Democracy and its vulnerabilities: Dynamics of democratic backsliding

Zhaotian Luo

Adam Przeworski

University of Chicago

New York University

luozhaotian@uchicago.edu

adam.przeworski@nyu.edu

August 4, 2020

Abstract

The puzzle entailed in erosion of democracy by backsliding is how a catastrophic situation can be gradually brought about by steps against which people who would be adversely affected do not react in time. We investigate conditions which render democracy impregnable to backsliding and conditions which make it vulnerable. Democracy is sustainable, free from the threat of backsliding, when opposing politicians are neither very attractive nor very unattractive to citizens. To sustain it, citizens must allow more appealing incumbents to gain some security in office. Backsliding occurs either under “populism,” when citizens knowingly consent to erosion of democracy because they find the incumbent highly appealing, or under “polarization,” when citizens oppose the incumbent regardless of the attractiveness of the challenger, so that the incumbent can remain in office only by backsliding.

1. INTRODUCTION

Most, if not all, democracies 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 institutional systems have varied but the goal everywhere was to establish 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

---

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

²As 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 puzzle entailed in destruction of democracy by backsliding is how a catastrophic situation can be gradually brought about by 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.” We think as follows. A government wins an election.³ This government is characterized by its attractiveness to citizens. Its appeal may be based on some policy outcomes, such as income growth,⁴ but it may be also be rooted in ideological affinities, such as Islamization in Turkey, “Bolivarianism” in Venezuela, “preserving the purity of the nation” in Hungary, “defending Christianity” in Poland, Hinduism in India, or xenophobia in the United States. The government decides whether to take steps to protect its tenure in office from any opposition, steps such as 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 whenever they would prefer a challenger, forward-looking citizens decide whether to retain or remove 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.

---

³The 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.

⁴Note 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).

The obvious question is what makes democracy vulnerable to backsliding. What are the different ways in which backsliding could occur? What may induce governments to take actions that limit the ability of the opposition to remove it? Once a government begins to backslide, can it be ever stopped short of realization of complete domination? Would the potential opposition be able to remove a backsliding government? Are there some conditions under which democracy is impregnable to backsliding?

As noted by Grillo and Prato (2019), the burgeoning literature on backsliding differs in two dimensions: the motivation of the incumbent, which can be either to increase policy discretion or to extend tenure in office, and the identity of the restraining agent, which can be either “horizontal” (other institutions, parties, elites) or “vertical” (citizens, voters). The distinction between seeking policy discretion and extending tenure in office is not perfectly sharp in that the incumbent may need policy discretion in order to advance his partisan advantage: Viktor Orbán, for example, centralized the revenue Hungarian cities receive from parking violations, in order to reduce the budget of the Budapest municipal government, controlled by the opposition. Yet as distinct from Howell and Wolton (2018) or Howell, Shepsle, and Wolton (2019), we model the process by which the government accumulates electoral advantage, not the process by which the incumbent weakens horizontal checks on his policy discretion. This distinction is important because the erosion of institutional constraints is a lasting legacy, while partisan advantage of an incumbent is not inherited by a partisan opponent.

Table 1: Literature on democratic backsliding

		Restrainer	
		<i>Horizontal</i>	<i>Vertical</i>
Motivation	<i>Discretion</i>	Howell and Wolton (2018)	Graham and Svolik (2019)
		Howell, Shepsle, and Wolton (2019)	Grillo and Prato (2019)
	<i>Tenure</i>	Helmke, Kroeger, and Paine (2019)	Nalepa, Vanberg, and Chiopris (2019)
			<i>This paper</i>

While our model fits into tenure-voter cell, as distinct from static models (Graham and Svulik 2019, Grillo and Prato 2019, Nalepa, Vanberg, and Chiopris 2019), we treat the value of democracy as evolving endogenously. For us, 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, and this value evolves endogenously as a consequence of actions taken by the incumbent rather than being fixed once and for all.

In what follows, we first preview the assumptions of the model and summarize the central results in non-technical terms. The model is presented next. A conclusion closes the paper. Most proofs are in the Appendix.

2. DEMOCRACY AND ITS VULNERABILITIES

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, whether materially or ideologically? People want to be governed by politicians they find appealing, so that even if they do not care about institutions per se, they value being able to decide who would govern them. In turn, because citizens would try to remove the incumbent government from office when they find a competitor more attractive, incumbents are insecure in office, which encourages them to take steps to protect their tenure from the voice of the people. Hence, even under democracy, citizens may have to allow incumbents to gain some security in office. This reward must be sufficient for the incumbent government not want to push its domination too far, while still being tolerable for citizens.

This system is vulnerable, however, when the incumbent government is either very appealing or very unattractive to citizens, relatively to their expectations about potential challengers. When people are highly satisfied with the current government, they see it as

unlikely that a competitor would be better, so they retain the incumbent, which in turn makes the incumbent free to take steps that increase its chances to remain in office. When the incumbent government is relatively unappealing, citizens are afraid that it would hold onto office, and to preempt this possibility they want to remove the incumbent even if they see the current challenger as less attractive. In turn, because the incumbent knows that citizens want to remove it no matter what, it does everything possible to solidify its hold on office. Hence, in both situations the incumbent engages in *backsliding*, that is, takes every opportunity to undermine democracy in its partisan interest, 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”) generates some fixed level of satisfaction to a representative citizen (“citizen, she”).⁵ This level of satisfaction may be due to some policies of the incumbent, to manipulation of beliefs, or just to ideological congruence of values. At each time the incumbent faces a challenger whom the citizen finds more or less attractive, better or worse, than the incumbent. The probability that any future challenger would be better is given. At each time, the citizen decides whether to retain the incumbent. “Incumbent advantage” is the probability that the incumbent remains in office when the citizen wants to remove him: this 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. 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, a step which he may or may not take. The citizen does not observe immediately whether the incumbent took a step of a particular magnitude and learns about it only when the advantage has been realized, so that the incumbent can

---

⁵To the extent to which the incumbent may take actions that affect the level of satisfaction of the citizen, perhaps the best way to think is that the flow payoff represents the highest level of appeal the incumbent can achieve.

proceed by “stealth” (Varol 2015).

Stealth operates in a subtle way. Most acts by governments require legislation or executive decisions and, as such, are observable when they are announced.⁶ But whether these acts increase incumbent advantage and by how much can often 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 adopts legislation to require additional documentation at the polling place, or 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. Democracy is undermined by these measures because their accumulation makes it more difficult for citizens to remove the incumbent government when they so desire. But this effect is not visible immediately.

These assumptions are sufficient to identify the conditions under which democracy is sustainable and when it is vulnerable to two situations which induce the incumbent to engage in backsliding, to which we refer as “*populism*” and “*polarization*.”

Democracy is sustainable, free from the danger of backsliding, only if there exists a threshold level of advantage that is simultaneously satisfactory to the incumbent, so that the incumbent increases his advantage up to this level but stops short of complete domination,

---

⁶Some 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).

and acceptable to the citizen, so that the citizen supports the incumbent against worse challengers as long as the threshold level is not passed. Such a threshold level exists, so that democracy is sustainable, when incumbents are neither very attractive or very unattractive to the citizen. Democracy is ideal when the incumbent is sufficiently secure in office even without any advantage, which can be sustained only when neither the incumbent nor the potential challengers are expected to be attractive. But citizens face a tradeoff between the attractiveness of the competing candidates and the ability to replace the incumbent. To sustain democracy, citizens must be willing to tolerate a higher level of incumbent advantage for more attractive politicians.

Democracy is vulnerable in two situations. Whenever the citizen finds the incumbent highly appealing, the unique equilibrium is populism. Under populism the citizen retains the incumbent whenever she finds him more attractive than the challenger and the incumbent takes every opportunity to undermine democracy. Because the incumbent is highly attractive, the citizen believes that any challenger is unlikely to be better, so that she values relatively little her future ability to remove the incumbent. She wants the government to continue to govern even at the cost of weakening the checks on its authority, as in Acemoglu, Robinson, and Torvik (2013). Because the citizen wants to retain the attractive incumbent, this incumbent feels *free* to use all opportunities to increase his advantage. The fact that citizens do not observe the actions of the incumbent immediately, stealth, plays no role here: even if the citizen immediately observed what the incumbent is doing, she would still retain him. Populism is a situation in which the incumbent undermines democracy with the consent of citizens.

In contrast, polarization may emerge gradually from democracy. Polarization is a situation in which the citizen wants to remove the incumbent regardless of the attractiveness of the challenger and the incumbent undermines democracy as long as he remains in office. The relation between the incumbent and the citizen is polarized in the

sense that both actors behave unconditionally.⁷ Polarization emerges when the citizen finds the incumbent relatively unappealing and the incumbent has already managed to achieve considerable advantage, above the threshold which sustains democracy. Because the incumbent is unattractive, the citizen fears that he would further increase his advantage, so that she wants to remove him even when she finds the challenger less attractive. Because the citizen unconditionally opposes him, the incumbent finds it *necessary* to use all opportunities to accumulate advantage. Here stealth is important. 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. Had the citizen observed immediately that the incumbent is about to take this step, she would remove him preemptively, but she observes the step only after its effect has been realized.

In both cases, the ability of the citizen to remove the incumbent declines and the citizen becomes worse off as backsliding proceeds, yet the citizen does not react in time. Under populism this is true because the citizen values retaining the incumbent so highly that she wants him to remain in office even at the cost of being unable to remove him in the unlikely possibility that a better challenger would appear. Even though her welfare declines as the incumbent backslides, she wants to retain him even when he reaches the highest advantage he can attain. Under polarization, however, the citizen would want to unconditionally remove the incumbent if he would cross some tolerable level of advantage but she cannot react in

---

⁷We recognize that it is a bit strange to speak of “polarization” in a model with a single representative voter. We use the term because of its powerful intuition, namely, that there are situations in which some people unconditionally support and some unconditionally oppose the incumbent and this division is insensitive to actions by the incumbent. The US is a good example: according to *538* (June 11, 2020), “Trump’s approval ratings have hovered between 40 and 45 percent for most of his presidency, making his approval rating the steadiest of any president since World War II.” Poland, with its urban-rural split is another example.

time because she does not immediately observe the consequences of incumbents' actions. Hence, the answer to the puzzle differs across situations.

