Democracy and Its Vulnerabilities: 
Dynamics of Democratic Backsliding

Zhaotian Luo  
University of Chicago  
luozhaotian@uchicago.edu

Adam Przeworski  
New York University  
adam.przeworski@nyu.edu

December 8, 2019

Abstract

The dream of all politicians is to remain for ever in office. Most governments attempt to advance this goal by building popular support within the established institutional framework. Some, however, seek to protect their tenure in office by undermining institutions and disabling all opposition. The striking lesson of the successful cases of backsliding is that governments need not take unconstitutional or undemocratic steps to secure domination and yet the cumulative effect is that unless citizens react early they may lose the ability to remove the incumbent government by democratic means. We investigate what makes democracy vulnerable to such steps, specifically, whether a government would take anti-democratic steps, whether it can be stopped short of realization of complete domination, and whether it is likely to be removed at any stage of the process.

*For comments on earlier versions we thank John Ferejohn, Joanne Fox-Przeworski, Diego Gambetta, Roberto Gargarella, Robert Gulotty, Steven Holmes, Andrew Little, Beatriz Magaloni, Bernard Manin, Pasquale Pasquino, Arturas Rozenas, Milan Svolik, and Andrea Vindigni.
1. **Introduction**

Most, if not all, democracies in history were established as a reaction against “despotic,” “tyrannical,” or “autocratic” rule. Their institutional systems were designed to prevent incumbents from holding onto office independently of popular sentiments or from adopting measures that would curtail individual freedoms. The resulting institutions have varied but the goal everywhere was to design a system in which each part of the government would want to and have the means to prevent usurpation of power by any other part. The father of constitutionalism, Montesquieu (1995 [1748]: 326), insisted that “For the abuse of power to be impossible, it is necessary that by the disposition of things, the power stop the power.” Or, in an often cited Madison’s passage (*Federalist #51*), “the great security against a gradual concentration of several powers in the same department consists in giving to those who administer each department the necessary constitutional means and personal motives to resist encroachments of the others.... Ambition must be made to counteract ambition.” The effect of the separation of powers would be “limited” or “moderate” government.\(^1\)

Not everyone was confident that institutional checks would be sufficient to maintain the balance of powers.\(^2\) But if these internal controls were to fail, if governments were to commit flagrantly unconstitutional acts, people would rise in a revolution aimed at restoring the

---

\(^1\)For models in which separation of powers, treated as unitary actors, generates moderate equilibriums see Persson, Roland, and Tabelini (1996) and Dragu, Fan, and Kuklinski (2014).

\(^2\)As Palmer (1959: 262) observed,

The real problem (and it was a real problem) was to prevent the powers thus constituted from usurping more authority than they have been granted. According to one school, the several constituted powers of government, by watching and balancing and checking one another, were to prevent such usurpation. According to another school, which regarded the first school as undemocratic or mistrustful of the people, the people itself must maintain a constant vigilance and restraint upon the powers of government.
constitutional status quo. Montesquieu (1995: 19) thought that if any power succeeded to violate fundamental laws, “everything would unite against it”; there would be a revolution, “which would not change the form of government or its constitution: for revolutions shaped by liberty are but a confirmation of liberty.” In this tradition, Weingast (1997, 2015) argued that if a government were to conspicuously violate the constitution, cross a “bright line,” citizens would coordinate against it and, anticipating this reaction, the government would not commit such violations. Fearon (2011) thought that the same would occur if a government were not to hold an election or commit flagrant fraud. Hence, the combination of internal and external controls would make democratic institutions impregnable to the “encroaching spirit of power” (Madison, Federalist #48), the desire of politicians for enduring and unlimited power.

This is the view of democracy we inherited and this is the view we are now forced to question. By now we have seen Turkey under the government of AKP, Venezuela under Chavez and Maduro, Hungary under the second government of Fidesz, Poland under the second government of PiS, India under Narendra Modi, as well as the United States under Donald Trump. All these are, albeit to different extent, instances of democratic “backsliding” (or “deconsolidation,” “erosion,” “retrogression”): “a process of incremental (but ultimately still substantial) decay in the three basic predicates of democracy — competitive elections, liberal rights to speech and association, and the rule of law” (Huq and Ginsburg 2018a: 17). As this process advances, the opposition becomes unable to win elections or assume office if it wins, established institutions lose the capacity to control the executive, while manifestations of popular protest are repressed by force.

The first lesson we are learning from these experiences is that democratic institutions may not provide safeguards that protect them from being subverted by duly elected governments observing constitutional norms. When Hitler came to power, through an “authoritarian gap in the Weimar Constitution” (the Article 48 which allowed the President to empower the government to rule by decree; Bracher 1966: 119), the possibility
of a legal path to dictatorship was seen as a flaw of this particular constitution. Yet such
gaps may be generic. *Pace* Madison (Federalist #51), checks and balances do not operate
effectively when different powers of the government are controlled by the same party: as
Madison himself was almost immediately to discover (Dunn 2004: 47-61), constitutional
separation of powers is vulnerable to partisan interests. Courts, constitutional as well as
ordinary, can be packed, intimidated, or circumvented. Wholesale changes of constitutions,
amendments, or referendums can constitutionally overcome extant constitutional obstacles.
Public bureaucracies, including security agencies, can be instrumentalized for partisan
purposes. Public media can be controlled by partisan regulatory bodies, while private
media can be intimidated or destroyed financially. All such measures can be taken legally.
As Landau (2013: 192-3) observes, “The set of formal rules found in constitutions is
proving to be a mere parchment barrier against authoritarian and quasi-authoritarian
regimes. There is even worse news: existing democracy-protecting mechanisms in
international and comparative constitutional law have proven ineffective against this new
threat.”

Democratic deconsolidation need not entail violations of constitutionality. Thinking
about the United States, a constitutional lawyer writes, “If it happens here, it won’t
happen all at once.... Each step might be objectionable but not, by itself, alarming... there
will have been no single, cataclysmic point at which democratic institutions were
demolished... the steps toward authoritarianism will not always, or even usually, be
obviously illegal.... In fact, each step might conform to the letter of the law. But each step,
legal in itself, might undermine liberal democracy a little bit more.” (Strauss 2018: 365-6).
In a broader context, another constitutional lawyer concludes, “it is difficult to identify a
tipping point during the events: no single new law, decision or transformation seems
sufficient to cry wolf; only ex-post do we realize that the line dividing liberal democracy
from a fake one has been crossed: threshold moments are not seen as such when we live in
them” (Sadurski 2017: 5).
The puzzle entailed in destruction of democracy by backsliding is how a catastrophic situation can be gradually brought about by small steps, against which people who would be adversely affected do not react in time. As Ginsburg and Huq (2018b: 91) pose it, “The key to understanding democratic erosion is to see how discrete measures, which either in isolation or in the abstract might be justified as consistent with democratic norms, can nevertheless be deployed as mechanisms to unravel liberal constitutional democracy.” Why would citizens not react against such measures? One possibility is that they may acquiesce to them even when they fully realize their consequences, just not care sufficienly about preserving democracy. The other is that they may not realize in time that the government is consolidating its political domination. The latter possibility entails “stealth” (Varol 2015), which operates in a subtle way. Most acts by governments require legislation or executive decisions and, as such, are observable when they are announced. But whether these acts increase incumbent advantage and by how much can be observed only when their effects materialize. Imagine that a government extends voting rights to citizens residing abroad (Erdogan did it, but also Berlusconi), or it adopts legislation to require additional documentation at the polling place, or it relaxes the rules regulating private political financing. These measures are adopted according to constitutional provisions, so they are not unconstitutional in the sense of violating procedural norms. Moreover, governments offer democratic arguments in their favor: “We want to extend rights to all citizens,” “We want to prevent fraud,” “We are protecting the freedom of expression.” Only ex-post we learn that Erdogan won an election by the vote of Turks in Berlin, that Republicans won because poor people who did not have the required documents were prevented from voting, or that the Indian BNP enjoyed a massive financial advantage in the election. Such measures are anti-democratic only because

---

3Some measures taken by backsliding governments may not be observed by citizens. For example, the Polish ruling party, PiS, gradually altered the parliamentary procedure concerning introduction of new bills: the parliamentary rules say that bills proposed by the government must be subject to public hearings, while private member bills are not, and the government shifted to offering its proposal as private bills of its deputies (Sadurski 2017: 6).
their accumulation makes it more difficult for citizens to remove the incumbent government when they so desire. But this effect is not visible immediately and when it is realized it may be too late.

We think as follows. A government wins an election. This government decides whether to take steps to protect its tenure in office from any opposition. As Lust and Waldner (2015: 7) put it, “Backsliding occurs through a series of discrete changes in the rules and informal procedures that shape elections, rights and accountability. These take place over time, separated by months or even years.” Examples include changing electoral formulae, redistricting, changing voting qualifications, harassing the partisan opposition, imposing restrictions on NGOs, reducing judicial independence, using referendums to overcome constitutional barriers, imposing partisan control over the state apparatuses, or controlling the media. Concerned that such steps reduce their future ability to remove the incumbent government, citizens may turn against it. Yet whether the government can be removed from office depends on how far this process has already advanced, on how much advantage the incumbent government has already mustered. The obvious question is why some governments enter on this path while most refrain from it. The second is whether once a government takes such steps it can be ever stopped short of realization of complete domination. The third is whether the potential opposition would be able to remove the government and reverse this process.

The definition of “victory” is not as simple as it may appear. Electoral laws play an important role: in Turkey, the AKP won 34.3 percent of votes to obtain 66.0 percent of seats when it first assumed power in 2002, in Hungary Fidesz won 53 percent of votes and 68 of seats in 2010, in Poland PiS got 37.5 percent of votes and 51.0 of seats, in the 2016 election in the United States Donald Trump won with 46.09 percent of popular vote against 48.18 for his opponent. The ascension of Chavez to office in Venezuela was convoluted: the traditional parties actually won the legislative election of 1998, Chavez won the presidential election with 56.4 percent, the referendum for a new constitution was passed by 71.8 percent, then Chavez won a new election with 59.8 percent while his party obtained 44.4 percent of votes and 55.7 of seats in the legislative election of 2000.
The model we present below differs in two important aspects from the existing literature. In our model, the trade-off citizens face emerges endogenously, even if citizens do not value democracy for reasons such as liberty or political equality, as assumed by Svolik (2017), Graham and Svolik (2018), Miller 2018, Grillo and Prato (2019), or Nalepa, Vanberg, and Chiopris (2019). In our model, the value of democracy is the ability to remove any incumbent from office by democratic procedures when citizens believe that a different government would be better for them. Thus, when the incumbent government delivers outcomes that citizens enjoy, they must consider the possibility that if they retain the government, the incumbent may take steps that will make removing it more difficult in the future. Moreover, we model backsliding as a dynamic process in which at any time the incumbent government may or may not take steps of varying magnitude to consolidate its power and citizens decide at successive times whether or not to retain the incumbent given the advantage the incumbent has accumulated. (Other dynamic models of backsliding include Helmke, Kroeger, and Paine 2019 and Howell, Shepsle, and Walton 2019).

2. **Democracy and Its Vulnerabilities**

People want to be governed by competent politicians, so that even if they do not care about institutions per se, they value being able to select governments. In turn, because citizens would try to remove the incumbent government from office when they expect that a prospective competitor would be better, incumbents are insecure in office, which encourages them to take steps to protect their tenure from the voice of the people.

Democracy is an institutional arrangement in which citizens are able to replace incumbents whenever they believe that a different government would be better for them. Ideally, the outcomes of democratic procedures by which citizens select their rulers would depend only on citizens’ comparisons of the current and the prospective governments: which would make them better off? Yet even under democracy, citizens have to allow reasonably performing incumbents to gain some security in office. This reward must be
sufficient for the incumbent government not to want to push its domination too far, while still being tolerable for citizens.

