$ increases from 0 to q whenever $\nu > \hat{\nu}$. □