These results characterize the actions of the incumbent and of the representative citizen. The dynamics of backsliding, however, depend on whether the citizen gets a chance to decide whether the incumbent remains in office. Under democracy, the incumbent gains advantage sufficient to make him feel secure in office and stops at some critical level which the citizen tolerates, so that citizens preserve some ability to choose governments by democratic means. But under populism and polarization the incumbent takes every opportunity to increase his advantage, backsliding all the way to the level at which citizens lose all hope of being able to remove him by democratic means and conflicts spill outside institutional boundaries.⁸ Because under populism the citizen wants to retain the incumbent as long as the challenger is not better while under polarization she wants to remove him unconditionally, the probability that the incumbent would complete more 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 can hardly be removed by democratic means later.

3. THE MODEL

An *incumbent* ("he") and a representative *citizen* ("she") interact in infinitely many periods $t = 0, 1, 2, \dots$. In each period, the citizen decides whether to keep the incumbent in office or try to remove him, $k_t = 0, 1$, where $k_t = 1$ indicates keeping the incumbent and $k_t = 0$ trying to remove him.

The incumbent provides the citizen with a flow payoff of $x \in (0, 1)$ each period he holds office, where x measures how appealing the incumbent is to the citizen. A *challenger*, who

---

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, run 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.

can be more or less attractive than the incumbent, *better* or *worse* than the incumbent, is drawn in each period to compete for office. If in office, a better challenger brings the citizen a flow payoff of $1 > x$ while a worse challenger brings $0 < x$. In every period, the probability that the challenger is better is $\gamma \in (0, 1)$, which measures how appealing the citizen expects challengers to be.⁹

The citizen prefers more appealing politicians. As a result, she would like the incumbent to be removed when a better challenger is competing for office. The citizen, however, may not be perfectly able to remove the incumbent, either because she is not decisive or because the incumbent has means to stay in power against her will. Formally, if the citizen tries 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 his partisan *advantage*.

The goal of the incumbent is to stay in power as long as possible. A greater advantage advances this goal by making the tenure of the incumbent more robust against opposition by the citizen. The incumbent enters office under *perfect* democracy, that is, with no advantage, $p_0 = 0$. But in each period, he gets an opportunity to *undermine* democracy, a step that increases his advantage in the next period from p_t to some $q_t \geq p_t$, drawn from a distribution F_{p_t} conditional on p_t . Letting $b_t = 0, 1$ denote the action of the incumbent, where $b_t = 1$ indicates undermining democracy,

$$p_{t+1} = (1 - b_t)p_t + b_t q_t.$$

The following two assumptions with regard to F_p are maintained throughout. First,

---

⁹Allowing the challenger's quality to be continuously distributed in $[0, 1]$ would require the citizen to form an expectation about their quality, greatly complicating the algebra. But in the end the equilibriums would still depend on the comparison between the appeal of the incumbent and the expected attractiveness of challengers and at each time the decision of the citizen would still depend on a particular draw from the distribution. Hence, qualitative results would remain in the same.

the support of F_p is $[p, \bar{p}]$, where $\bar{p} < 1$. This implies that $\bar{p}$ is the largest advantage the incumbent can harvest through undermining democracy. Advantage beyond $\bar{p}$ is attainable only through committing flagrantly unconstitutional acts which expose the incumbent to being removed by coups or popular rebellions. Second, $F_{p'}$ stochastically dominates F_p if $p' > p$. This assumption means with a larger advantage in the current period, the incumbent cannot be less able to increase his advantage in the next period.

At the beginning of each period, p_t is observed both by the incumbent and the citizen. Only the incumbent, however, knows whether he undermines democracy, b_t , and the consequence if he does, q_t . The citizen is uncertain about both. She becomes aware only after the consequence, a greater advantage of the incumbent, materializes in the next period. This delay constitutes “stealth.”

The stage game in any period t the incumbent is not removed has the following sequence of moves: 1) p_t is observed by both the incumbent and the citizen; 2) q_t and a challenger are drawn; 3) the incumbent observes q_t and chooses whether to undermine democracy, $b_t = 0, 1$; 4) the citizen observes if the challenger is better or worse and chooses whether to keep the incumbent, $k_t = 0, 1$.

If the incumbent is removed, he loses power forever and a challenger assumes office. If a challenger enters office, he becomes a new incumbent and in the next period a new challenger is drawn and the citizen chooses who would hold office. For convenience, only the incumbent can undermine democracy and accumulate partisan advantage.¹⁰

The incumbent cares only about staying in power while the citizen cares only about the quality of the politician in office. Therefore, the flow payoff of the incumbent in each period

---

¹⁰If we allowed victorious challengers to start with some partisan advantage and/or able to accumulate it, democracy would be more vulnerable to the danger of backsliding but the qualitative results would be the same.

is

$$v_t = \begin{cases} 0, & \text{incumbent is out of office} \\ 1, & \text{incumbent is in office} \end{cases}$$

and that of the citizen is

$$u_t = \begin{cases} 0, & \text{worse challenger is in office} \\ x, & \text{incumbent is in office} \\ 1, & \text{better challenger is in office} \end{cases}.$$

Both actors are forward looking and share a common discount factor of $\delta \in (0, 1)$. They respectively seek to maximize the discounted average payoff of

$$(1 - \delta) \sum_{t=0}^{\infty} \delta^t v_t$$

and

$$(1 - \delta) \sum_{t=0}^{\infty} \delta^t u_t.$$

3.1. Benchmark

To provide intuition, consider first a benchmark in which the incumbent has a fixed advantage p , with no option to increase it. The incumbent is not strategic, so that what matters is only the decision of the citizen.

The preference of the citizen is clear when a challenger competing for office is more attractive to the citizen than the incumbent. In this case, the citizen always prefers to remove the incumbent, replacing him with the more appealing challenger. Removal of the incumbent fails with probability p , but if it is successful, the better challenger assumes office and brings the citizen the best flow payoff of 1 in all subsequent periods.

When the challenger is worse than the incumbent, however, the citizen faces a tradeoff. If the incumbent is replaced by a less appealing challenger, the citizen gets a lower flow payoff of 0 instead of x in the current period. But once the incumbent is removed, the challenger who enters office has no advantage at all. This implies that if a better challenger appears in any future period, the citizen would be perfectly able to place him into office and enjoy the best flow payoff of 1 in every period thereafter. Therefore, by replacing the incumbent with a worse challenger the citizen loses in the current period in terms of less appealing office holder but gains in the future periods in terms of higher ability to select a more appealing politician whenever such a challenger is available.

To find out how the citizen resolves this tradeoff, let W denote the continuation value of the citizen in any period that begins with a worse challenger holding office. Two cases are possible in such a period. With probability γ , a better new challenger appears and is chosen. In this case, the citizen gets a flow payoff of 1 in this and in all subsequent periods. Her continuation value, therefore, is 1. With probability $1 - \gamma$, a new challenger is worse and no matter whom the citizen selects, a politician less appealing than the incumbent holds office. The citizen gets a flow payoff of 0 in the current period and anticipates W in the future periods. As a result, the continuation value of the citizen in this case would be $(1 - \delta)0 + \delta W = \delta W$. Aggregating the two cases with their respective likelihoods yields

$$(1) \quad W = \gamma + (1 - \gamma)\delta W = \frac{\gamma}{1 - \delta(1 - \gamma)}.$$

Similarly, let $\bar{U}(p)$ denote the continuation value of the citizen in any period the incumbent is not yet removed, depending on his advantage p . Due to incumbent advantage, three cases are possible in such a period. First, with probability p , the incumbent stays in office, regardless of who the challenger is and whom the citizen chooses. In this case, the citizen faces exactly the same situation in the next period. She gets a flow payoff of x in the current period and expects to get $\bar{U}(p)$ in the future periods. The

continuation value of the citizen in this case is $(1 - \delta)x + \delta\bar{U}(p)$. Second, with probability $(1 - p)\gamma$, a better challenger competes against the incumbent and the citizen successfully places this challenger into office. In this case, the citizen gets the continuation value of 1. Third, with probability $(1 - p)(1 - \gamma)$, a worse challenger competes for power and the citizen decides whom to place in office. Her continuation value is $(1 - \delta)x + \delta\bar{U}(p)$ if she keeps the incumbent and $\frac{\delta\gamma}{1 - \delta(1 - \gamma)}$ if she chooses the less appealing challenger. The citizen would choose whoever is associated with a higher continuation value and, therefore, her continuation value in this case is

$$\max \left\{ (1 - \delta)x + \delta\bar{U}(p), \frac{\delta\gamma}{1 - \delta(1 - \gamma)} \right\}.$$

Aggregating all these three possible cases with their respective probabilities yields

$$(2) \quad \bar{U}(p) = p \left((1 - \delta)x + \delta\bar{U}(p) \right) + (1 - p) \left(\gamma + (1 - \gamma) \max \left\{ (1 - \delta)x + \delta\bar{U}(p), \frac{\delta\gamma}{1 - \delta(1 - \gamma)} \right\} \right).$$

Now suppose a worse challenger competes against the incumbent. By removing the incumbent, the citizen gets 0 in the current period and anticipates the continuation value of $\frac{\gamma}{1 - \delta(1 - \gamma)}$ in the subsequent periods. In total, she gets

$$(1 - \delta)0 + \delta \left(\frac{\gamma}{1 - \delta(1 - \gamma)} \right) = \frac{\delta\gamma}{1 - \delta(1 - \gamma)}.$$

In turn, by keeping the incumbent, the citizen gets x in the current period and expects $\bar{U}(p)$ in the subsequent periods. She gets

$$(1 - \delta)x + \delta\bar{U}(p),$$

where $\bar{U}(p)$ solves (2). Therefore, the citizen would try to replace the incumbent with a

less appealing challenger, in exchange for better ability to place in office a more appealing politician in the future but at the expense of her current flow payoff, if and only if

$$(1 - \delta)x + \delta\bar{U}(p) < \frac{\delta\gamma}{1 - \delta(1 - \gamma)}.$$

Otherwise, she would let the incumbent stay and wait for a better challenger to try to replace him.