This system is vulnerable, however, when the benefits delivered by the incumbent government are either very high or very low, relatively to the expectations about potential challengers. When people enjoy high current benefits, they see it as unlikely that a competitor would be better, so that they retain the incumbent, which in turn makes the incumbent free to exploit every opportunity to increase its advantage. When the incumbent government generates relatively low benefits, citizens are afraid that it would hold onto office. To preempt this possibility, citizens want to remove the incumbent even if they believe that the current challenger would be inferior. In turn, because the incumbent expects that citizens would want to remove it no matter what, it does everything possible to prevent it. Hence, in both situations the incumbent engages in backsliding, that is, it takes every opportunity to increase its chances to remain in office, all the way to the level at which it risks being removed by non-democratic means, such as coups or popular uprisings.

To flesh out intuitively the logic of this analysis, consider the following model, presented formally below. The incumbent government (“incumbent, he”) pursues policies that deliver some fixed level of well-being to a representative citizen. At each time the incumbent faces a challenger who may deliver a lower or a higher level of welfare, be worse or better than the incumbent. The probability that any future challenger would be better is given. At each time, the representative citizen (“she”) decides whether to retain the incumbent. “Incumbent advantage” is the probability that the incumbent remains in office when the representative citizen wants to remove him. This probability, in turn, depends on the actions of the incumbent: in each period the incumbent faces an opportunity to take a step of varying magnitude to increase his advantage, “subvert,” a step which he may or may not take.

---

5This definition encompasses situations in which the representative citizen is not decisive as well as those in which the incumbent loses an election and manages to stay in office.
Citizens do not observe immediately whether the incumbent took a step of a particular magnitude and learn about it only when the advantage has been realized.

These assumptions are sufficient to identify the conditions under which all the possible outcomes may occur. We characterize them by the competence of the incumbent, always relatively to the expected competence of potential challengers:

1) *Democracy*, in which the incumbent increases his advantage up to some critical level but stops short of complete domination and the citizen retains him as long as the challenger is inferior and this critical level is not passed, is sustained when the policies of the incumbent generate intermediate results. In these equilibriums, the incumbent feels sufficiently secure in office and the citizen preserves some chances of removing him when she believes the challenger would be better. Democracy is ideal when the incumbent is sufficiently secure in office even without any advantage, which occurs only when the potential challengers are expected not to be competent. Yet citizens face a trade-off between between the competence of the incumbent and the ability to replace him whenever a better challenger become available. Hence, the citizen allows the incumbent to accumulate some advantage as a reward for being reasonably competent. But if the incumbent generates some level of advantage which the citizen considers excessive, she wants to remove the incumbent regardless of the competence of the prospective challengers, and democracy deteriorates into polarization, characterized below.

2) *Populism*, in which the citizen retains the incumbent whenever she believes that the challenger would be inferior and the incumbent subverts in each period he holds office, emerges when the incumbent generates high benefits. Under such conditions, the citizen wants the government to be able to govern, even at the cost of weakening the checks on his authority, as in Acemoglu, Robinson, and Torbik (2013). Because the incumbent is highly competent, the citizen believes that any challenger is unlikely to be better, so that she values relatively little her future ability to remove the incumbent. Because the citizen is likely to retain the competent incumbent, this incumbent feels free to use all opportunities to increase
his advantage.

3) *Polarization*, in which the citizen always wants to remove the incumbent and the incumbent subverts in each period he holds office, emerges when the citizen does not benefit much from the policies of the incumbent and the incumbent already has considerable advantage. Under such conditions the relation between the incumbent and the citizen is polarized, in the sense that both actors behave unconditionally. Because the incumbent is incompetent, the citizen fears that he would increase his advantage, so that she wants to remove him preemptively, even when she believes that the challenger would be worse. Because the citizen unconditionally wants to remove him, the incumbent can only gain by taking every opportunity to increase his advantage.

Democracy is vulnerable in different ways to the two situations in which the incumbent takes every available step to increase his advantage. Whenever a newly elected government is highly successful in delivering whatever citizens want — say incomes grow rapidly\(^6\) or citizens intensely share the ideology of the government\(^7\) — populism is the unique equilibrium. Stealth plays no role: even if the citizen immediately observed what the incumbent is doing, she would retain him. In contrast, polarization may emerge gradually from democracy. Say the incumbent has already gained some level of advantage which the citizen still tolerates and that the incumbent sees a chance to make a big step that would make him almost immune from the preference of the citizen. Then the incumbent takes this step and democracy deteriorates into polarization, in which the incumbent does all to stay in office and the citizen wants to remove him unconditionally. Here stealth is important: had the citizen observed immediately that the incumbent is about to take this

\(^6\)Note that in Turkey per capita incomes grew at the annual rate of 4.4 under AKP government, Venezuela enjoyed spectacular growth between 2004 and 2011 (except for 2009) due to oil prices, Hungary grew at the rate of 3.5 under Fidesz, and the Polish incomes grew the fastest in Europe under PiS. (Data from PWT 9.0, end in 2014).

\(^7\)Examples include Islamization in Turkey, “Bolivarianism” in Venezuela, “preserving the purity of the nation” in Hungary, or “defending Christianity” in Poland,
step, she would remove him preemptively. But she observes the step only after its effect has been realized.

Populism, in turn, is vulnerable to downward shocks to the incumbent’s performance. When the incumbent’s competence falls below a critical threshold, the incumbent who has not yet crossed the threshold of advantage tolerable for the citizen stops taking additional steps, so that the equilibrium is democracy, while if the incumbent has already passed this threshold, the equilibrium is polarization.

These results characterize the actions of the incumbent and of the representative citizen. The dynamic outcomes, however, depend on whether the citizen gets a chance to decide whether the incumbent remains in office. Imagine that a new government is elected and it has no or little advantage, so that its tenure in office is subject only to the decision of the representative citizen. This state occurs as the initial condition in all the three types of equilibriums. Yet the dynamics of advantage differs. Under democracy, the incumbent gains advantage sufficient to make him feel secure in office and stops at some critical level which the citizen tolerates. But under populism and polarization the incumbent takes every opportunity to increase his advantage, backsliding all the way to the level at which he becomes vulnerable to being removed by non-democratic means, the level at which he crosses “the bright line.” Because under populism the incumbent is competent and the citizens wants to retain the incumbent only when she expects the challenger not to be better, while under polarization she wants to remove him unconditionally, the probability that the incumbent would complete all the seemingly democratic steps toward complete domination is higher under populism than under polarization. Yet in both situations, unless the incumbent is removed early into the process, he cannot be removed by

---

8Evo Morales in Bolivia is a good example. He was term-limited, called for a referendum to abolish the term limits, lost it, appealed to the constitutional courts which he previously appointed, received a favorable ruling, ran for office again, and declared his victory in spite of widespread allegations of fraud. The result was a popular rebellion, in which he was abandoned by the police and the armed forces.
3. **Model Setup**

An *incumbent* ("he") and a representative *citizen* ("she") interact in infinitely many periods $t = 0, 1, 2, \ldots$. In each period, the citizen decides whether she wants to keep the incumbent in office or try to remove him, $k_t = 0, 1$, where $k_t = 1$ indicates keeping the incumbent and $k_t = 0$ indicates wanting to remove him. Once the incumbent is removed, he loses power forever and a challenger assumes office.

Politicians vary in competence. The incumbent has a fixed level of competence $x$ and in each period $t$ he competes for office against a challenger whose level of competence $y_t = y, \overline{y}$ can be higher or lower than his, $y < x < \overline{y}$, and it is drawn from a stationary distribution $Pr\{y_t = \overline{y}\} = \gamma \in (0, 1)$. A challenger with $\overline{y}$ is more competent than the incumbent and is referred to as *superior*, while one with $y$ is less competent and referred to as *inferior*.

The citizen prefers more competent politicians. Her flow payoff in each period $t$ is the level of competence of the politician holding office in that period, represented by $x_t = y, x, \overline{y}$. As a result, the citizen would like the incumbent to be removed when a superior challenger is competing for office. The representative citizen, however, may be unable to remove the incumbent, either because she is not decisive or because the incumbent has means to stay in power against her wish. If the citizen wants to remove the incumbent, she fails with probability $p_t$. This probability measures the ability of the incumbent to stay in power against the will of the citizen and is referred to as incumbent *advantage*.

The goal of the incumbent is to stay in power as long as possible. His flow payoff in each period $t$ is $h_t = 0, 1$, where $h_t = 1$ indicates holding office. A greater advantage advances this goal by making the tenure of the incumbent more robust against opposition by the citizen. The incumbent begins with no advantage at all, $p_0 = 0$. In each period he is in office, the incumbent gets an opportunity to take a step that increases his advantage, "*subvert*." Letting
$b_t = 0, 1$ denote the action of the incumbent in period $t$, where $b_t = 1$ indicates subversion,

$$p_{t+1} = (p_t + (\pi - p_t)\theta_t b_t) h_t,$$

where $\pi \in (0, 1)$ and $\theta_t \in [0, 1]$ are drawn from a stationary distribution $F$ with full support. The magnitude of the subversion opportunity in period $t$ is given by $\theta_t$. If the incumbent subverts and manages to stay in office in period $t$, his advantage increases by $(\pi - p_t)\theta_t$ in the next period. The largest advantage the incumbent can possibly achieve by seemingly democratic steps is $\pi$, beyond which he can increase his advantage only by committing flagrantly unconstitutional acts and exposing himself to being removed by coups or popular rebellions.

At the beginning of each period, $p_t$ is observed both by the incumbent and the citizen. Only the incumbent, however, knows the scale of his subversion opportunity $\theta_t$ and his action $b_t$. In other words, the incumbent is able to subvert by stealth: the citizen is uncertain about whether the incumbent attempts to subvert and the extent to which he can subvert. She becomes aware of subversion only after its outcome, a greater advantage of the incumbent, is realized in the next period.

If the incumbent holds office in period $t$, in period $t + 1$ he gets a new opportunity to subvert and encounters a new challenger. If a challenger enters office in period $t$, he competes against a new challenger in the next period. The stage game in any period $t$ has the following sequence of moves.

1. $p_t$ is realized and observed by both the incumbent and the citizen.
2. $\theta_t$ and $y_t$ are drawn.
3. The incumbent observes $\theta_t$, then chooses whether to subvert, $b_t = 0, 1$.
4. The citizen observes $y_t$, then chooses whether to keep the incumbent, $k_t = 0, 1$.

The incumbent and the citizen discount future payoffs by $\beta \in (0, 1)$ and $\delta \in (0, 1)$,
respectively. The aggregate discounted payoff of the incumbent is

\[(1 - \beta) \sum_{t=0}^{\infty} \beta^t h_t\]

and that of the citizen is

\[(1 - \delta) \sum_{t=0}^{\infty} \delta^t x_t.\]

To ease exposition, \(y, \overline{y}, x\) can be rescaled to \(y = 0, \overline{y} = 1, \) and \(x \in (0, 1)\). The solution concept is Markov perfect equilibrium (equilibrium). The state variable is the current-period advantage of the incumbent, \(p\).