Lemma 1. *Suppose the incumbent has a fixed advantage p . The citizen prefers to replace the incumbent with a worse challenger if and only if $p > \frac{x}{\frac{\delta\gamma}{1 - \delta(1 - \gamma)}}$. Moreover,*

$$\bar{U}(p) = \begin{cases} \frac{(1 - (1 - p)\gamma)(1 - \delta)x + (1 - p)\gamma}{1 - \delta(1 - (1 - p)\gamma)}, & p \leq \frac{x}{\frac{\delta\gamma}{1 - \delta(1 - \gamma)}} \\ \frac{p(1 - \delta(1 - \gamma))(1 - \delta)x + (1 - p)\gamma}{(1 - \delta p)(1 - \delta(1 - \gamma))}, & p > \frac{x}{\frac{\delta\gamma}{1 - \delta(1 - \gamma)}} \end{cases}$$

and is strictly decreasing in p .

The above lemma has two important implications. First, when the challenger is less appealing than the incumbent, the citizen still prefers the incumbent to be removed if he has a sufficiently large advantage. The threshold for p is exactly the ratio between what the citizen loses in the current period by removing the incumbent, x , and what she gains in the future periods, $\frac{\delta\gamma}{1 - \delta(1 - \gamma)}$. Second, the continuation value of the citizen under the incumbent, $\bar{U}(p)$, is strictly decreasing in his advantage p . This implies that the citizen is worse off when the incumbent has a greater advantage.

Hence, even if citizens do not value democracy because of some ideals, such as political equality or liberty, with which they associate it, caring only about the flow payoffs governments provide, they value the ability to choose a better office holder whenever in the future one would be available. This value provides citizens with an incentive to remove an incumbent who deprives them of this opportunity. When democracy is threatened by an overly large advantage of the incumbent, this incentive is strong enough to induce the

citizen to remove the incumbent as soon as possible, even if doing so means having a less appealing politician in office.

3.2. *Vulnerabilities of democracy*

The ability of the citizen to choose whoever she prefers to rule is a defining feature of democracy. When citizens are able to do so, given that they may find some challengers more attractive than the incumbent, the incumbent suffers from office insecurity. This insecurity, in turn, encourages the incumbent to seek advantage through undermining democracy, making his tenure in office more robust against the potential opposition by the citizen.

In the rest of the paper, we analyze the full model by endogenizing advantage of the incumbent. The incumbent begins with no advantage and in each period holding office he chooses whether to increase his advantage. We first characterize two vulnerabilities of democracy, corresponding to two ways in which *democratic backsliding* – persistent subversion of democracy caused by the incumbent – may occur, then analyze the consequences of backsliding, derive the conditions under which democracy is sustainable, free from the threat of backsliding, and at last discuss the cost necessary to sustain democracy.

The equilibrium concept is *Markov perfect equilibrium* (“equilibrium” hereafter) with the state variable being the advantage of the incumbent in the current period, p_t . Each strategy of the incumbent maps his current advantage, p_t , and his advantage in the next period if he undermines democracy, q_t , to a decision $b_t = \beta(q_t|p_t) = 0, 1$ on whether or not to undermine democracy. The citizen always tries to remove the incumbent in any period with a better challenger. Hence, each strategy of the citizen maps the current advantage of the incumbent, p_t , to a decision $k_t = \kappa(p_t) = 0, 1$ on whether or not to try to remove the incumbent when the challenger is worse. To ease exposition, the time subscript t is omitted in what follows whenever it causes no confusion.

Democratic backsliding occurs if the incumbent takes every opportunity to undermine democracy. Formally, it means that the incumbent pursues a strategy $\beta(q|p) = 1$ for all p, q .

Proposition 1 (vulnerabilities).

1. (*populism*) If $x \geq \frac{\bar{p}\delta\gamma}{1-\delta(1-\gamma)}$, there exists a unique equilibrium (β^*, κ^*) such that $\beta^*(q|p) = 1$ and $\kappa^*(p) = 1$ for all p, q .
2. (*polarization*) If $x < \frac{\bar{p}\delta\gamma}{1-\delta(1-\gamma)}$, for any equilibrium (β^*, κ^*) , there exists a minimal $p^* \leq \frac{x}{\frac{\delta\gamma}{1-\delta(1-\gamma)}}$ such that $\beta^*(q|p) = 1$ and $\kappa^*(p) = 0$ for all $q \geq p > p^*$.

Backsliding can happen in two ways. First, it may happen under “*populism*,” which is when the incumbent persistently undermines democracy while the citizen opposes him only if a better challenger is available. In other words, populism is a scenario in which the incumbent always tries to erode democracy while the citizen never tries to defend it. If the incumbent is lucky enough to survive several periods, either by facing worse challengers or through the advantage he has already accumulated, he becomes very difficult to remove.

Second, backsliding may also happen under “*polarization*,” when the incumbent takes every opportunity to undermine democracy while the citizen takes every chance trying to remove him from office, regardless whether the challenger is more appealing or less. In other words, polarization is a scenario in which the incumbent and the citizen unconditionally oppose each other. The incumbent always tries to weaken the ability of the citizen to remove him from office, undermining democracy, while the citizen always tries to remove the incumbent from office, defending democracy. Because the opposition of the citizen does not depend on the attractiveness of the challenger, the incumbent is less likely to survive in office under polarization than under populism. Yet if the incumbent is lucky enough to remain in power for some number of periods, it can still become difficult to remove him.

Populism emerges when the incumbent is sufficiently attractive to the citizen. The condition $x \geq \frac{\bar{p}\delta\gamma}{1-\delta(1-\gamma)}$ is equivalent to $\bar{p} \leq \frac{x}{\frac{\delta\gamma}{1-\delta(1-\gamma)}}$. Because $\bar{p}$ is the greatest advantage the incumbent can possibly achieve through undermining democracy, this condition implies

$p \leq \bar{p} \leq \frac{x}{\frac{\delta\gamma}{1-\delta(1-\gamma)}}$ for all possible p . This, by Lemma 1, means that as long as the challenger is worse, the citizen prefers to retain the incumbent regardless of how much advantage the incumbent has already accumulated and how much he will attempt to achieve in the future. When the current challenger is worse, the citizen waits for a better one to appear in the future before replacing the incumbent, despite knowing that the incumbent is backsliding and, therefore, will be more difficult to remove. In turn, given that the citizen never tries to defend democracy against backsliding, undermining it is a *costless* option for the incumbent to obtain extra office security in case a more appealing challenger appears. As a result, the incumbent takes every conceivable opportunity to undermine democracy as long as he is in power.

Polarization is triggered when the incumbent is not appealing enough to enjoy populist support yet manages to gain a sufficiently large advantage. When $x < \frac{\bar{p}\delta\gamma}{1-\delta(1-\gamma)}$ or $\bar{p} > \frac{x}{\frac{\delta\gamma}{1-\delta(1-\gamma)}}$, it is possible for the incumbent to get opportunities to increase his advantage to some $p > \frac{x}{\frac{\delta\gamma}{1-\delta(1-\gamma)}}$. According to Lemma 1, if the incumbent does so, the citizen would have a strong enough incentive to defend democracy by unconditionally trying to remove him from office even if the the incumbent does not seek to further increase his advantage above p . This incentive to defend democracy is only stronger if the incumbent may attempt to achieve even more advantage. As a result, the citizen never waits to oppose the incumbent, even if doing so requires placing a less appealing challenger in office, which in turn makes undermining democracy a *necessary* option for the incumbent. He has no choice but to rely on his advantage as the only way to stay in power. Consequently, when the incumbent has already accumulated enough advantage to trigger polarization, he exploits every possible chance to reinforce his advantage.

Although the incumbent acts similarly under populism and polarization, he does so for different reasons. The incumbent backslides under populism because of the leniency of the citizen, which makes undermining democracy costless, while under polarization he backslides because of the unconditional opposition by the citizen, which makes undermining democracy

necessary.

3.3. *Consequence of backsliding*

To see how backsliding affects welfare of the citizen, let $\underline{U}(p)$ denote the continuation value of the citizen when the incumbent is backsliding, that is, when he has the strategy β such that $\beta(q|p) = 1$ for all p, q , and currently has the advantage of p . To derive $\underline{U}(p)$, note again that there are three possible cases in each period that begins with the incumbent holding power. First, with probability p , the incumbent stays in office no matter what, so that the citizen gets a flow payoff of x in the current period. Because the incumbent is backsliding, the continuation value the citizen expects in the next period depends on the advantage the incumbent will get through undermining democracy, q . Because the citizen knows only the distribution of q , F_p , her expectation of the continuation value in the next period is

$$\int_p^{\bar{p}} \underline{U}(q) dF_p(q).$$

In this case, the citizen gets in total

$$(1 - \delta)x + \delta \int_p^{\bar{p}} \underline{U}(q) dF_p(q).$$

Second, with probability $(1 - p)\gamma$, a better challenger competes against the incumbent and would be placed into office by the citizen, who would thereafter get a flow payoff of 1 in every period. At last, with probability $(1 - p)(1 - \gamma)$, a worse challenger competes against the incumbent and the citizen decides whom to place in office. She gets $(1 - \delta)x + \delta \int_p^{\bar{p}} \underline{U}(q) dF_p(q)$ by allowing the incumbent to remain and $\frac{\delta\gamma}{1 - \delta(1 - \gamma)}$ by replacing him with the worse challenger. The continuation value of the citizen in this case is

$$\max \left\{ (1 - \delta)x + \delta \int_p^{\bar{p}} \underline{U}(q) dF_p(q), \frac{\delta\gamma}{1 - \delta(1 - \gamma)} \right\}.$$

Aggregating these three cases with their respective probabilities yields

$$(3) \quad \underline{U}(p) = p \left((1 - \delta)x + \delta \int_p^{\bar{p}} \underline{U}(q) dF_p(q) \right) + (1 - p) \left(\gamma + (1 - \gamma) \max \left\{ (1 - \delta)x + \delta \int_p^{\bar{p}} \underline{U}(q) dF_p(q), \frac{\delta\gamma}{1 - \delta(1 - \gamma)} \right\} \right).$$

Proposition 2 (welfare consequence of backsliding). *Suppose the incumbent is backsliding.*

The welfare of the citizen as measured by $\underline{U}(p)$ is strictly decreasing in p . Moreover,

$$\overline{U}(\bar{p}) \leq \underline{U}(p) \leq \overline{U}(p)$$

for all p and the equalities hold if and only if $p = \bar{p}$.

Therefore, the citizen suffers in the process of backsliding. The continuation value of the citizen, which measures her welfare, continuously declines as the advantage of the incumbent grows. This continuation value $\underline{U}(p)$, however, is bounded from below by $\overline{U}(\bar{p})$. This implies that the citizen is always *better off* in the process of backsliding than if the incumbent already achieved the largest possible advantage, $\bar{p}$. Under populism, the citizen prefers to retain the incumbent against a worse challenger even when the incumbent already attains $\bar{p}$, as $x \geq \frac{\bar{p}\delta\gamma}{1 - \delta(1 - \gamma)}$ implies $(1 - \delta)x + \delta\overline{U}(\bar{p}) \geq \frac{\delta\gamma}{1 - \delta(1 - \gamma)}$. She prefers to do so during the process of backsliding, as

$$(1 - \delta)x + \delta \int_p^{\bar{p}} \underline{U}(q) dF_p(q) \geq (1 - \delta)x + \delta\overline{U}(\bar{p}) \geq \frac{\delta\gamma}{1 - \delta(1 - \gamma)}$$

holds for all p .

$\underline{U}(p)$ is bounded from above by $\overline{U}(p)$, that is, the continuation value the citizen gets when the incumbent has a fixed advantage of p . This implies that the citizen is always *worse off* in the process of backsliding than if the incumbent stops further undermining democracy. Under polarization, the citizen is willing to unconditionally oppose the incumbent when he

stops further undermining democracy, as $p > \frac{x}{\frac{\delta\gamma}{1-\delta(1-\gamma)}}$ implies $(1-\delta)x + \delta\bar{U}(p) < \frac{\delta\gamma}{1-\delta(1-\gamma)}$. She would be more than willing to do so when during the process of backsliding, as

$$(1-\delta)x + \delta \int_p^{\bar{p}} \underline{U}(q) dF_p(q) \leq (1-\delta)x + \delta\bar{U}(p) < \frac{\delta\gamma}{1-\delta(1-\gamma)}.$$

3.4. Sustaining democracy

Democracy is vulnerable to backsliding, either by a confident popular incumbent in the form of populism or by an anxious unpopular incumbent in the form of polarization. Moreover, backsliding leads to detrimental consequences for the welfare of the citizen. A natural question, therefore, is whether, how, and under what conditions democracy is sustainable, free from the threat of backsliding.

Democracy is sustainable if backsliding never occurs. Formally,

Definition 1 (sustainability). *Democracy is sustained in an equilibrium (β^*, κ^*) if there exists a minimal threshold $p^* < \bar{p}$ such that $\beta^*(q|p) = 0$ for all $q > p^* \geq p$. Democracy is sustainable if there exists an equilibrium that sustains it.*

Democracy may be sustainable even if the incumbent takes some steps to increase his advantage. It is still sustained if there exists an equilibrium path on which, even if the incumbent takes some steps, he never attempts to achieve an advantage beyond a threshold p^* .

This definition of sustainable democracy requires only the incumbent not to backslide, while leaving the action of the citizen unspecified. The following result characterizes the action of the citizen on the path of any equilibrium that sustains democracy.

Lemma 2. *If (β^*, κ^*) is an equilibrium that sustains democracy with threshold of p^* , then $p^* \leq \frac{x}{\frac{\delta\gamma}{1-\delta(1-\gamma)}}$ and $\kappa^*(p) = 1$ for all $p \leq p^*$.*

According to Lemma 2, under sustainable democracy the citizen always supports her preferred candidate. As long as advantage of the incumbent does not exceed p^* , she keeps

the incumbent in office whenever the challenger is worse and tries to remove him whenever the challenger is better. In other words, the citizen allows the incumbent to accumulate advantage up to p^* . Replacing the incumbent with a worse challenger would be costly for the citizen. Because the citizen always sides with her preferred politician, under sustainable democracy she never incurs such a cost. When democracy is sustained, citizens are able to place in office politicians they prefer and actually execute this preference.

Lemma 2 also shows that the maximal level of advantage the incumbent can achieve under sustainable democracy: p^* has an upper bound $\frac{x}{\frac{\delta\gamma}{1-\delta(1-\gamma)}}$. According to Proposition 1, any advantage of the incumbent above this upper bound immediately triggers polarization, in which the incumbent backslides anticipating the unconditional opposition by the citizen. As a result, to prevent backsliding and, therefore, sustain democracy, the incumbent cannot be allowed to have any advantage larger than $\frac{x}{\frac{\delta\gamma}{1-\delta(1-\gamma)}}$.

The key for any equilibrium that sustains democracy is the threshold p^* . Most importantly, this threshold represents a level of advantage that is satisfactory to the incumbent, so that he is willing to stop further undermining democracy once reaching at it. But to stop, the incumbent needs to be incentivized: he must face *punishment* for increasing his advantage over p^* . As shown in the discussion about populism, if the citizen continues supporting the incumbent against worse challengers when he moves beyond p^* , he would do so to gain extra office security at no cost. The harshest possible punishment the citizen can deliver is to unconditionally oppose the incumbent regardless of who is the challenger, which she does under polarization. In turn, the threat of triggering polarization whenever the incumbent attains any $p > p^*$ provides the strongest incentive to stop the incumbent at p^* . This threat works under two conditions.

On the one hand, the threat to trigger polarization when $p > p^*$ has to be *incentive compatible* for the citizen. This means that when the incumbent indeed takes a step to increase his advantage over p^* , it is in the best interest of the citizen to oppose him unconditionally. Otherwise, the threat would not be credible. The incumbent backslides

under polarization, so that by retaining him in office, the citizen gets $(1 - \delta)x + \delta \int_p^{\bar{p}} \underline{U}(q) dF_p(q)$. She gets $\frac{\delta\gamma}{1 - \delta(1 - \gamma)}$ by replacing the incumbent with a worse challenger. Therefore, the threat to trigger polarization when $p > p^*$ is incentive compatible if and only if

$$(4) \quad (1 - \delta)x + \delta \int_p^{\bar{p}} \underline{U}(q) dF_p(q) < \frac{\delta\gamma}{1 - \delta(1 - \gamma)}$$

holds for all $p > p^*$.

On the other hand, the threat has to be *effective* for the incumbent. Effectiveness requires that, given that the citizen would trigger polarization when the incumbent gains any advantage over p^* , he finds it optimal not to take opportunities that would increase his advantage beyond this threshold. Specifically, if the incumbent currently has $p \leq p^*$, it has to be better for him to stay with p than to increase his advantage to any $q > p^*$.

To characterize effectiveness, first let $\underline{V}(p)$ denote the continuation value of the incumbent when he is unconditionally opposed by the citizen. Under the unconditional opposition by the citizen, the incumbent survives in office with probability p in any period his advantage is p . Given that the incumbent survives, he gets a flow payoff of 1 in the current period, while he anticipates a continuation value which is either $\underline{V}(q)$, if he takes an opportunity to increase his advantage to q , or $\underline{V}(p)$, if he gives up such an opportunity, that is, $\max \{\underline{V}(q), \underline{V}(p)\}$. Hence,

$$(5) \quad \underline{V}(p) = p \left(1 - \delta + \delta \int_p^{\bar{p}} \max \{\underline{V}(q), \underline{V}(p)\} dF_p(q) \right).$$

Similarly, for any $p \leq p^*$, let $V^*(p|p^*)$ denote the continuation value of the incumbent when the citizen supports him against worse challengers if he has any advantage below p^* while unconditionally opposes him for any advantage above p^* , that is, when she has the strategy κ such that $\kappa(p) = 1$ for any $p \leq p^*$ and $\kappa(p) = 0$ for any $p > p^*$. Under this strategy of the citizen, the incumbent survives with probability $1 - (1 - p)\gamma$ in any period

his advantage is $p \leq p^*$. Given that the incumbent survives, he gets a flow payoff of 1 in the current period. If the opportunity to undermine democracy is small, so that $q \leq p^*$, the incumbent anticipates in the future periods the greater of the values $V^*(q|p^*)$, which is what he gets by taking the opportunity, and $V^*(p|p^*)$, which is what he gets by giving up, that is, $\max \{V^*(q|p^*), V^*(p|p^*)\}$. If the opportunity to undermine democracy is large, so that $q > p^*$, taking such an opportunity triggers unconditional opposition to the incumbent and leaves him with the continuation value of $\underline{V}(q)$. In this case, the incumbent anticipates getting in the future the larger of $\underline{V}(q)$ and $V^*(p|p^*)$, that is, $\max \{\underline{V}(q), V^*(p|p^*)\}$. It follows that

$$(6) \quad V^*(p|p^*) = (1 - (1 - p)\gamma) \left(1 - \delta + \delta \left(\int_p^{p^*} \max \{V^*(q|p^*), V^*(p|p^*)\} dF_p(q) + \int_{p^*}^{\bar{p}} \max \{\underline{V}(q), V^*(p|p^*)\} dF_p(q) \right) \right).$$

Now suppose the incumbent has an advantage of $p \leq p^*$ and the citizen supports him against worse challengers as long as his advantage remains below p^* while credibly threatens unconditional opposition for any advantage exceeding p^* . Consider the case when the incumbent gets an opportunity to increase his advantage to some $q > p^*$. By taking this opportunity to undermine democracy, the incumbent triggers polarization under which his continuation value in the subsequent periods would be $\underline{V}(q)$. By staying at p , the incumbent maintains support by the citizen and his continuation value in the subsequent periods would be $V^*(p|p^*)$. Effectiveness requires the incumbent to resist any temptation to increase his advantage beyond p^* for any current level of advantage. Therefore, the threat of the citizen to trigger polarization when $p > p^*$ is effective for the incumbent if and only if

$$(7) \quad \underline{V}(q) \leq V^*(p|p^*)$$

holds for all $q > p^* \geq p$, where $\underline{V}(q)$ solves (5) and $V^*(p|p^*)$ solves (6).