4. Strategies, Continuation Values, and Equilibrium

The incumbent decides whether to subvert conditional on his advantage \(p\) and the scale of his subversion opportunity \(\theta\) in the current period. This is equivalent to conditioning on \(p\) and \(q := p + (\pi - p)\theta\), the advantage of the incumbent in the next period given that he subverts and stays in office. Let

\[\mathcal{P} := \{(p, q) : p \in [0, \pi], q \in [p, \pi]\}\]

denote all the possible pairs of the current-period and attainable next-period advantage of the incumbent. Each strategy of the incumbent can be characterized by a function \(\sigma : \mathcal{P} \to \{0, 1\}\) that specifies a decision \(\sigma(q|p) = 0, 1\) whether to subvert in any period in which his advantage is \(p\) and will be \(q\) in the next period given that he subverts and stays in office. Conditional on the current-period advantage \(p\), the next-period advantage \(q\) of the incumbent given that
he subverts and stays in office is drawn from $F_p$ such that for each $q \in [p, \pi]$,

$$F_p(q) := \Pr \{ p + (\pi - p)\theta \leq q \} = F \left( \frac{q - p}{\pi - p} \right).$$

Clearly, $F_p$ stochastically dominates any $F_{p'}$ such that $p' < p$, which implies that with a greater advantage in the current period, the incumbent is more likely to get a greater advantage in the next period.

The citizen always prefers to remove the incumbent when the challenger is superior. When the challenger is inferior, the citizen decides whether to keep the incumbent in office conditional on the current-period advantage of the incumbent $p$. Each strategy of the citizen can be fully characterized by a function $\kappa : [0, \pi] \to \{0, 1\}$ that specifies a decision $\kappa(p) = 0, 1$ whether to keep the incumbent in any period the challenger is inferior and the incumbent has the advantage of $p$.

Let $v_\kappa(p)$ be the continuation value of the incumbent when he has the advantage of $p$, provided that the citizen has strategy $\kappa$. Note first that no matter what the incumbent does, the probability for him to stay in office is $p + (1 - p)(1 - \gamma)\kappa(p)$, which depends on his advantage $p$ and the decision of the citizen $\kappa(p)$. If the incumbent stays in office, he gets 1 from the current period. He expects to get $v_\kappa(q)$ in the future periods by taking an opportunity to subvert that improves the next-period advantage to $q$ and $v_\kappa(p)$ by giving up this opportunity. The incumbent subverts if and only if

$$v_\kappa(q) > v_\kappa(p).$$

(1)

In turn, $v_\kappa$ solves the Bellman equation:

$$v_\kappa(p) = (p + (1 - p)(1 - \gamma)\kappa(p)) \left( 1 - \beta + \beta \int_{p}^{\pi} \max \{ v_\kappa(q), v_\kappa(p) \} \, dF_p(q) \right).$$

(2)

Let $u_\sigma(p)$ be the continuation value of the citizen when the incumbent has the advantage
of $p$ and pursues strategy $\sigma$. If the incumbent stays in office, the citizen expects to get

$$(1 - \delta)x + \delta U_\sigma(p),$$

where

$$U_\sigma(p) := \int_p^\pi ((1 - \sigma(q|p)) u_\sigma(p) + \sigma(q|p)u_\sigma(q)) F_p(q)$$

is her expected payoff in the future periods while considering the possibility of subversion. If the incumbent is replaced by a challenger, the citizen gets 1 if the challenger is superior, while if the challenger is inferior, the citizen gets 0 in the current period and expects to get

$$\frac{\gamma}{1 - \delta(1 - \gamma)}$$

by choosing a superior challenger into office in the future periods. Therefore, the citizen prefers to keep the incumbent in office if and only if the challenger is inferior and

$$(1 - \delta)x + \delta U_\sigma(p) \geq \frac{\delta\gamma}{1 - \delta(1 - \gamma)}.$$  \hspace{1cm} (3)

In turn, $u_\sigma$ solves the Bellman equation:

$$u_\sigma(p) = p \begin{cases} (1 - \delta)x + \delta U_\sigma(p) \\ \gamma + (1 - \gamma) \max \left\{ (1 - \delta)x + \delta U_\sigma(p), \frac{\delta\gamma}{1 - \delta(1 - \gamma)} \right\} \end{cases} + (1 - p) \begin{cases} \text{citizen does not matter} \\ \text{citizen matters} \end{cases}.$$  \hspace{1cm} (4)

To interpret the right hand side of this equation, note that in each period the incumbent is not yet removed, the citizen faces two possibilities. First, with probability $p$, the incumbent holds office regardless of whether the citizen wants to keep him or tries to remove him. In this case, the choice of the citizen does not matter for who holds office and her expected
payoff is given by the first additive term in (4). Second, with probability $1 - p$, the tenure of the incumbent is subject to the decision of the citizen. In this case, the choice of the citizen matters and her expected payoff is given by the second additive term in (4).

An equilibrium is a strategy profile $(\sigma^*, \kappa^*)$ that is consistent with the optimality conditions, (1) for the incumbent and (3) for the citizen, in which the continuation values are derived from the Bellman equations (2) and (4). Formally,

**Definition 1 (Equilibrium).** $(\sigma^*, \kappa^*)$ constitutes an equilibrium if for each $(p, q) \in \mathcal{P}$,

1. $\sigma^*(q|p) = 1$ if and only if $v_{\kappa^*}(q) > v_{\kappa^*}(p)$, where $v_{\kappa^*}$ solves (2);

2. $\kappa^*(p) = 1$ if and only if $(1 - \delta)x + \delta U_{\sigma^*}(p) \geq \frac{\delta \gamma}{1 - \delta(1 - \gamma)}$, where $u_{\sigma^*}$ solves (4).

In what follows, we analyze the types of equilibrium under different values of two exogenous parameters: the level of competence of the incumbent $x$ and the expected level of competence of challengers $\gamma$.

### 5. Sustainability of Democracy

Democracy is a system in which citizens are able to choose governments by some established procedures. The incumbent may have some advantage but this advantage cannot exceed a critical level that citizens tolerate. Yet because citizens prefer more competent politicians and it is always possible for a challenger to be superior, the incumbent suffers from office insecurity. To mitigate this insecurity, the incumbent can accumulate advantage through subversion, reducing the ability of the citizen to remove him. In turn, although the citizen does not care about democracy per se, she has a derived preference, valuing the ability to select governments in the future. Indeed, in (4),

$$
\gamma + (1 - \gamma) \max \left\{ (1 - \delta)x + \delta U_{\sigma}(p), \frac{\delta \gamma}{1 - \delta(1 - \gamma)} \right\} \geq (1 - \delta)x + \delta U_{\sigma}(p),
$$

17
which implies that the citizen always prefers her choice to matter for who holds office. Hence, she opposes any attempt of the incumbent to increase his advantage. Due to this derived preference, the citizen may have an incentive to defend democracy by trying to remove the incumbent \textit{unconditionally}, preempting the risk of subversion, even at the cost of placing in office a challenger who the citizen knows to be worse than the incumbent. The interactive effects of the insecurity of the incumbent and the derived preference of the citizen make democracy vulnerable.

In this section, we first determine the stationary level of advantage that citizens tolerate under different conditions. Then we characterize two different situations under which democracy is vulnerable to backsliding, that is, in which the incumbents exploit every opportunity to subvert. Finally, we derive the conditions under which democracy is sustainable, robust against these vulnerabilities.

5.1. \textit{Stationary Advantage}

As a benchmark, we first analyze the case when the incumbent has a \textit{stationary} advantage of $p$, that is, when he refrains from subverting farther once he reaches this level. The question is whether the citizen would find a particular level of stationary advantage \textit{acceptable}, so that she would be willing to keep the incumbent in office as long as the challenger is worse in terms of competence.

Suppose the challenger is inferior. Given that the incumbent has a fixed level of advantage $p$, the citizen faces a trade-off between the competence of the politician in office in the current period and the ability to select more competent politicians in the future. If the incumbent is removed, the citizen suffers in the current period from the incompetence of an inferior challenger. The benefit, however, is that if a superior challenger appears in the future periods, the citizen would be to place him in office. Formally, the citizen prefers the incumbent to stay in office, so that $p$ is acceptable as a stationary advantage of the
incumbent, if and only if

\[(1 - \delta)x + \delta \bar{u}(p) \geq \frac{\delta \gamma}{1 - \delta (1 - \gamma)},\]  

(5)

where

\[
\bar{u}(p) = p \left( (1 - \delta)x + \delta \bar{u}(p) \right) \\
+ (1 - p) \left( \gamma + (1 - \gamma) \max \left\{ (1 - \delta)x + \delta \bar{u}(p), \frac{\delta \gamma}{1 - \delta (1 - \gamma)} \right\} \right)
\]

(6)

is the continuation value she gets in the future periods given that the incumbent holds office and stays with the advantage of \(p\).

**Lemma 1** (Acceptable Stationary Advantage). A stationary advantage of the incumbent \(p\) is acceptable for the citizen if and only if

\[p \leq \mathcal{P}(x, \gamma) := \left( \frac{1 - \delta (1 - \gamma)}{\delta \gamma} \right) x.\]

Moreover,

\[
\bar{u}(p) = \begin{cases} 
(1 - \delta)px + (1 - p)\gamma \frac{1 - \delta (1 - \gamma)}{1 - \delta (1 - (1 - p)\gamma)} & \text{if } p \leq \mathcal{P}(x, \gamma) \\
(1 - \delta)px + (1 - p)\gamma \frac{1 - \delta (1 - \gamma)}{1 - \delta p} & \text{if } p > \mathcal{P}(x, \gamma)
\end{cases}
\]

and for any strategy \(\sigma\) of the incumbent and \(p \in [0, \pi]\),

\[\bar{u}(\pi) \leq u_\sigma(p) \leq \bar{u}(p).\]

Therefore, the citizen would allow the incumbent to have a stationary advantage as long as this advantage is not too large. The critical level \(\mathcal{P}(x, \gamma)\) is exactly the ratio between the competence of the incumbent \(x\) and the current value of selecting a superior challenger into office in the future periods, \(\frac{\delta \gamma}{1 - \delta (1 - \gamma)}\). The advantage of the incumbent is the price the citizen
pays for his competence: she trades the ability to select better politicians in the future in exchange for the competence of the current government. The highest price the citizen is willing to pay, \( p(x, \gamma) \), in turn, is determined by the competence of the incumbent and by her value of the ability to select future governments.

The benchmark with a stationary advantage facilitates the subsequent analysis by providing two bounds on the continuation value of the citizen. Specifically, regardless of the strategy of the incumbent, the continuation value the citizen gets when the incumbent has an advantage of \( p \) cannot be higher than \( \pi(p) \), which is her continuation value if the incumbent stops at \( p \), and it cannot be lower than \( \pi(\pi) \), which is her continuation value if the incumbent already has the largest possible advantage. The intuition is as follows. Each strategy of the incumbent entails a path along which his advantage grows. But because the advantage of the incumbent can only harm the citizen, regardless of how it grows, the citizen cannot be better off than in the case when this advantage does not grow at all and she cannot be worse off than in the case when this advantage reaches its maximum. These two bounds help in identifying the two vulnerabilities of democracy.

5.2. Vulnerabilities of Democracy: Populism and Polarization

**Proposition 1** (Vulnerabilities of Democracy).

1. (Populism) If \( x \geq \frac{\pi \delta \gamma}{1 - \delta (1 - \gamma)} \), then there exists a unique equilibrium \((\sigma^*, \kappa^*)\) such that \( \sigma^*(q|p) = 1 \) and \( \kappa^*(p) = 1 \) for all \((p, q) \in \mathcal{P}\).

2. (Polarization) If \( x < \frac{\pi \delta \gamma}{1 - \delta (1 - \gamma)} \), then for any equilibrium \((\sigma^*, \kappa^*)\), \( \sigma^*(q|p) = 1 \) and \( \kappa^*(p) = 0 \) for all \((p, q) \in \mathcal{P}\) such that \( p > \bar{p}(x, \gamma) \).

Democracy is vulnerable to two situations in which the incumbent never stops short of maximizing his advantage for as long as he remains in office: populism and polarization. The difference between these two situations is the behavior of the citizen. Under populism, the citizen keeps the incumbent in office as long as the challenger is less competent. In contrast,
under polarization, the citizen unconditionally opposes the incumbent, wanting to remove him regardless of the quality of the challenger. We refer to this situation as “polarization” because both actors behave unconditionally: the incumbent always subverts and the citizen always wants to remove him.