Proposition 3 (sustainability). *For any $p^* < \bar{p}$, there exists an equilibrium that sustains democracy with the associated threshold of p^* if and only if $x \geq \frac{p^*\delta\gamma}{1-\delta(1-\gamma)}$ and*

1. *(incentive compatibility) $x \leq x^*(\gamma, p^*)$, where $x^*(\gamma, p^*) \in \left(\frac{p^*\delta\gamma}{1-\delta(1-\gamma)}, \frac{\bar{p}\delta\gamma}{1-\delta(1-\gamma)}\right)$ is strictly increasing in γ and p^* ;*
2. *(effectiveness) $\gamma \leq \gamma^*(p^*)$, where $\gamma^*(p^*) \in \left[1 - \bar{p}, \frac{1-\bar{p}}{1-p^*}\right)$ is strictly increasing in p^* and $\gamma^*(0) = 1 - \bar{p}$.*

The above proposition implies that in addition to the condition $p^* \leq \frac{x}{\frac{\delta\gamma}{1-\delta(1-\gamma)}}$ identified in Lemma 2 or restated as $x \geq \frac{p^*\delta\gamma}{1-\delta(1-\gamma)}$, two more conditions are necessary for any p^* to be the threshold in an equilibrium that sustains democracy. These two new conditions respectively correspond to the incentive compatibility and effectiveness for the citizen to threaten triggering polarization when $p > p^*$. Specifically, for this threat to be incentive compatible for the citizen, the incumbent must be sufficiently unappealing to her, so that $x \leq x^*(\gamma, p^*)$. The intuition is straightforward. The more the citizen likes the incumbent, the more she suffers by replacing him with a less appealing candidate, and, therefore, the more reluctant she would be to trigger polarization by unconditionally opposing the incumbent.

For the threat of triggering polarization when $p > p^*$ to be effective for the incumbent, it has to be sufficiently unlikely that a challenger more appealing than the incumbent would compete against him, so that $\gamma \leq \gamma^*(p)$. This is because, even though the citizen supports him against worse challengers, the incumbent still faces office insecurity due to the possibility of competing against a better challenger. If this happens, the citizen cannot commit to retain the incumbent and the incumbent, in turn, relies only on his advantage to survive. Maintaining support of the citizen helps the incumbent to stay in power only in the case of a worse challenger. The more likely it is that the incumbent would compete against a better challenger, the less useful it is for him to maintain support of the citizen by resisting from undermining democracy, and, therefore, the more tempted he would be to take a big step to increase his advantage beyond p^* whenever such an opportunity arises.

The three conditions in Proposition 3, $x \geq \frac{p^*\delta\gamma}{1-\delta(1-\gamma)}$, $x \leq x^*(\gamma, p^*)$, and $\gamma \leq \gamma^*(p^*)$ together are sufficient to ensure that democracy can be sustained in an equilibrium with the associated threshold of p^* . Under these conditions, polarization happens only off the equilibrium path, as a threat to restrain the incumbent. Because $\gamma \leq \gamma^*(p^*)$, the threat is effective and, therefore, the incumbent does not dare crossing the line.

In general, democracy is sustainable as long as a $p^* < \bar{p}$ that satisfies all the three conditions in Proposition 3 exists.

Corollary 1. *Democracy is sustainable if and only if*

1. (no populism) $x < \frac{\bar{p}\delta\gamma}{1-\delta(1-\gamma)}$;
2. (no polarization) $\gamma \leq 1 - \bar{p}$ or $1 - \bar{p} < \gamma < \bar{\gamma}$ and $x \geq \underline{x}(\gamma)$, where $\bar{\gamma} \in (1 - \bar{p}, 1)$ and $\underline{x}(\gamma) \in \left(0, \frac{\bar{p}\delta\gamma}{1-\delta(1-\gamma)}\right)$ is strictly increasing in γ .

Democracy is sustainable under two conditions that respectively prevent backsliding from happening either in the form of populism or polarization. Backsliding occurs through populism if the citizen too much likes the incumbent or too much dislikes potential challengers. In this case, the citizen can never credibly threaten the incumbent to punish him by unconditional removal. Backsliding can occur in the form of polarization if the citizen has an overly high expectation about potential challengers or if she finds the incumbent to be overly unappealing. In this case, the incumbent faces severe office insecurity and is willing to take a big step to increase his advantage when he has the opportunity to do so, even if the citizen, once she realizes it, would unconditionally oppose him.

3.5. Cost of democracy

Democracy is sustainable if it is free from the danger of backsliding. As mentioned with regard to Definition 1, this notion of sustainable democracy is not incompatible with incumbent taking some steps to increase his advantage. Sustainability is guaranteed as

long as there is a threshold level of advantage $p^* < \bar{p}$ that is simultaneously acceptable to the citizen and satisfactory for the incumbent. The threshold at which the incumbent would stop undermining democracy measures how democratic institutionally a particular sustainable democracy is. A lower p^* means better ability of the citizen to select her preferred office holder which, in turn, corresponds to a more democratic regime. In the extreme, if there exists an equilibrium with the threshold of $p^* = 0$, *perfect* democracy, in which the incumbent begins with no advantage at all and never attempts to get any, is sustainable.

Because $x^*(\gamma, p^*)$ is strictly increasing in p^* , the citizen is able to have a more appealing politician in office if she allows the incumbent to accumulate a higher level of advantage and, therefore, to become more difficult to remove. Intuitively, the citizen is more reluctant to remove a more appealing incumbent and, therefore, is more tolerant of his advantage, which implies that the threshold of tolerable advantage has to be higher. Likewise, because $\gamma^*(p^*)$ is strictly increasing in p^* , when the incumbent is allowed to accumulate a larger advantage, the citizen has a better chance to see a challenger more attractive than the incumbent competing for office. When the incumbent has a larger probability to compete against a better challenger, he faces more severe office insecurity, which would make him more tempted to gain overly large advantage at the risk of triggering polarization. This implies that to deter the incumbent from doing so, the citizen allows him to accumulate advantage up to a higher threshold to compensate his office insecurity.

Therefore, in all sustainable democracies, the citizen faces a *tradeoff* between the attractiveness of politicians and the level of democracy, specifically, the ability to place any politician she wants into office. If the citizen wants to have a better pool of politicians to select from, she has to bear with weaker ability to place whoever she prefers into office. If, instead, the citizen wants more control over who occupies office, she has to bear with a pool of not so attractive candidates to choose from.

Less than perfect politicians are a necessary evil to sustain an institutionally perfect

democracy. To sustain it, the incumbent has to be unattractive, so that the citizen is able to impose a credible threat on the incumbent that if he obtains any advantage, however small, he would have to cope with her unconditional opposition. The potential challengers have to be quite unappealing in expectation, making it not worth for the incumbent to take the risk of being unconditionally opposed by the citizen to gain any advantage. An institutionally imperfect, perhaps really existing, democracy is a system, in which the incumbent is allowed to achieve a moderately high level of advantage as a reward for being more appealing.

4. CONCLUSION: COULD IT HAPPEN HERE?

One lesson we are learning from the instances of backsliding 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. 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 ... have proven ineffective against this new threat.”

The contribution of our analysis is another lesson, namely that defending democracy imposes a difficult challenge on individual citizens. To act against a government that is destroying democracy, people must weigh their satisfaction with the incumbent government against the effect of backsliding on their future ability to replace it by a better one. Even if individuals have consistent time preferences (Akerlof 1991) and even if they fully anticipate

that the government will not stop short of seeking complete domination, they may still prefer the incumbent over potential opponents, giving up future possibilities to remove it. In turn, when the effects of anti-democratic measures are not visible immediately, it may be too late to remove the incumbent once these effects materialize. 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 effect is to protect the incumbent from being defeated in the future.

The fact is that backsliding governments have managed to remain in power. 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 and maintained the presidency in July 2020. 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 2102 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.

The optimism that citizens would effectively threaten to punish 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 steps, it may be too late to prevent it from doing whatever it wants.

REFERENCES

- Acemoglu, Daron, James Robinson, and Ragnar Torvik. 2013. “Why Do Voters Dismantle Checks and Balances?” *Review of Economic Studies* 80: 845–875.
- Akerlof, George A. 1991. “Procrastination and Obedience.” *American Economic Review* 81: 1-19.
- Bracher, Karl Dietrich. 1966. “The Technique of the National Socialist Seizure of Power.” in Fritz Stern (ed.), *The Path to Dictatorship 1918-1933: Ten Essays by German Scholars*. Garden City, NY: Anchor Books. Pages 113-132.
- Dragu, Tiberiu, Xiaochen Fan, and James Kuklinski. 2014. “Designing Checks and Balances.” *Quarterly Journal of Political Science* 9: 1–42.
- Dunn, Susan. 2004. *Jefferson’s Second Revolution: the Election Crisis of 1800 and the Triumph of Republicanism*. Boston: Houghton Mifflin.
- Fearon, James. 2011. “Self-enforcing Democracy.” *Quarterly Journal of Economics* 126: 1661-1708.
- Ginsburg, Tom and Aziz Huq. 2018a. “How to Lose a Constitutional Democracy.” *UCLA Law Review* 65(1): 78-169.
- Ginsburg, Tom and Aziz Huq. 2018b. *How to Save a Constitutional Democracy*. Chicago: University of Chicago Press.
- Graham, Matthew, and Milan Svolik. 2018. “Democracy in America? Partisanship, Polarization, and the Robustness of Support for Democracy in the United States.” Working Paper.
- Grillo, Edoardo and Carlo Prato. 2019. “Opportunistic Authoritarians, Reference-dependent Preferences and Democratic Backsliding.” Working Paper.
- Helmke, Gretchen, Mary Kroeger, and Jack Paine. 2019. “Exploiting Asymmetries: A Theory of Democratic Constitutional Hardball.” Working Paper.
- Howell, William, Kenneth Shepsle, and Stephane Wolton. 2019. “Executive Absolutism: A

Model.” Working Paper.