Populism emerges when the incumbent is sufficiently competent. Because the incumbent is competent, the citizen accepts any advantage the incumbent can possibly achieve through subversion. The condition for populism is equivalent to $\bar{p}(x, \gamma) \geq \pi$, which means that any stationary advantage of the incumbent is acceptable for the citizen.

The worst possible case for the citizen is when the advantage reaches its upper limit $\pi$ or, equivalently, when the citizen has the lowest possible ability to select politicians by democratic procedures. In this case, the citizen has the strongest incentive to remove the incumbent. Yet even this incentive may not be strong enough to overcome the temptation to procrastinate with removing the incumbent whenever the challenger in the current period is inferior. If the incumbent is replaced by an inferior challenger, the citizen gets 0 in the current period and a continuation value of $\frac{\gamma}{1-\delta(1-\gamma)}$. If the citizen waits one period, she gets a better payoff $x > 0$ in the current period but a worse continuation value,

$$\frac{(1 - \delta)\pi x + (1 - \pi)\left(\frac{\gamma}{1-\delta(1-\gamma)}\right)}{1 - \delta\pi} < \frac{\gamma}{1 - \delta(1 - \gamma)}$$

in the future periods due to the reduced ability to select governments. The citizen prefers to wait for one period if and only if

$$\frac{(1 - \delta)x + \delta\left(\frac{(1 - \delta)\pi x + (1 - \pi)\left(\frac{\gamma}{1-\delta(1-\gamma)}\right)}{1 - \delta\pi}\right)}{1 - \delta(1 - \gamma)} = \frac{\delta\gamma}{1 - \delta(1 - \gamma)} + \frac{1 - \delta}{1 - \delta\pi}\left(x - \frac{\pi\delta\gamma}{1 - \delta(1 - \gamma)}\right)$$

$$\geq \frac{\delta\gamma}{1 - \delta(1 - \gamma)},$$

which is equivalent to $x \geq \frac{\pi\delta\gamma}{1 - \delta(1 - \gamma)}$. When this condition holds, the citizen procrastinates
with removing the incumbent whenever the challenger is inferior. In turn, because the citizen procrastinates, subversion is a costless option for the incumbent and he takes every opportunity to do so.

Polarization is triggered when the incumbent is not too competent and has an overly large advantage. When the incumbent is not competent enough to enjoy populist support, \( \bar{p}(x, \gamma) < \pi \), so that the citizen cannot allow the incumbent to have too much advantage. Any advantage of the incumbent exceeding the critical level, \( p > \bar{p}(x, \gamma) \), would make the citizen want to remove him regardless of the quality of the challenger.

When the current challenger is inferior, the benefit of keeping the incumbent in office is his still superior competence. The cost, however, is that if the incumbent increases his advantage, the citizen would have a lower chance to select a more competent politician when a superior challenger appears. If this opportunity cost is sufficiently large, the citizen would be willing to replace the incumbent even with an inferior challenger, tolerating incompetence in the current period. The opportunity cost of keeping the incumbent in office is lowest when the incumbent stays with the current-period advantage, not subverting to increase it. In this case, the citizen is most tempted to put off removing the incumbent. By waiting for one period to remove the incumbent, the citizen gets \( x \) in the current period and a continuation value of

\[
\frac{(1 - \delta)x + (1 - p)\left(\frac{\gamma}{1 - \delta(1 - \gamma)}\right)}{1 - \delta p},
\]

which is strictly decreasing in \( p \). The citizen would rather replace the incumbent with an
inferior challenger immediately rather than wait for one period to do so if and only if

\[
(1 - \delta)x + \delta \left( \frac{(1 - \delta)px + (1 - p) \left( \frac{\gamma}{1 - \delta(1 - \gamma)} \right)}{1 - \delta p} \right) > \frac{\delta \gamma}{1 - \delta(1 - \gamma)} \left( 1 - \frac{1 - \delta}{1 - \delta p} \left( p - \bar{p}(x, \gamma) \right) \right) > \frac{\delta \gamma}{1 - \delta(1 - \gamma)},
\]

which is equivalent to \( p > \bar{p}(x, \gamma) \). Therefore, when the advantage of the incumbent is sufficiently large, the citizen always wants to remove the incumbent right away, regardless of who the challenger is. She is punishing the incumbent as much as she can. Such a harsh punishment, in turn, leaves the incumbent no choice but to rely on his advantage as the only way to secure his position in office. Therefore, if the incumbent has already gained an advantage over the critical level, anticipating unconditional opposition by the citizen, he exploits every possible chance to reinforce it.

Although the incumbent acts similarly in the populism and polarization, he does so for different reasons. Under populism, the incumbent subverts due to the leniency of the citizen, which makes subversion costless; while under polarization, the incumbent subverts due to the unconditional opposition by the citizen, which makes subversion necessary.

5.3. Sustainable Democracy

The possibility for the incumbent to subvert by stealth imposes an additional difficulty for the sustainability of democracy. On the one hand, to sustain democracy, the subversion by the incumbent cannot go unpunished. Given the incentive of the incumbent to subvert, the leniency of the citizen results in populism. Punishing subversion, on the other hand, triggers polarization, in which the unconditional opposition by the citizen encourages the incumbent to subvert and the citizen may suffer from the incompetence of inferior challengers. Hence, for democracy to be sustainable, the punishment against subversion must be credible, but it
should be implementable only \textit{off equilibrium}, as a threat to keep the incumbent behave.

Therefore, the first requirement of sustainable democracy is that the incumbent not be too competent, $x < \frac{\pi \delta \gamma}{1 - \delta (1 - \gamma)}$, so that populism would not arise. If this requirement is satisfied, any advantage of the incumbent above the critical level $\mathcal{P}(x, \gamma)$ would trigger polarization as a punishment against increased subversion. The second requirement is that the punishment polarization entails be sufficiently severe, so that fearing to trigger it, the incumbent would restrain himself. Formally,

**Definition 2 (Sustainability of Democracy).** \textit{Democracy is sustainable if}

1. \textit{(No Populism)} $x < \frac{\pi \delta \gamma}{1 - \delta (1 - \gamma)}$;

2. \textit{(No Risk of Polarization)} there exists an equilibrium $(\sigma^*, \kappa^*)$ such that $\sigma^*(q|p) = 0$ for all $(p, q) \in \mathcal{P}$ such that $p \leq \mathcal{P}(x, \gamma)$ and $q > \mathcal{P}(x, \gamma)$.

Two consequences follow when these two requirements are met. First, the incumbent refrains from farther subversion once achieving a level of advantage that is acceptable for the citizen as a stationary advantage. Second, given that the incumbent never attempts to gain an advantage above the critical level, the citizen keeps the incumbent in office as long as the challenger is inferior. Therefore, under sustainable democracy, the citizen is willing to select whoever she believes to be better and she reserves some ability to do so.

**Proposition 2 (Sustainability of Democracy).** \textit{Democracy is sustainable if and only if}

$$(\gamma, x) \in \mathcal{D} := (0, \overline{\gamma}) \times \left( \mathcal{X}(\gamma), \frac{\pi \delta \gamma}{1 - \delta (1 - \gamma)} \right),$$

where $\overline{\gamma} \in (1 - \pi, 1)$, $\mathcal{X}(\gamma) = 0$ if $\gamma \leq 1 - \pi$, and $\mathcal{X}(\gamma) \in \left(0, \frac{\pi \delta \gamma}{1 - \delta (1 - \gamma)}\right)$ is strictly increasing in $\gamma$ if $1 - \pi < \gamma < \overline{\gamma}$.

As shown in Figure 1, democracy is sustainable when both the expected competence of challengers and of the incumbent fall into the region $\mathcal{D}$ enclosed by the blue and red curves.
This implies that for democracy to be sustainable, both the incumbent and the challenger must have intermediate levels of competence. If the incumbent is highly competent or challengers are too incompetent in expectation (outside of the blue curve), populism emerges, and the incumbent can subvert without being punished. If the incumbent is too incompetent or challengers are competent in expectation (outside of the red curve), democracy is under the risk of polarization: any time opportunities are available, the incumbent would take them, crossing the critical level at the cost of triggering polarization. The intuition of the first case is clear, but that of the second case is subtle. When the incumbent is incompetent or when he has to compete against challengers who are expected to be quite competent, the citizen is eager to remove the incumbent. In this case, the incumbent prefers to increase his advantage, rather than relying on support of the citizen which he would obtain only when the challenger is even less competent. Stealth plays a role here: the incumbent is able to take a polarizing step only because the citizen cannot immediately observe the additional advantage the incumbent would gain.

To conclude, because of the office insecurity of the incumbent and the derived preference of the citizen, they have conflictive preferences over democratic backsliding. This conflict exposes democracy to the danger of populism and polarization. But when both the incumbent and the expected challenger have intermediate competence, democracy is sustainable, meaning that it is free from the danger of backsliding. Although the incumbent would take small steps for extra office security, fearing to trigger unconditional opposition by the citizen, he would stop short of completely compromising democratic means of being replaced.
Figure 1: Sustainability of Democracy

\[
\text{Blue: } x = \frac{\pi \delta \gamma}{1 - \delta (1 - \gamma)}; \quad \text{Red: } x = \bar{x}(\gamma);
\]

Region enclosed by blue and red: \(D\), democracy is sustainable.

6. **Competence-Democracy Trade-off**

Backsliding is not a danger when the incumbent has no intention to completely subvert democratic means of selecting governments. But to increase office security, the incumbent may still take steps to gain moderate advantage. The question, then, is to what extent the incumbent can be restrained. In this section, we answer this question by characterizing the conditions under which the incumbent can be stopped from farther subverting once he achieves a particular level of advantage.

**Definition 3.** \(p < \pi\) is an equilibrium stationary advantage of the incumbent if there exists an equilibrium \((\sigma^*, \kappa^*)\) such that \(\sigma^*(q|p') = 0\) for all \((p', q) \in \mathcal{P}\) such that \(p' \leq p\) and \(q > p\).

Democracy is sustainable if and only if an equilibrium stationary advantage exits. In particular, if \(p = 0\) is an equilibrium stationary advantage, there exists an equilibrium in which the incumbent starts with no advantage at all and never attempts to achieve any through subversion. In this equilibrium, the citizen is perfectly able to select whoever
she believes to be better and this ability is never compromised. Democracy, therefore, is 
institutionally perfect.

**Proposition 3** (Competence-Democracy Trade-off). \( p < \pi \) is an equilibrium stationary advantage of the incumbent if and only if

\[
(\gamma, x) \in D^*(p) = (0, \gamma^*(p)) \times \left[ \frac{p\delta \gamma}{1 - \delta (1 - \gamma)}, x^*(\gamma|p) \right] \subseteq D,
\]

where \( \gamma^*(0) = 1 - \pi \) and \( \gamma^*(p) \in (1 - \pi, \pi) \) is strictly increasing in \( p \) if \( p > 0 \) while \( x^*(\gamma|p) \in \left( \frac{p\delta \gamma}{1 - \delta (1 - \gamma)}, \frac{\pi \delta \gamma}{1 - \delta (1 - \gamma)} \right) \) is strictly increasing in \( \gamma \) and \( p \).

A particular \( p \) can be sustained as an equilibrium stationary advantage under three conditions. First, \( p \) can be sustained in equilibrium only if it is acceptable, \( p \leq p(x, \gamma) \) or, equivalently, \( x \geq \frac{p\delta \gamma}{1 - \delta (1 - \gamma)} \). Given that \( p \) is acceptable, it can be an equilibrium stationary advantage of the incumbent if and only if he is expected to compete against sufficiently incompetent challengers, \( \gamma < \gamma^*(p) \), and he is not too competent himself, \( x \leq x^*(\gamma|p) \).

Intuitively, the incumbent can be stopped at \( p \) only by the threat of the citizen to trigger polarization if the incumbent would gain any advantage greater than \( p \). This threat can be effective in stopping the incumbent only when it is sufficient and credible at the same time. The threat is sufficient if the incumbent prefers to stay with \( p \) rather than gaining any advantage larger than \( p \) and facing unconditional opposition by the citizen. The incumbent has enough office security under \( p \), so that he would not be willing to move into polarization, if and only if it is sufficiently unlikely for him to compete against a superior challenger. For the threat of unconditional opposition to be credible, it has to be in the best interest of the citizen to remove the incumbent, regardless of the quality of the challenger, once the incumbent gains any advantage larger than \( p \) and would continue subverting. Keeping \( p \) fixed, the citizen has a stronger incentive to remove the incumbent when he is less competent, which is when the citizen gains more from selecting a superior challenger. Therefore, it is the incompetence of the incumbent that renders the threat of the citizen credible.
A more interesting implication of Proposition 3 is that the citizen always faces a competence-democracy trade-off, that is, a trade-off between enjoying a better government at the present and being able to choose better governments in the future. First, $\gamma^*(p)$ is strictly increasing in $p$. The advantage of the incumbent and the incompetence of challengers both help the incumbent to be secure in office. When the incumbent has a greater advantage, the threat of unconditional opposition by the citizen can be sufficient to deter the incumbent from farther subverting, even if he has to compete against more competent challengers. Second, $x^* (\gamma|p)$ is strictly increasing in $p$. The advantage and the incompetence of the incumbent both incentivize the citizen to remove him from office. When the incumbent has a greater advantage, the threat of unconditional removal by the citizen can be credible even if the incumbent is more competent and thus the citizen benefits more by keeping him in office. Therefore, if the citizen faces more competent politicians, either the incumbent or the challenger, she must bear with a greater advantage of the incumbent.

Hence, far less than perfect politicians are the necessary evil for democracy to be institutionally perfect. The sustainability of perfect democracy, in which the incumbent has an equilibrium stationary advantage of $p = 0$, requires both the incumbent and the challengers to be highly incompetent. The challengers have to be highly incompetent in expectation, making it unnecessary for the incumbent to gain any advantage. The incumbent has to be highly incompetent, so that the citizen is able to impose a credible threat on the incumbent that once he obtains any advantage, however small, he would have to face her unconditional opposition. An institutionally imperfect, perhaps really existing, democracy is a system, in which the incumbent is allowed to gain a moderately high level of advantage as a reward of being more competent.
7. Could It Happen Here?

An obvious avenue for future research is to investigate whether different systems of separation of power are equally vulnerable to the breakdown of purely institutional controls. Aguilar Rivera (2008) explains why systems of strict separation of powers are particularly vulnerable to usurpation of power by the president who controls the armed forces. Maeda (2010) finds that presidential systems are more vulnerable to subversion from above than parliamentary ones. Yet as of now we know little empirically about the vulnerability of particular institutional arrangements to backsliding.

What we did learn is that defending democracy imposes a difficult challenge on individual citizens. To act against the government that may be destroying democracy, people must consider the effect of the government’s actions on their future ability to replace it by a better one. Even if individuals have consistent time preferences (Ackerlof 1991), they must
be able to calculate the effect of seemingly democratic steps. This is a formidable task and it should not be surprising if many people could not perform it. Consider a sequence of events in which the government first adopts legislation to require additional documentation at the polling place, then has its cronies buy a major opposition newspaper, then re-maps electoral districts, and then stuffs the bodies that administer and supervise elections. To prevent backsliding, people must both value their ability to choose governments and understand that while each of the backsliding measures is perfectly legal, their cumulative effect is to protect the incumbent from being defeated in the future.

The fact is that backsliding governments have enjoyed continued popular support. To our best knowledge, the only case in which a backsliding government lost an election and left office is of Sri Lanka in 2015, and this outcome resulted from massive defections from the ruling coalition, with the winner previously a minister in the outgoing government. Other backsliding governments suffered temporary reversals but were able to recover and continue: With 40.9 percent of votes, the AKP failed to win a majority of seats in the election of June 7, 2015 but it called for a new election and won 49.5 percent of the vote five months later. Three years later, in June 2018, Erdogan won the presidential election with 52.6 percent. In Poland, PiS won an absolute majority of parliamentary seats in October 2019. In Hungary, Fidesz and its allies won a reelection in April 2018 with 44.9 percent of the vote. In Venezuela, Chavez won a re-election in 2006 with 62.8 percent of the vote and again in 2012 with 55.1 percent. He enjoyed majority support in the polls and the opposition became majoritarian only after his death (Venezuelabarometro). And in the United States, the popularity of President Trump hovers narrowly around 40 percent regardless of anything. The implication must be that either many people do not care at all about preserving the ability to remove him or that they do not see the consequences of supporting him when they answer survey questions.

The optimism that citizens would effectively threaten governments that commit transgressions against democracy and thus prevent them from taking this path is sadly
unfounded. This view is based on the assumption that when a government commits some acts that flagrantly threaten liberty, violate constitutional norms, or undermine democracy, people will unify against it. Yet people may not react to such violations even when they observe them or they may be unable to assess their consequences. And if citizens do not stop the government from taking some series of legal steps, it may be too late to prevent it from doing whatever it wants.
REFERENCES


Appendix to “Democracy and Its Vulnerabilities: Dynamics of Democratic Backsliding”

A.1. Lemma 1

Proof of Lemma 1. First, suppose \( p > \overline{p}(x, \gamma) \). Assume
\[
(1 - \delta)x + \delta \overline{u}(p) \geq \frac{\delta \gamma}{1 - \delta (1 - \gamma)}.
\]
Then,
\[
\overline{u}(p) = \frac{(1 - \delta)px + (1 - p)\gamma}{1 - \delta (1 - (1 - p)\gamma)},
\]
so that
\[
(1 - \delta)x + \delta \overline{u}(p) = \frac{\delta \gamma}{1 - \delta (1 - \gamma)} + \left( \frac{1 - \delta}{1 - \delta (1 - (1 - p)\gamma)} \right) \left( x - \frac{p \delta \gamma}{1 - \delta (1 - \gamma)} \right)
\]
\[
< \frac{\delta \gamma}{1 - \delta (1 - \gamma)},
\]
a contradiction. Hence, it must be true that
\[
(1 - \delta)x + \delta \overline{u}(p) < \frac{\delta \gamma}{1 - \delta (1 - \gamma)}
\]
and
\[
\overline{u}(p) = \frac{(1 - \delta)px + (1 - p)\left( \frac{\gamma}{1 - \delta (1 - \gamma)} \right)}{1 - \delta p}.
\]
The same steps prove that if \( p \leq \overline{p}(x, \gamma) \),
\[
(1 - \delta)x + \delta \overline{u}(p) \geq \frac{\delta \gamma}{1 - \delta (1 - \gamma)},
\]
\[
\overline{u}(p) = \frac{(1 - \delta)px + (1 - p)\gamma}{1 - \delta (1 - (1 - p)\gamma)}.
\]
Second, consider any $\sigma$ and $p$. For any $p' \geq p$,

\[
    u_\sigma(p') \leq p' \left( (1 - \delta)x + \delta \max_{q \in [p', \pi]} u_\sigma(q) \right)
    + (1 - p') \left( \gamma + (1 - \gamma) \max \left\{ (1 - \delta)x + \delta \max_{q \in [p', \pi]} u_\sigma(q), \frac{\delta \gamma}{1 - \delta(1 - \gamma)} \right\} \right)
    \leq p \left( (1 - \delta)x + \delta \max_{q \in [p, \pi]} u_\sigma(q) \right)
    + (1 - p) \left( \gamma + (1 - \gamma) \max \left\{ (1 - \delta)x + \delta \max_{q \in [p, \pi]} u_\sigma(q), \frac{\delta \gamma}{1 - \delta(1 - \gamma)} \right\} \right),
\]

so that

\[
    \max_{q \in [p, \pi]} u_\sigma(q) \leq p \left( (1 - \delta)x + \delta \max_{q \in [p, \pi]} u_\sigma(q) \right)
    + (1 - p) \left( \gamma + (1 - \gamma) \max \left\{ (1 - \delta)x + \delta \max_{q \in [p, \pi]} u_\sigma(q), \frac{\delta \gamma}{1 - \delta(1 - \gamma)} \right\} \right). \tag{1}
\]

By the definition of $\pi(p)$, this implies that $u_\sigma(p) \leq \max_{q \in [p, \pi]} u_\sigma(q) \leq \pi(p)$. Similarly, for any $p' \geq p$,

\[
    u_\sigma(p') \geq p' \left( (1 - \delta)x + \delta \min_{q \in [p', \pi]} u_\sigma(q) \right)
    + (1 - p') \left( \gamma + (1 - \gamma) \max \left\{ (1 - \delta)x + \delta \min_{q \in [p', \pi]} u_\sigma(q), \frac{\delta \gamma}{1 - \delta(1 - \gamma)} \right\} \right)
    \geq \pi \left( (1 - \delta)x + \delta \min_{q \in [p, \pi]} u_\sigma(q) \right)
    + (1 - \pi) \left( \gamma + (1 - \gamma) \max \left\{ (1 - \delta)x + \delta \min_{q \in [p, \pi]} u_\sigma(q), \frac{\delta \gamma}{1 - \delta(1 - \gamma)} \right\} \right),
\]

so that

\[
    \min_{q \in [p, \pi]} u_\sigma(q) \geq \pi \left( (1 - \delta)x + \delta \min_{q \in [p, \pi]} u_\sigma(q) \right)
    + (1 - \pi) \left( \gamma + (1 - \gamma) \max \left\{ (1 - \delta)x + \delta \min_{q \in [p, \pi]} u_\sigma(q), \frac{\delta \gamma}{1 - \delta(1 - \gamma)} \right\} \right). \tag{2}
\]
By the definition of $\pi(\pi)$, this implies that $u_\sigma(p) \geq \min_{q \in [p, \pi]} u_\sigma(q) \geq \pi(\pi)$. ■

A.2. Proposition 1

Proof of Proposition 1. First, assume $x \geq \frac{\pi \delta \gamma}{1 - \delta (1 - \gamma)}$, so that $\pi \leq \bar{p}(x, \gamma)$. Then, for any strategy $\sigma$ of the incumbent, because $u_\sigma(p) \geq \pi(\pi)$ holds for all $p$, 

$$U_\sigma(p) = \int_p^\pi \left( (1 - \sigma(q|p))u_\sigma(p) + \sigma(q|p)u_\sigma(q) \right) dF_p(q) \geq \pi(\pi),$$

so that 

$$(1 - \delta)x + \delta U_\sigma(p) \geq (1 - \delta)x + \delta \pi(\pi) \geq \frac{\delta \gamma}{1 - \delta (1 - \gamma)}$$

holds for all $p$. Hence, the citizen has a dominant strategy $\kappa^*(p) = 1$ for all $p$. Given that the citizen has this strategy, 

$$v_{\kappa^*}(p) = (p + (1 - p)(1 - \gamma)) \left( 1 - \beta + \beta \int_p^\pi \max\{v_{\kappa^*}(q), v_{\kappa^*}(p)\} dF_p(q) \right).$$

The right hand side is strictly increasing in $p$ if $v_{\kappa^*}$ is increasing, which implies that $v_{\kappa^*}$ is strictly increasing. In turn, $v_{\kappa^*}(q) \geq v_{\kappa^*}(p)$ for any $(p, q) \in \mathcal{P}$. Hence, it is optimal for the incumbent to have $\sigma^*(p, q) = 1$ for all $(p, q) \in \mathcal{P}$.