Howell, William and Stephane Wolton. 2018. “The Politician’s Province.” *Quarterly Journal of Political Science* 13(2): 119-146.

Landau, David. 2013. “Abusive Constitutionalism ” *University of California at Davis Law Review* 47: 189–260.

Madison, James. 1982 [1788]. *The Federalist Papers by Alexander Hamilton, James Madison, and John Jay*. Edited by Gary Wills. New York: Bantam Books.

Montesquieu, Baron de. 1995 [1748]. *De l’esprit des lois*. Paris: Gallimard.

Nalepa, Monika, Georg Vanberg, and Caterina Chiopris. 2019. “Authoritarian Backsliding.” Working Paper.

Palmer, Robert Roswell. 1959. *The Age of the Democratic Revolution. Vol. I. The Challenge*. Princeton: Princeton University Press.

Persson, Torsten, Gerard Roland, and Guido Tabetini. 1996. “Separation of Powers and Accountability: Towards a Formal Approach to Comparative Politics.” Discussion Paper No. 1475. London: Centre for Economic Policy Research.

Sadurski, Wojciech. 2018. “How Democracy Dies (in Poland): A Case Study of Anti-Constitutional Populist Backsliding.” Legal Studies Research Paper No. 18/01. Sydney Law School.

Svolik, Milan. 2015. “Which Democracies Will Last? Coups, Incumbent Takeovers, and the Dynamic of Democratic Consolidation.” *British Journal of Political Science* 45(4): 715-38.

Svolik, Milan. 2020. “When Polarization Trumps Civic Virtue: Partisan Conflict and the Subversion of Democracy by Incumbents.” *Quarterly Journal of Political Science* 15(1): 3-31.

Varol, Ozan. 2015. “Stealth Authoritarianism.” *Iowa Law Review* 100: 1673-1742.

Weingast, Barry. 1997. “Political Foundations of Democracy and the Rule of Law.” *American Political Science Review* 91: 245-63.

Weingast, Barry. 2015. “Capitalism, Democracy, and Countermajoritarian Institutions.”

Supreme Court Economic Review 23: 255-277.

Appendix to “Democracy and its vulnerabilities: Dynamics of democratic backsliding”

A.1. LEMMA 1

Proof of Lemma 1. Recall that by definition,

$$\bar{U}(p) = p \left((1 - \delta)x + \delta \bar{U}(p) \right) + (1 - p) \left(\gamma + (1 - \gamma) \max \left\{ (1 - \delta)x + \delta \bar{U}(p), \frac{\delta \gamma}{1 - \delta(1 - \gamma)} \right\} \right).$$

First, suppose $p > \frac{x}{\frac{\delta \gamma}{1 - \delta(1 - \gamma)}}$. Assume $(1 - \delta)x + \delta \bar{U}(p) \geq \frac{\delta \gamma}{1 - \delta(1 - \gamma)}$. Then,

$$\bar{U}(p) = \frac{(1 - (1 - p)\gamma)(1 - \delta)x + (1 - p)\gamma}{1 - \delta(1 - (1 - p)\gamma)},$$

so that

$$\begin{aligned} (1 - \delta)x + \delta \bar{U}(p) &= \frac{(1 - \delta)x + \delta(1 - p)\gamma}{1 - \delta(1 - (1 - p)\gamma)} \\ &= \frac{\delta \gamma}{1 - \delta(1 - \gamma)} + \frac{1 - \delta}{1 - \delta(1 - (1 - p)\gamma)} \left(x - \frac{p\delta \gamma}{1 - \delta(1 - \gamma)} \right) \\ &< \frac{\delta \gamma}{1 - \delta(1 - \gamma)}, \end{aligned}$$

a contradiction. Hence, it must be true that $(1 - \delta)x + \delta \bar{U}(p) < \frac{\delta \gamma}{1 - \delta(1 - \gamma)}$, so that

$$\bar{U}(p) = \frac{p(1 - \delta)x + (1 - p)\frac{\gamma}{1 - \delta(1 - \gamma)}}{1 - \delta p} = \frac{p(1 - \delta(1 - \gamma))(1 - \delta)x + (1 - p)\gamma}{(1 - \delta p)(1 - \delta(1 - \gamma))}$$

Second, suppose $p \leq \frac{x}{\frac{\delta\gamma}{1-\delta(1-\gamma)}}$. Assume $(1-\delta)x + \delta\bar{U}(p) < \frac{\delta\gamma}{1-\delta(1-\gamma)}$. Then,

$$\bar{U}(p) = \frac{p(1-\delta)x + (1-p)\frac{\delta\gamma}{1-\delta(1-\gamma)}}{1-\delta p},$$

so that

$$\begin{aligned} (1-\delta)x + \delta\bar{U}(p) &= \frac{(1-\delta)x + (1-p)\frac{\delta\gamma}{1-\delta(1-\gamma)}}{1-\delta p} \\ &= \frac{\delta\gamma}{1-\delta\gamma} + \frac{1-\delta}{1-\delta p} \left(x - \frac{p\delta\gamma}{1-\delta(1-\gamma)} \right) \\ &\geq \frac{\delta\gamma}{1-\delta(1-\gamma)}, \end{aligned}$$

a contradiction. Hence, it must be true that $(1-\delta)x + \delta\bar{U}(p) \geq \frac{\delta\gamma}{1-\delta(1-\gamma)}$, so that

$$\bar{U}(p) = \frac{(1-(1-p)\gamma)(1-\delta)x + (1-p)\gamma}{1-\delta(1-(1-p)\gamma)}.$$

At last, given the expression of $\bar{U}(p)$, it is clear that it is strictly decreasing in p . ■

A.2. PROPOSITION 1 AND 2

Let $U_\beta(p)$ denote the continuation value of the citizen when the incumbent has the strategy of β and the current advantage of p . Then,

$$\begin{aligned} (A.1) \quad U_\beta(p) &= p((1-\delta)x + \delta E_\beta U_\beta(p)) \\ &\quad + (1-p) \left(\gamma + (1-\gamma) \max \left\{ (1-\delta)x + \delta E_\beta U_\beta(p), \frac{\delta\gamma}{1-\delta(1-\gamma)} \right\} \right), \end{aligned}$$

where

$$E_\beta U_\beta(p) = \int_p^{\bar{p}} ((1-\beta(q|p)) U_\beta(p) + \beta(q|p) U_\beta(q)) dF_p(q).$$

Specifically, with probability p , the incumbent stays in office independent to the choice of the citizen, in which case the citizen gets the expected continuation value of $(1 - \delta)x + \delta E_\beta U_\beta(p)$. With probability $1 - p$ the choice of the citizen matters. Given that her choice matters, with probability γ , a better challenger appears and would be placed in to office, bringing the citizen the continuation value of 1; and with probability $1 - \gamma$, the challenger is worse and the citizen would choose the better option between staying with the incumbent, who would bring the expected continuation value of $(1 - \delta)x + \delta E_\beta U_\beta(p)$, and placing the challenger into office, who would bring the expected continuation value of $\frac{\delta\gamma}{1 - \delta(1 - \gamma)}$. The citizen prefers to replace the incumbent with a worse challenger if and only if

$$(A.2) \quad (1 - \delta)x + \delta E_\beta U_\beta(p) \geq \frac{\delta\gamma}{1 - \delta(1 - \gamma)}.$$

The “max” in (A.1) reflects the optimal choice of the citizen.

Let $V_\kappa(p)$ denote the continuation value of the incumbent when the citizen has the strategy of κ and the current advantage of the incumbent is p . Then,

$$(A.3) \quad V_\kappa(p) = (p + (1 - p)(1 - \gamma)\kappa(p)) \left(1 - \delta + \delta \int_p^{\bar{p}} \max \{V_\kappa(p), V_\kappa(q)\} dF_p(q) \right).$$

Specifically, the incumbent stays in office with probability $p + (1 - p)(1 - \gamma)\kappa(p)$. Given that the incumbent stays, he gets 1 in the current period and expects the better one between staying with p , which brings the continuation value of $V_\kappa(p)$ in the subsequent periods, and increasing his advantage to q , which brings the continuation value of $V_\kappa(q)$ in the subsequent periods. The incumbent prefers to increase his advantage by undermining democracy if and only if

$$(A.4) \quad V_\kappa(q) \geq V_\kappa(p).$$

The “max” in (A.3) reflects the optimal choice of the incumbent. In expectation, the

continuation value given that the incumbent stays in office is

$$1 - \delta + \delta \int_p^{\bar{p}} \max \{V_\kappa(p), V_\kappa(q)\} dF_p(q).$$

Lemma A.1. *For any β and p ,*

$$\bar{U}(\bar{p}) \leq U_\beta(p) \leq \bar{U}(p).$$

Proof. Let β and p be given. For any $p' \geq p$,

$$\begin{aligned} E_\beta U_\beta(p') &= \int_p^{\bar{p}} ((1 - \beta(q|p))p + \beta(q|p)q) dF_p(q) \\ &\leq \max_{q \in [p', \bar{p}]} U_\beta(q) \\ &\leq \max_{q \in [p, \bar{p}]} U_\beta(q), \end{aligned}$$

so that

$$\begin{aligned} U_\beta(p') &\leq p' \left((1 - \delta)x + \delta \max_{q \in [p, \bar{p}]} U_\beta(q) \right) \\ &\quad + (1 - p') \left(\gamma + (1 - \gamma) \max \left\{ (1 - \delta)x + \delta \max_{q \in [p, \bar{p}]} U_\beta(q), \frac{\delta\gamma}{1 - \delta(1 - \gamma)} \right\} \right) \\ &\leq p \left((1 - \delta)x + \delta \max_{q \in [p, \bar{p}]} U_\beta(q) \right) \\ &\quad + (1 - p) \left(\gamma + (1 - \gamma) \max \left\{ (1 - \delta)x + \delta \max_{q \in [p, \bar{p}]} U_\beta(q), \frac{\delta\gamma}{1 - \delta(1 - \gamma)} \right\} \right), \end{aligned}$$

where the second inequality is due to $p' \geq p$ and

$$\gamma + (1 - \gamma) \max \left\{ (1 - \delta)x + \delta \max_{q \in [p, \bar{p}]} U_\beta(q), \frac{\delta\gamma}{1 - \delta(1 - \gamma)} \right\} \geq (1 - \delta)x + \delta \max_{q \in [p, \bar{p}]} U_\beta(q).$$

In turn, because

$$U_\beta(p') \leq p \left((1 - \delta)x + \delta \max_{q \in [p, \bar{p}]} U_\beta(q) \right) + (1 - p) \left(\gamma + (1 - \gamma) \max \left\{ (1 - \delta)x + \delta \max_{q \in [p, \bar{p}]} U_\beta(q), \frac{\delta\gamma}{1 - \delta(1 - \gamma)} \right\} \right),$$

holds for all $p' \geq p$,

$$\max_{q \in [p, \bar{p}]} U_\beta(q) \leq p \left((1 - \delta)x + \delta \max_{q \in [p, \bar{p}]} U_\beta(q) \right) + (1 - p) \left(\gamma + (1 - \gamma) \max \left\{ (1 - \delta)x + \delta \max_{q \in [p, \bar{p}]} U_\beta(q), \frac{\delta\gamma}{1 - \delta(1 - \gamma)} \right\} \right).$$

By the definition of $\bar{U}(p)$, this implies that $U_\beta(p) \leq \max_{q \in [p, \bar{p}]} U_\beta(q) \leq \bar{U}(p)$. Similarly, for any $p' \geq p$,

$$\begin{aligned} E_\beta U_\beta(p') &= \int_p^{\bar{p}} ((1 - \beta(q|p))p + \beta(q|p)q) dF_p(q) \\ &\geq \min_{q \in [p', \bar{p}]} U_\beta(q) \\ &\geq \min_{q \in [p, \bar{p}]} U_\beta(q), \end{aligned}$$

so that

$$\begin{aligned} U_\beta(p') &\geq p' \left((1 - \delta)x + \delta \min_{q \in [p, \bar{p}]} U_\beta(q) \right) \\ &\quad + (1 - p') \left(\gamma + (1 - \gamma) \max \left\{ (1 - \delta)x + \delta \min_{q \in [p, \bar{p}]} U_\beta(q), \frac{\delta\gamma}{1 - \delta(1 - \gamma)} \right\} \right) \\ &\geq \bar{p} \left((1 - \delta)x + \delta \min_{q \in [p, \bar{p}]} U_\beta(q) \right) \\ &\quad + (1 - \bar{p}) \left(\gamma + (1 - \gamma) \max \left\{ (1 - \delta)x + \delta \min_{q \in [p, \bar{p}]} U_\beta(q), \frac{\delta\gamma}{1 - \delta(1 - \gamma)} \right\} \right). \end{aligned}$$

Because the last inequality holds for all $p' \geq p$,

$$\begin{aligned} \min_{q \in [p, \bar{p}]} U_\beta(q) &\geq \bar{p} \left((1 - \delta)x + \delta \min_{q \in [p, \bar{p}]} U_\beta(q) \right) \\ &\quad + (1 - \bar{p}) \left(\gamma + (1 - \gamma) \max \left\{ (1 - \delta)x + \delta \min_{q \in [p, \bar{p}]} U_\beta(q), \frac{\delta\gamma}{1 - \delta(1 - \gamma)} \right\} \right). \end{aligned}$$

By the definition of $\bar{U}(\bar{p})$, this implies that $U_\beta(p) \geq \min_{q \in [p, \bar{p}]} U_\beta(q) \geq \bar{U}(\bar{p})$. ■

Proof of Proposition 1. First, assume $x \geq \frac{\bar{p}\delta\gamma}{1 - \delta(1 - \gamma)}$, so that $\bar{p} \leq \frac{x}{\frac{\delta\gamma}{1 - \delta(1 - \gamma)}}$. Then, according to Lemma A.1, for any β and p , $U_\beta(p) \geq \bar{U}(\bar{p})$, so that

$$(1 - \delta)x + \delta U_\beta(p) \geq (1 - \delta)x + \delta \bar{U}(\bar{p}) \geq \frac{\delta\gamma}{1 - \delta(1 - \gamma)}.$$

Hence, for any β and p ,

$$(1 - \delta)x + \delta E_\beta U_\beta(p) \geq \frac{\delta\gamma}{1 - \delta(1 - \gamma)}.$$

This implies that always keeping the incumbent in office against a worse challenger, $\kappa^*(p) = 1$ for all p , is a dominant strategy of the citizen. Given that the citizen has this strategy,

$$V_{\kappa^*}(p) = (p + (1 - p)(1 - \gamma)) \left(1 - \delta + \delta \int_p^{\bar{p}} \max \{V_{\kappa^*}(p), V_{\kappa^*}(q)\} dF_p(q) \right).$$

The operator on the right hand side maps every increasing function f to a function

$$(p + (1 - p)(1 - \gamma)) \left(1 - \delta + \delta \int_p^{\bar{p}} \max \{f(p), f(q)\} dF_p(q) \right),$$

that is strictly increasing in p . Hence, $V_{\kappa^*}(p)$ is strictly increasing in p . In turn, $V_{\kappa^*}(q) \geq V_{\kappa^*}(p)$ for all p, q . Therefore, it is optimal for the incumbent to choose $\beta^*(q|p) = 1$ for all p, q .

Second, assume $x < \frac{\bar{p}\delta\gamma}{1-\delta(1-\gamma)}$, so that $\bar{p} > \frac{x}{\frac{\delta\gamma}{1-\delta(1-\gamma)}}$. Consider any equilibrium (β^*, κ^*) and any $p^* \geq \frac{x}{\frac{\delta\gamma}{1-\delta(1-\gamma)}}$. Then, because $\bar{U}$ is strictly decreasing, for all $p > p^*$, $U_{\beta^*}(p) \leq \bar{U}(p) < \bar{U}(p^*)$ and

$$(1-\delta)x + \delta U_{\beta^*}(p) < (1-\delta)x + \delta \bar{U}(p^*) < \frac{\delta\gamma}{1-\delta(1-\gamma)}.$$

Hence, for all $p > p^*$, because

$$(1-\delta)x + \delta E_{\beta^*} U_{\beta^*}(p) < \frac{\delta\gamma}{1-\delta(1-\gamma)},$$

it must be true that $\kappa^*(p) = 0$. Given that the citizen has $\kappa^*(p) = 0$ for all $p > p^*$,

$$V_{\kappa^*}(p) = p \left(1 - \delta + \delta \int_p^{\bar{p}} \max \{V_{\kappa^*}(p), V_{\kappa^*}(q)\} dF_p(q) \right)$$

for all $p > p^*$. The right hand side is strictly increasing in $p > p^*$ if V_{κ^*} is increasing in $p > p^*$, which implies that V_{κ^*} is strictly increasing in $p > p^*$. In turn, $V_{\kappa^*}(q) \geq V_{\kappa^*}(p)$ for all $q \geq p > p^*$. Therefore, it must be true that $\beta^*(q|p) = 1$ for all $q \geq p > p^*$. ■

Proof of Proposition 2. First, $\underline{U}(p) = U_{\beta}(p)$, where $\beta(q|p) = 1$ for all p, q . This together with Lemma A.1 imply that $\bar{U}(\bar{p}) \leq \underline{U}(p) \leq \bar{U}(p)$ for all p and the equalities hold if and only if $p = \bar{p}$.

To prove that $\underline{U}(p)$ is strictly decreasing in p , remember that

$$\begin{aligned} \underline{U}(p) = & p \left((1-\delta)x + \delta \int_p^{\bar{p}} \underline{U}(q) dF_p(q) \right) \\ & + (1-p) \left(\gamma + (1-\gamma) \max \left\{ (1-\delta)x + \delta \int_p^{\bar{p}} \underline{U}(q) dF_p(q), \frac{\delta\gamma}{1-\delta(1-\gamma)} \right\} \right). \end{aligned}$$

The operator on the right hand side maps any decreasing function f to a function

$$p \left((1 - \delta)x + \delta \int_p^{\bar{p}} f(q) dF_p(q) \right) + (1 - p) \left(\gamma + (1 - \gamma) \max \left\{ (1 - \delta)x + \delta \int_p^{\bar{p}} f(q) dF_p(q), \frac{\delta\gamma}{1 - \delta(1 - \gamma)} \right\} \right)$$

that is strictly decreasing in p . This is because

$$\int_p^{\bar{p}} f(q) dF_p(q)$$

is decreasing in p given that $f(p)$ is decreasing in p and for any p ,

$$\begin{aligned} & \gamma + (1 - \gamma) \max \left\{ (1 - \delta)x + \delta \int_p^{\bar{p}} f(q) dF_p(q), \frac{\delta\gamma}{1 - \delta(1 - \gamma)} \right\} \\ & \geq (1 - \delta)x + \delta \int_p^{\bar{p}} f(q) dF_p(q). \end{aligned}$$

Hence, $\underline{U}(p)$ is strictly decreasing in p . ■

A.3. LEMMA 2

Proof of Lemma 2. Let (β^*, κ^*) be an equilibrium that sustains democracy with the threshold of p^* . First, Lemma 1 and Lemma A.1 together imply that for any $p > \frac{x}{\frac{\delta\gamma}{1 - \delta(1 - \gamma)}}$, $U_{\beta^*}(p) \leq \bar{U}(p)$ and

$$(1 - \delta)x + \delta U_{\beta^*}(p) \leq (1 - \delta)x + \delta \bar{U}(p) < \frac{\delta\gamma}{1 - \delta(1 - \gamma)},$$

so that

$$(1 - \delta)x + \delta E_{\beta^*} U_{\beta^*}(p) < \frac{\delta\gamma}{1 - \delta(1 - \gamma)}.$$

Hence, $\kappa^*(p) = 0$ for all $p > \frac{x}{\frac{\delta\gamma}{1-\delta(1-\gamma)}}$. But given κ^* , for any $p > \frac{x}{\frac{\delta\gamma}{1-\delta(1-\gamma)}}$,

$$V_{\kappa^*}(p) = p \left(1 - \delta + \delta \int_p^{\bar{p}} \max \{V_{\kappa^*}(p), V_{\kappa^*}(q)\} dF_p(q) \right)$$

is strictly increasing in p . This in turn implies that $\beta^*(q|p) = 1$ for all $q \geq p > \frac{x}{\frac{\delta\gamma}{1-\delta(1-\gamma)}}$.