Second, assume $x < \frac{\pi \delta \gamma}{1 - \delta (1 - \gamma)}$, so that $\bar{p}(x, \gamma) < \pi$. Consider any equilibrium $(\sigma^*, \kappa^*)$ and any $p > \bar{p}(x, \gamma)$. Then, because $\pi$ is strictly decreasing, $u_{\sigma^*}(p') \leq \pi(p') < \pi(p)$ for all $p' > p$, so that 

$$U_\sigma(p) = \int_p^\pi \left( (1 - \sigma(q|p))u_\sigma(p) + \sigma(q|p)u_\sigma(q) \right) dF_p(q) < \pi(p).$$
and

\((1 - \delta)x + \delta U_\sigma(p) < (1 - \delta)x + \delta \overline{u}(p) < \frac{\delta \gamma}{1 - \delta(1 - \gamma)}\).

Hence, it must be true that \(\kappa^*(p) = 0\). Given that the citizen has \(\kappa^*(p) = 0\) for all \(p > \overline{p}(x, \gamma)\),

\(v_{\kappa^*}(p) = p \left( 1 - \beta + \beta \int_p^\pi \max \left\{ v_{\kappa^*}(q), v_{\kappa^*}(p) \right\} dF_p(q) \right)\)

for all \(p > \overline{p}(x, \gamma)\). The right hand side is strictly increasing in \(p > \overline{p}(x, \gamma)\) if \(v_{\kappa^*}\) is increasing in \(p > \overline{p}(x, \gamma)\), which implies that \(v_{\kappa^*}\) is strictly increasing in \(p > \overline{p}(x, \gamma)\). In turn, \(v_{\kappa^*}(q) \geq v_{\kappa^*}(p)\) for all \((p, q) \in \mathcal{P}\) such that \(p > \overline{p}(x, \gamma)\). Hence, it must be true that \(\sigma^*(q|p) = 1\) for all \((p, q) \in \mathcal{P}\) such that \(p > \overline{p}(x, \gamma)\). ■

### A.3. Proposition 2

In the proof of Proposition 2, we use the following functions and their properties. First, for each \(p\), let

\(\psi(p) = p \left( 1 - \beta + \beta \int_p^\pi \max \left\{ \psi(q), \psi(p) \right\} dF_p(q) \right)\). \hspace{1cm} (A.1)

\(\psi(p)\) is the continuation value of the incumbent when he has the advantage of \(p\) and the citizen tries to remove him whenever he has any advantage at least as large as \(p\), that is, \(\kappa(p') = 0\) for all \(p' \geq p\).

Second, for each \((p', p) \in \mathcal{P}\) and \(\gamma\), let

\(\hat{\psi}(p', p|\gamma) = (p' + (1 - p')(1 - \gamma)) \left( 1 - \beta + \beta \left( \int_{p'}^p \max \left\{ \hat{\psi}(q, p|\gamma), \hat{\psi}(p', p|\gamma) \right\} dF_{p'}(q) \right) \right. \right. \left. \left. + \int_{p}^\pi \max \left\{ \psi(q), \hat{\psi}(p', p|\gamma) \right\} dF_{p'}(q) \right) \right)\). \hspace{1cm} (A.2)
\( \hat{v}(p', p|\gamma) \) is the continuation value of the incumbent when he has the advantage of \( p' \) and the citizen retains him if he has any advantage at most as large as \( p \) and tries to remove him if he has any advantage larger than \( p \), that is, \( \kappa(p') = 1 \) if \( p' \leq p \) and \( \kappa(p') = 0 \) if \( p' > p \).

**Lemma A.1.** \( v(p) \) is strictly increasing in \( p \) and \( \hat{v}(p', p|\gamma) \) is strictly increasing in \( p' \), \( p \) and strictly decreasing in \( \gamma \).

**Proof.** The right hand side of (A.1) is strictly increasing in \( p \) if \( v \) is increasing, which implies that \( v \) is strictly increasing. The right hand side of (A.2) is strictly increasing in \( p' \) if \( \hat{v}(p', p|\gamma) \) is increasing in \( p' \), which implies that \( \hat{v}(p', p|\gamma) \) is strictly increasing in \( p' \). Similarly, the right hand side of (A.2) is strictly decreasing in \( \gamma \) if \( \hat{v}(p', p|\gamma) \) is decreasing in \( \gamma \), which implies that \( \hat{v}(p', p|\gamma) \) is strictly decreasing in \( \gamma \). At last, because \( \hat{v}(p', p|\gamma) \) is strictly increasing in \( p' \), \( \hat{v}(p, p|\gamma) > \hat{v}(p', p|\gamma) \) for any \( p' < p \). Because \( v \) is strictly increasing,

\[
\begin{align*}
 v(p) &= p \left( 1 - \beta + \beta \int_p^\pi v(q) dF_p(q) \right) \\
 &\leq p \left( 1 - \beta + \beta \int_p^\pi \max \{v(q), \hat{v}(p, p|\gamma)\} dF_p(q) \right) \\
 &< (p + (1 - p)(1 - \gamma)) \left( 1 - \beta + \beta \int_p^\pi \max \{v(q), \hat{v}(p, p|\gamma)\} dF_p(q) \right) \\
 &= \hat{v}(p, p|\gamma).
\end{align*}
\]

Hence, \( \hat{v}(p, p|\gamma) > \max \{v(p), \hat{v}(p', p|\gamma)\} \) for any \( p' < p \), so that the right hand side of (A.2) is strictly increasing in \( p \) provided that \( \hat{v}(p', p|\gamma) \) is increasing in \( p \). This implies that \( \hat{v}(p', p|\gamma) \) is strictly increasing in \( p \). \( \blacksquare \)

**Lemma A.2.** If \( x < \frac{\pi \delta \gamma}{1 - \delta(1 - \gamma)} \), then for any rationalizable \( \kappa \),

1. \( v_\kappa(p) = v(p) \) for all \( p > \bar{p}(x, \gamma) \);

2. \( v_\kappa(p) \leq \bar{v}(p|x, \gamma) := \hat{v}(p, \bar{p}(x, \gamma)|\gamma) \) for all \( p \leq \bar{p}(x, \gamma) \).
Proof. Let \( \kappa \) be a rationalizable strategy of the citizen. First, for any \( p > p(x, \gamma) \), \( \kappa(p) = 0 \), so that

\[
v_\kappa(p) = p \left( 1 - \beta + \beta \int_p^\pi \max \{v_\kappa(q), v_\kappa(p)\} \, dF_p(q) \right),
\]

which implies that \( v_\kappa(p) = \nu(p) \). Second, because \( v_\kappa(p) = \nu(p) \) for all \( p > p(x, \gamma) \),

\[
v_\kappa(p) = (p + (1 - p)(1 - \gamma)\kappa(p)) \left( 1 - \beta + \beta \int_p^\pi \max \{v_\kappa(q), v_\kappa(p)\} \, dF_p(q) \right)
\]

\[
= (p + (1 - p)(1 - \gamma)\kappa(p)) \left( 1 - \beta + \beta \left( \int_p^{\pi(x, \gamma)} \max \{v_\kappa(q), v_\kappa(p)\} \, dF_p(q) + \int_{\pi(x, \gamma)}^\pi \max \{\nu(q), v_\kappa(q)\} \, dF_p(q) \right) \right)
\]

\[
\leq (p + (1 - p)(1 - \gamma)) \left( 1 - \beta + \beta \left( \int_p^{\pi(x, \gamma)} \max \{v_\kappa(q), v_\kappa(p)\} \, dF_p(q) + \int_{\pi(x, \gamma)}^\pi \max \{\nu(q), v_\kappa(q)\} \, dF_p(q) \right) \right).
\]

holds for all \( p \leq \pi(x, \gamma) \). This implies that \( v_\kappa(p) \leq \hat{\nu}(p, \pi(x, \gamma)|\gamma) = \nu(p|x, \gamma) \) for all \( p \leq \pi(x, \gamma) \). ■

In the rest of this section, we prove Proposition 2 in two steps. First, we show that democracy is sustainable if and only if \( x < \frac{\pi \delta \gamma}{1 - \delta(1 - \gamma)} \) and \( \nu(0|x, \gamma) > \nu(\pi) \). Second, we derive the conditions in terms of \((\gamma, x)\) under which \( \nu(p|x, \gamma) > \nu(\pi) \).

Lemma A.3. Democracy is sustainable if and only if \( x < \frac{\pi \delta \gamma}{1 - \delta(1 - \gamma)} \) and \( \nu(0|x, \gamma) > \nu(\pi) \).

Proof. It is clear that democracy is sustainable only if \( x < \frac{\pi \delta \gamma}{1 - \delta(1 - \gamma)} \). Now we prove that provided \( x < \frac{\pi \delta \gamma}{1 - \delta(1 - \gamma)} \), democracy is sustainable if and only if \( \nu(0|x, \gamma) > \nu(\pi) \).

For necessity, suppose democracy is sustainable and let \((\sigma^*, \kappa^*)\) be an equilibrium for which \( \sigma^*(q|p) = 0 \) for all \((p, q) \in P\) such that \( p \leq \pi(x, \gamma) \) and \( q > \pi(x, \gamma) \). Because \( \sigma^*(\pi|0) = 0 \), it must be true that \( v_{\kappa^*}(0) > v_{\kappa^*}(\pi) = \nu(\pi) \). In turn, \( \nu(0|x, \gamma) \geq v_{\kappa^*}(0) > \nu(\pi) \).
For sufficiency, suppose \( v(0|x, \gamma) > v(\pi) \) and consider \((\sigma^*, \kappa^*)\) such that
\[
\sigma^*(q|p) = \begin{cases} 
0, & p \leq \overline{p}(x, \gamma), q > \overline{p}(x, \gamma) \\
1, & \text{otherwise}
\end{cases}
\]
and
\[
\kappa^*(p) = \begin{cases} 
0, & p > \overline{p}(x, \gamma) \\
1, & \text{otherwise}
\end{cases}
\]
for each \((p, q) \in \mathcal{P}\). Democracy is sustainable if \((\sigma^*, \kappa^*)\) constitutes an equilibrium. First,
\[
p((1 - \delta)x + \delta U_{\sigma^*}(p)) + (1 - p) \left( \gamma + (1 - \gamma) \max \left\{ (1 - \delta)x + \delta U_{\sigma^*}(p), \frac{\delta \gamma}{1 - \delta(1 - \gamma)} \right\} \right)
\]
is greater than \( u(\overline{p}(x, \gamma)) \) for each \( p < \overline{p}(x, \gamma) \) if \( u_{\sigma^*}(p) \) is at least as large as \( u(\overline{p}(x, \gamma)) \) for each \( p < \overline{p}(x, \gamma) \). This implies that \( u_{\sigma^*}(p) > u(\overline{p}(x, \gamma)) \) for each \( p < \overline{p}(x, \gamma) \). In turn,
\[
(1 - \delta)x + \delta U_{\sigma^*}(p) = (1 - \delta)x + \delta \left( \int_{\overline{p}(x, \gamma)}^{p} u_{\sigma^*}(q) dF_p(q) + (1 - F_p(\overline{p}(x, \gamma))) u(\overline{p}(x, \gamma)) \right)
\]
\[
\geq (1 - \delta)x + \delta u(\overline{p}(x, \gamma))
\]
\[
= \frac{\delta \gamma}{1 - \delta(1 - \gamma)}
\]
for all \( p \leq \overline{p}(x, \gamma) \) and the inequality holds strictly if \( p < \overline{p}(x, \gamma) \). Hence, given that the incumbent has \( \sigma^* \), it is optimal for the citizen to pursue \( \kappa^* \). Second, by the definition of \( \hat{v}, v_{\kappa^*}(p) = \hat{v}(p, \overline{p}(x, \gamma)|\gamma) = v(p|x, \gamma) \) for each \( p \leq \overline{p}(x, \gamma) \). Because \( v(p|x, \gamma) \) is strictly increasing in \( p \leq \overline{p}(x, \gamma), \forall(p|x, \gamma) \leq v(q|x, \gamma) \) for all \((p, q) \in \mathcal{P}\) such that \( p \leq \overline{p}(x, \gamma) \) and \( q \leq \overline{p}(x, \gamma) \). In addition, because \( v(p) \) is strictly increasing in \( p > \overline{p}(x, \gamma), \)
\[
v(p|x, \gamma) \geq v(0|x, \gamma) > v(\pi) \geq v(q)
\]
for all \((p, q) \in \mathcal{P}\) such that \(p \leq \underline{p}(x, \gamma)\) and \(q > \overline{p}(x, \gamma)\). Hence, given that the citizen has \(\kappa^*\), it is optimal for the incumbent to pursue \(\sigma^*\). Therefore, \((\sigma^*, \kappa^*)\) constitutes an equilibrium and, thus, democracy is sustainable. ■

**Proof of Proposition 2.** Provided that \(x < \frac{\pi \delta \gamma}{1 - \delta (1 - \gamma)}\), democracy is sustainable if and only if \(\nu(0|x, \gamma) > \nu(\pi)\). Note that because \(\hat{\nu}(0, p, | \gamma)\) is strictly increasing in \(p\) and strictly decreasing in \(\gamma\), \(\nu(0|x, \gamma) = \hat{\nu}(0, \overline{p}(x, \gamma) | \gamma)\) is strictly increasing in \(x\) and strictly decreasing in \(\gamma\).

First, suppose \(\gamma \leq 1 - \pi\). Because \(\nu(p|x, \gamma)\) is strictly increasing in \(p\),

\[
\nu(0|x, \gamma) > (1 - \gamma)(1 - \beta + \beta \nu(0|x, \gamma)) \geq \frac{(1 - \beta)(1 - \gamma)}{1 - \beta (1 - \gamma)} \geq \frac{(1 - \beta)\pi}{1 - \beta \pi} = \nu(\pi).
\]

Second, because \(\hat{\nu}(0, \pi | \gamma)\) is strictly decreasing in \(\gamma\),

\[
\hat{\nu}(0, \pi | 1 - \pi) > (1 - \gamma)(1 - \beta + \beta \hat{\nu}(0, \pi | 1 - \pi)) = \frac{(1 - \beta)\pi}{1 - \beta \pi} = \nu(\pi),
\]

and \(\hat{\nu}(0, \pi | 1) = 0\), there exists a unique \(\gamma \in (1 - \pi, 1)\) such that \(\hat{\nu}(0, \pi | \gamma) = \nu(\pi)\) and \(\hat{\nu}(0, \pi | \gamma) > \nu(\pi)\) if and only if \(\gamma < \gamma\). Suppose \(\gamma \geq \gamma\), then \(\nu(p|x, \gamma) \leq \nu(\pi)\). Third, suppose \(1 - \pi < \gamma < \gamma\). Because \(\nu(0|x, \gamma)\) is strictly increasing in \(x\),

\[
\nu(0|0, \gamma) = \frac{(1 - \beta)(1 - \gamma)}{1 - \beta (1 - \gamma)} < \frac{(1 - \beta)\pi}{1 - \beta \pi} = \nu(\pi),
\]

and

\[
\nu \left(0| \frac{\pi \delta \gamma}{1 - \delta (1 - \gamma)}, \gamma \right) = \hat{\nu}(0, \pi | \gamma) > \nu(\pi),
\]

there exists a unique \(\underline{x}(\gamma) \in \left(0, \frac{\pi \delta \gamma}{1 - \delta (1 - \gamma)}\right)\) such that \(\nu(0|x(\gamma), \gamma) = \nu(\pi)\) and \(\nu(0|x, \gamma) > \nu(\pi)\) if and only if \(x > \underline{x}(\gamma)\). Because \(\nu(0|x, \gamma)\) is strictly increasing in \(x\) and
strictly decreasing in $\gamma$, $\varepsilon(\gamma)$ is strictly increasing in $\gamma$. At last, extending the definition of $\varepsilon(\gamma)$ so that $\varepsilon(\gamma) = 0$ for $\gamma \leq 1 - \pi$, then provided that $x < \frac{\pi \delta \gamma}{1-\delta(1-\gamma)}$, $\pi(0|x, \gamma) > \mu(\pi)$, so that democracy is sustainable, if and only if $\gamma < \overline{\gamma}$ and $x > \varepsilon(\gamma)$. ■

### A.4. Proposition 3

In the proof of Proposition 3, we use the following function and its properties. For each $p$, $x$, and $\gamma$, let

$$u(p|x, \gamma) = p \left( (1 - \delta)x + \delta U(p|x, \gamma) \right) + (1 - p) \left( \gamma + (1 - \gamma) \max \left\{ (1 - \delta)x + \delta U(p|x, \gamma), \frac{\delta \gamma}{1 - \delta(1 - \gamma)} \right\} \right),$$  \hspace{1cm} (A.3)

where

$$U(p|x, \gamma) := \int_{p}^{\pi} u(q|x, \gamma) dF_p(q)$$

$u(p|x, \gamma)$ is the continuation value of the citizen when the incumbent has the advantage of $p$ and does not stop subverting once achieving any advantage larger than $p$, that is, $\sigma(q|p') = 1$ for all $(p', q) \in P$ such that $p' > p$.

**Lemma A.4.** $u(p|x, \gamma)$ is strictly decreasing in $p$, strictly increasing $x, \gamma$, and $u(p|x, \gamma) - \frac{\gamma}{1-\delta(1-\gamma)}$ is strictly decreasing in $\gamma$.

**Proof.** The right hand side of (A.3) is strictly decreasing in $p$ and strictly increasing in $x$ and $\gamma$ if $u(p|x, \gamma)$ is decreasing in $p$ and increasing in $x, \gamma$. This implies that $u(p|x, \gamma)$ is strictly decreasing in $p$ and strictly increasing $x, \gamma$. For almost every $\gamma$, if $\frac{\partial}{\partial \gamma} u(p|x, \gamma) \leq \frac{\partial}{\partial \gamma} \left( \frac{\gamma}{1-\delta(1-\gamma)} \right)$
for all $p$, then differentiating the right hand side of (A.3) with respect to $\gamma$ either yields

$$
(p + (1 - p)(1 - \gamma)) \delta \int_p^\pi \frac{\partial}{\partial \gamma} u(p|x, \gamma) dF_p(q)
\leq (p + (1 - p)(1 - \gamma)) \delta \frac{\partial}{\partial \gamma} \left( \frac{\gamma}{1 - \delta(1 - \gamma)} \right)
< \frac{\partial}{\partial \gamma} \left( \frac{\gamma}{1 - \delta(1 - \gamma)} \right)
$$

or

$$
p\delta \int_p^\pi \frac{\partial}{\partial \gamma} u(p|x, \gamma) dF_p(q) + (1 - p) \frac{\partial}{\partial \gamma} \left( \frac{\gamma}{1 - \delta(1 - \gamma)} \right)
\leq (1 - p + p\delta) \frac{\partial}{\partial \gamma} \left( \frac{\gamma}{1 - \delta(1 - \gamma)} \right)
< \frac{\partial}{\partial \gamma} \left( \frac{\gamma}{1 - \delta(1 - \gamma)} \right).
$$

This implies that $\frac{\partial}{\partial \gamma} u(p|x, \gamma) < \frac{\partial}{\partial \gamma} \left( \frac{\gamma}{1 - \delta(1 - \gamma)} \right)$ for all $p$. In turn, $u(p|x, \gamma) - \frac{\gamma}{1 - \delta(1 - \gamma)}$ is strictly decreasing in $\gamma$. ■

**Lemma A.5.** For any $\sigma$, $u_{\sigma}(p) \geq u(p|x, \gamma)$ for all $p \in [0, \pi]$.

**Proof.** Consider any $\sigma$ and $p$. Because $u(q|x, \gamma)$ is strictly increasing in $q$,

$$
\int_p^\pi u(q|x, \gamma) dF_p(q) \leq \int_p^\pi \left( (1 - \sigma(q|p)) u(p|x, \gamma) + \sigma(q|p) u(q|x, \gamma) \right) dF_p(q) =: U_{\sigma}(p),
$$

so that

$$
u(p|x, \gamma) \leq (1 - p)((1 - \delta)x + \delta U_{\sigma}(p)) + p \left( \gamma + (1 - \gamma) \max \left\{ (1 - \delta)x + \delta U_{\sigma}(p), \frac{\delta \gamma}{1 - \delta(1 - \gamma)} \right\} \right).
$$

By the definition of $u_{\sigma}(p)$, this implies that $u_{\sigma}(p) \geq u(p|x, \gamma)$. ■
In the rest of this section, we prove Proposition 3 in two steps. First, we show that \( p < \pi \) is an equilibrium stationary advantage if and only if \( \frac{p \delta \gamma}{1 - \delta(1 - \gamma)} \leq x < \frac{\pi \delta \gamma}{1 - \delta(1 - \gamma)} \), \((1 - \delta)x + \delta U(p|x, \gamma) \leq \frac{\delta \gamma}{1 - \delta(1 - \gamma)}\), and \( \tilde{v}(0, p|\gamma) > \varphi(\pi) \). Second, we derive the conditions in terms of \((\gamma, x)\) under which \((1 - \delta)x + \delta U(p|x, \gamma) \leq \frac{\delta \gamma}{1 - \delta(1 - \gamma)}\) and \( \tilde{v}(0, p|\gamma) > \varphi(\pi) \).

Lemma A.6. \( p < \pi \) is an equilibrium stationary advantage if and only if \( \frac{p \delta \gamma}{1 - \delta(1 - \gamma)} \leq x < \frac{\pi \delta \gamma}{1 - \delta(1 - \gamma)}\), \((1 - \delta)x + \delta U(p|x, \gamma) \leq \frac{\delta \gamma}{1 - \delta(1 - \gamma)}\), and \( \tilde{v}(0, p|\gamma) > \varphi(\pi) \).

Proof. Clearly, \( p \) is an equilibrium stationary advantage only if it is acceptable, that is, \( p \leq \overline{p}(x, \gamma) \), which is equivalent to \( \frac{p \delta \gamma}{1 - \delta(1 - \gamma)} \leq x < \frac{\pi \delta \gamma}{1 - \delta(1 - \gamma)} \). Now we prove that provided \( x \geq \frac{p \delta \gamma}{1 - \delta(1 - \gamma)} \), \( p \) is an equilibrium stationary advantage if and only if \((1 - \delta)x + \delta U(p|x, \gamma) \leq \frac{\delta \gamma}{1 - \delta(1 - \gamma)}\) and \( \tilde{v}(0, p|\gamma) > \varphi(\pi) \).

For necessity, let \((\sigma^*, \kappa^*)\) be an equilibrium such that \( \sigma^*(q|p') = 0 \) for all \((p', q) \in \mathcal{P}\) such that \( p' \leq p \) and \( q > p \). First, given \( \sigma^* \) and \( \kappa^* \), for any \( p' \leq p \), the continuation value of the incumbent is

\[
v_{\kappa^*}(p') = (p + (1 - p)(1 - \gamma))\kappa^*(p')
\]

\[
(1 - \beta + \beta \left( \int_{p'}^p \max \{v_{\kappa^*}(q), v_{\kappa^*}(p')\} \, dF_{p'}(q) + \int_{p}^{\pi} v_{\kappa^*}(p') \, dF_{p'}(q) \right) \leq (p + (1 - p)(1 - \gamma))
\]

\[
(1 - \beta + \beta \left( \int_{p'}^p \max \{v_{\kappa^*}(q), v_{\kappa^*}(p')\} \, dF_{p'}(q) + \int_{p}^{\pi} \max \{v_{\kappa^*}(q), v_{\kappa^*}(p')\} \, dF_{p'}(q) \right).
\]

By the definition of \( \tilde{v} \), this implies that \( v_{\kappa^*}(p') \leq \tilde{v}(p', p|\gamma) \) for all \( p' \leq p \). But because \( \sigma^*(\pi|0) = 0 \), it must be true that \( v_{\kappa^*}(0) > v_{\kappa^*}(\pi) = \varphi(\pi) \). In turn, \( \tilde{v}(0, p|\gamma) \geq v_{\kappa^*}(0) > \varphi(\pi) \).

Second, assume that \((1 - \delta)x + \delta U(p|x, \gamma) > \frac{\delta \gamma}{1 - \delta(1 - \gamma)}\). Because \( u(q|x, \gamma) \) is strictly decreasing in \( q \),

\[
(1 - \delta)x + \delta U_{\sigma^*}(p) > \frac{\delta \gamma}{1 - \delta(1 - \gamma)}.
\]

45
But because $u_{\sigma^*}(q) \leq u(q|x,\gamma)$ for all $q \geq p$, 

$$(1 - \delta)x + \delta U_{\sigma^*}(p) \geq (1 - \delta)x + \delta U_{\sigma^*}(p) > \frac{\delta \gamma}{1 - \delta(1 - \gamma)}.$$ 

Hence, for some $p' > p$ sufficiently close to $p$, $\kappa^*(q) = 1$ for all $q < p'$. This implies that 

$$v_{\kappa^*}(q) = (q + (1 - q)(1 - \gamma))(1 - \beta + \beta \int_q^\pi \max \{v_{\kappa^*}(q'), v_{\kappa^*}(q)\} \, dF_q(q'))$$ 

is strictly increasing in $p \leq q < p'$. In turn, $v_{\kappa^*}(p) < v_{\kappa^*}(q)$ for all $p < q < p'$, so that $\sigma^*(q|p) = 1$ for all $p < q < p'$, a contradiction. Hence, $(1 - \delta)x + \delta U(p|x,\gamma) \leq \frac{\delta \gamma}{1 - \delta(1 - \gamma)}$.

For sufficiency, suppose $(1 - \delta)x + \delta U(p|x,\gamma) \leq \frac{\delta \gamma}{1 - \delta(1 - \gamma)}$ and $\bar{v}(0,p|\gamma) > \bar{v}(\pi)$ and consider $(\sigma^*,\kappa^*)$ such that 

$$\sigma^*(q|p') = \begin{cases} 
0, & p' \leq p, q > p \\
1, & \text{otherwise} 
\end{cases}$$

and 

$$\kappa^*(p') = \begin{cases} 
0, & p' > p \\
1, & \text{otherwise} 
\end{cases}$$

for each $(p',q) \in \mathcal{P}$. $p$ is an equilibrium stationary advantage if $(\sigma^*,\kappa^*)$ constitutes an equilibrium. First, 

$$p' ((1 - \delta)x + \delta U_{\sigma^*}(p')) + (1 - p') \left(\gamma + (1 - \gamma) \max \left\{ (1 - \delta)x + \delta U_{\sigma^*}(p'), \frac{\delta \gamma}{1 - \delta(1 - \gamma)} \right\} \right)$$

is greater than $\bar{\pi}(p)$ for each $p' < p$ if $u_{\sigma^*}(p')$ is at least as large as $\bar{\pi}(p)$ for each $p' < p$. This
implies that \( u_{\sigma^*}(p') > \overline{u}(p) \) for each \( p' < p \). In turn,

\[
(1 - \delta)x + \delta U_{\sigma^*}(p) = (1 - \delta)x + \delta \left( \int_{p'}^p u_{\sigma^*}(q)dF_{\nu'}(q) + (1 - F_{\nu'}(p)) \overline{u}(p) \right) \\
\geq (1 - \delta)x + \delta \overline{u}(p) \\
\geq \frac{\delta \gamma}{1 - \delta(1 - \gamma)}
\]

for all \( p' \leq p \). Clearly, for any \( p' > p \), \( u_{\sigma^*}(p') = \underline{u}(p'|x, \gamma) \), so that

\[
(1 - \delta)x + \delta U_{\sigma^*}(p') = (1 - \delta)x + \delta \underline{U}(p'|x, \gamma) \\
< (1 - \delta)x + \delta \underline{U}(p|x, \gamma) \\
\leq \frac{\delta \gamma}{1 - \delta(1 - \gamma)}.
\]

Hence, given that the incumbent has \( \sigma^* \), it is optimal for the citizen to pursue \( \kappa^* \). Second, by the definition of \( \hat{v} \) and \( \underline{v} \), \( v_{\kappa^*}(p') = \hat{v}(p', p|\gamma) \) for each \( p' \leq p \) and \( v_{\kappa^*}(p') = \underline{v}(p') \) for each \( p' > p \). Because \( \hat{v}(p', p|\gamma) \) is strictly increasing in \( p' \leq p \), \( \hat{v}(p', p|\gamma) \leq \hat{v}(q, p|\gamma) \) for all \( (p', q) \in P \) such that \( p' \leq p \) and \( q \leq p \). In addition, because \( \underline{v}(p') \) is strictly increasing in \( p' > p \),

\[
\hat{v}(p', p|\gamma) \geq \hat{v}(0, p|\gamma) \geq \underline{v}(p) \geq \underline{v}(q)
\]

for all \( (p', q) \in P \) such that \( p \leq p' \) and \( q > p' \). Hence, given that the citizen has \( \kappa^* \), it is optimal for the incumbent to pursue \( \sigma^* \). Therefore, \( (\sigma^*, \kappa^*) \) constitutes an equilibrium and, thus, \( p \) is an equilibrium stationary advantage.

**Proof of Proposition 3.** Consider a \( p < \pi \). Provided that \( \frac{\pi \delta \gamma}{1 - \delta(1 - \gamma)} \leq x < \frac{\pi \delta \gamma}{1 - \delta(1 - \gamma)} \), \( p \) is an equilibrium stationary advantage if and only if \( (1 - \delta)x + \delta \underline{U}(p|x, \gamma) \leq \frac{\delta \gamma}{1 - \delta(1 - \gamma)} \) and \( \hat{v}(0, p|\gamma) \geq \underline{v}(\pi) \). Note that because \( \underline{u}(p'|x, \gamma) - \frac{\delta \gamma}{1 - \delta(1 - \gamma)} \) is strictly decreasing in \( \gamma \) for all \( p' \),
\[(1 - \delta)x + \delta U(p|x, \gamma) - \frac{\delta\gamma}{1 - \delta(1 - \gamma)} \text{ is strictly decreasing in } \gamma.\]

First, for any \(x \leq \frac{p\delta\gamma}{1 - \delta(1 - \gamma)}\), \(p \geq p(x, \gamma)\), so that
\[
(1 - \delta)x + \delta U(p'|x, \gamma) \leq (1 - \delta)x + \delta \pi(p') < (1 - \delta)x + \delta \pi(p) \leq \frac{\delta\gamma}{1 - \delta(1 - \gamma)}
\]
holds for all \(p' \geq p\). Hence, \((1 - \delta)x + \delta U(p|x, \gamma) < \frac{\delta\gamma}{1 - \delta(1 - \gamma)}\). For any \(x \geq \frac{\pi\delta\gamma}{1 - \delta(1 - \gamma)}\),
\[
(1 - \delta)x + \delta U(p'|x, \gamma) > (1 - \delta)x + \delta \pi(\pi|x, \gamma) \geq (1 - \delta)x + \delta \pi(\pi) \geq \frac{\delta\gamma}{1 - \delta(1 - \gamma)}
\]
holds for all \(p' \leq \pi\). Hence, \((1 - \delta)x + \delta U(p|x, \gamma) > \frac{\delta\gamma}{1 - \delta(1 - \gamma)}\). Therefore, there exists a unique \(x^*(\gamma|p) \in \left(\frac{\pi\delta\gamma}{1 - \delta(1 - \gamma)}, \frac{\pi\delta\gamma}{1 - \delta(1 - \gamma)}\right)\) such that
\[
(1 - \delta)x^*(\gamma|p) + \delta U(p|x^*(\gamma|p), \gamma) = \frac{\delta\gamma}{1 - \delta(1 - \gamma)}
\]
and \((1 - \delta)x + \delta U(p|x, \gamma) \leq \frac{\delta\gamma}{1 - \delta(1 - \gamma)}\) if and only if \(x \leq x^*(\gamma|p)\). The monotonicity properties of \(x^*(\gamma|p)\) are due to \((1 - \delta)x + \delta U(p|x, \gamma) - \frac{\delta\gamma}{1 - \delta(1 - \gamma)}\) being strictly decreasing in \(p\), strictly increasing in \(x\), and strictly decreasing in \(\gamma\).

Second, at \(\gamma = 1 - \pi\),
\[
\hat{\nu}(0, p|1 - \pi) \geq (1 - \gamma) (1 - \beta + \beta\hat{\nu}(0, p|1 - \pi)) \geq \frac{(1 - \beta)\pi}{1 - \beta\pi} = \nu(\pi).
\]
The equalities hold if and only if \(p = 0\). At \(\gamma = \gamma\),
\[
\hat{\nu}(0, p|\gamma) < \hat{\nu}(0, p(x, \gamma)|\gamma) = \nu(0|x, \gamma) = \nu(\pi).
\]
Hence, \(\hat{\nu}(0, p|\gamma) \leq \nu(\pi)\) and the equality holds if and only if \(p = \overline{p}(x, \gamma)\). Therefore, there exists a unique \(\gamma^*(p) \in [1 - \pi, \gamma]\) such that \(\hat{\nu}(0, p|\gamma^*(p)) = \nu(\pi)\) and \(\hat{\nu}(0, p|\gamma) > \nu(\pi)\) if and only if \(\gamma < \gamma^*(p)\). Moreover, \(\gamma^*(0) = 1 - \pi\) and because \(\hat{\nu}(0, p|\gamma)\) is strictly increasing in \(p\) and strictly decreasing in \(\gamma\), \(\gamma^*(p)\) is strictly increasing in \(p\).
Therefore, provided that \( \frac{p\delta\gamma}{1-\delta(1-\gamma)} \leq x < \frac{\pi\delta\gamma}{1-\delta(1-\gamma)} \), (1 - \( \delta \))x + \delta U(x, \gamma) \leq \frac{\delta\gamma}{1-\delta(1-\gamma)} \) and \( \tilde{v}(0, p|\gamma) > v(\pi) \), so that \( p \) is an equilibrium stationary advantage, if and only if \( x \leq x^*(\gamma|p) \) and \( \gamma < \gamma^*(p) \).

At last,

\[
\varpi \left( 0 \left| \frac{p\delta\gamma}{1-\delta(1-\gamma)}, \gamma \right. \right) = \tilde{v}(0, p|\gamma) > v(\pi)
\]

if and only if \( \gamma < \gamma^*(p) \). By the definition of \( \varpi(\gamma) \), this implies that \( \varpi(\gamma) < \frac{p\delta\gamma}{1-\delta(1-\gamma)} \) if and only if \( \gamma < \gamma^*(p) \). Hence, \( x^*(p|\gamma) > \frac{p\delta\gamma}{1-\delta(1-\gamma)} > \varpi(\gamma) \) for all \( \gamma < \gamma^*(p) \), so that \( D^*(p) \subseteq D \).