Therefore, given the definition of p^* , it must be true that $p^* \leq \frac{x}{\frac{\delta\gamma}{1-\delta(1-\gamma)}}$.

Second, because $p^* \leq \frac{x}{\frac{\delta\gamma}{1-\delta(1-\gamma)}}$, Lemma A.1 implies that

$$(1 - \delta)x + \delta \bar{U}(p^*) \geq \frac{\delta\gamma}{1 - \delta(1 - \gamma)}.$$

Moreover, because $\beta^*(q|p) = 0$ for all $q > p^* \geq p$,

$$E_{\beta^*} U_{\beta^*}(p) = \int_p^{p^*} ((1 - \beta^*(q|p)) U_{\beta^*}(p) + \beta^*(q|p) U_{\beta^*}(q)) dF_p(q) + (1 - F_p(p^*)) U_{\beta^*}(p)$$

Remember in the definition of U_{β^*} , (A.1), the operator on the right hand side maps any function f that is bounded from below by $\bar{U}(p^*)$ on $[0, p^*]$ to

$$p((1 - \delta)x + \delta E_{\beta^*} f(p)) + (1 - p) \left(\gamma + (1 - \gamma) \max \left\{ (1 - \delta)x + \delta E_{\beta^*} f(p), \frac{\delta\gamma}{1 - \delta(1 - \gamma)} \right\} \right),$$

where

$$E_{\beta^*} f(p) = \int_p^{p^*} ((1 - \beta^*(q|p)) f(p) + \beta^*(q|p) f(q)) dF_p(q) + (1 - F_p(p^*)) f(p) \geq \bar{U}(p^*)$$

for any $p \leq p^*$. It follows that for any $p \leq p^*$,

$$\begin{aligned}
& p((1-\delta)x + \delta E_{\beta^*} f(p)) + (1-p) \left(\gamma + (1-\gamma) \max \left\{ (1-\delta)x + \delta E_{\beta^*} f(p), \frac{\delta\gamma}{1-\delta(1-\gamma)} \right\} \right) \\
& \geq p((1-\delta)x + \delta \bar{U}(p^*)) + (1-p) \left(\gamma + (1-\gamma) \max \left\{ (1-\delta)x + \delta \bar{U}(p^*), \frac{\delta\gamma}{1-\delta(1-\gamma)} \right\} \right) \\
& \geq p^*((1-\delta)x + \delta \bar{U}(p^*)) + (1-p^*) \left(\gamma + (1-\gamma) \max \left\{ (1-\delta)x + \delta \bar{U}(p^*), \frac{\delta\gamma}{1-\delta(1-\gamma)} \right\} \right) \\
& = \bar{U}(p^*),
\end{aligned}$$

where the second inequality is due to $p \leq p^*$ and

$$\gamma + (1-\gamma) \max \left\{ (1-\delta)x + \delta \bar{U}(p^*), \frac{\delta\gamma}{1-\delta(1-\gamma)} \right\} \geq (1-\delta)x + \delta \bar{U}(p^*).$$

Hence, the operator on the right hand side in the definition of U_{β^*} maps any function that is bounded from below by $\bar{U}(p^*)$ on $[0, p^*]$ to a function that is bounded from below by $\bar{U}(p^*)$ on $[0, p^*]$. Therefore, it must be true that $U_{\beta^*}(p) \geq \bar{U}(p^*)$ for all $p \leq p^*$. In turn, for any $p \leq p^*$, $E_{\beta^*} U_{\beta^*}(p) \geq \bar{U}(p^*)$ and

$$(1-\delta)x + \delta E_{\beta^*} U_{\beta^*}(p) \geq (1-\delta)x + \delta \bar{U}(p^*) \geq \frac{\delta\gamma}{1-\delta(1-\gamma)}.$$

This implies that for any $p \leq p^*$, the citizen prefers to keep the incumbent in office against a worse challenger and thus $\kappa^*(p) = 1$. ■

A.4. PROPOSITION 3 AND COROLLARY 1

Lemma A.2 (incentive compatibility). *For any $p^* < \bar{p}$, there exists a unique $x^*(\gamma, p^*) \in \left(\frac{p^* \delta \gamma}{1 - \delta(1 - \gamma)}, \frac{\bar{p} \delta \gamma}{1 - \delta(1 - \gamma)}\right)$ such that*

$$(1 - \delta)x + \delta \int_p^{\bar{p}} \underline{U}(q) dF_p(q) < \frac{\delta \gamma}{1 - \delta(1 - \gamma)}$$

holds for all $p > p^*$ if and only if $x \leq x^*(\gamma, p^*)$. Moreover, $x^*(\gamma, p^*)$ is strictly increasing in γ and p^* .

Proof. First, by definition,

$$\begin{aligned} \underline{U}(p) = & p \left((1 - \delta)x + \delta \int_p^{\bar{p}} \underline{U}(q) dF_p(q) \right) \\ & + (1 - p) \left(\gamma + (1 - \gamma) \max \left\{ (1 - \delta)x + \delta \int_p^{\bar{p}} \underline{U}(q) dF_p(q), \frac{\delta \gamma}{1 - \delta(1 - \gamma)} \right\} \right). \end{aligned}$$

The operator on the right hand side maps any function f that is increasing in x to a function that is strictly increasing in x . This implies that $\underline{U}(p)$ must be strictly increasing in x for all p , which in turn implies that

$$(1 - \delta)x + \delta \int_{p^*}^{\bar{p}} \underline{U}(q) dF_{p^*}(q)$$

is strictly increasing in x . Moreover, at $x = \frac{p^* \delta \gamma}{1 - \delta(1 - \gamma)}$ or when $p^* = \frac{x}{\frac{\delta \gamma}{1 - \delta(1 - \gamma)}}$, Lemma 1 and Lemma A.1 together imply that

$$(1 - \delta)x + \delta \int_{p^*}^{\bar{p}} \underline{U}(q) dF_{p^*}(q) < (1 - \delta)x + \delta \bar{U}(p^*) = \frac{\delta \gamma}{1 - \delta(1 - \gamma)}.$$

At $x = \frac{\bar{p} \delta \gamma}{1 - \delta(1 - \gamma)}$ or when $\bar{p} = \frac{x}{\frac{\delta \gamma}{1 - \delta(1 - \gamma)}}$,

$$(1 - \delta)x + \delta \int_{p^*}^{\bar{p}} \underline{U}(q) dF_{p^*}(q) > (1 - \delta)x + \delta \bar{U}(\bar{p}) = \frac{\delta \gamma}{1 - \delta(1 - \gamma)}.$$

Therefore, there exists a unique $x^*(\gamma, p^*) \in \left(\frac{p^*\delta\gamma}{1-\delta(1-\gamma)}, \frac{\bar{p}\delta\gamma}{1-\delta(1-\gamma)}\right)$ such that

$$(1-\delta)x + \delta \int_{p^*}^{\bar{p}} \underline{U}(q) dF_{p^*}(q) = \frac{\delta\gamma}{1-\delta(1-\gamma)}$$

at $x = x^*(\gamma, p^*)$, so that

$$(1-\delta)x + \delta \int_p^{\bar{p}} \underline{U}(q) dF_p(q) < \frac{\delta\gamma}{1-\delta(1-\gamma)}$$

holds for all $p > p^*$ if and only if $x \leq x^*(\gamma, p^*)$.

Second, according to Proposition 2, $\underline{U}(p)$ is strictly decreasing in p , so that

$$(1-\delta)x + \delta \int_{p^*}^{\bar{p}} \underline{U}(q) dF_{p^*}(q)$$

is strictly decreasing in p^* . Hence, by the definition of $x^*(\gamma, p^*)$, it must be strictly increasing in p^* .

Third, differentiate both sides in the definition of $\underline{U}(p)$ yields either

$$\frac{\partial}{\partial\gamma}\underline{U}(p) = (p + (1-p)(1-\gamma)) \delta \int_p^{\bar{p}} \frac{\partial}{\partial\gamma}\underline{U}(q) dF_p(q)$$

or

$$\frac{\partial}{\partial\gamma}\underline{U}(p) = p\delta \int_p^{\bar{p}} \frac{\partial}{\partial\gamma}\underline{U}(q) dF_p(q) + (1-p) \frac{\partial}{\partial\gamma} \left(\frac{\gamma}{1-\delta(1-\gamma)} \right).$$

For any function f that is bounded from above by $\frac{\partial}{\partial\gamma} \left(\frac{\gamma}{1-\delta(1-\gamma)} \right) > 0$,

$$\begin{aligned} (p + (1-p)(1-\gamma)) \delta \int_p^{\bar{p}} f(q) dF_p(q) &\leq (p + (1-p)(1-\gamma)) \delta \frac{\partial}{\partial\gamma} \left(\frac{\gamma}{1-\delta(1-\gamma)} \right) \\ &< \frac{\partial}{\partial\gamma} \left(\frac{\gamma}{1-\delta(1-\gamma)} \right) \end{aligned}$$

and

$$\begin{aligned} p\delta \int_p^{\bar{p}} f(q)dF_p(q) + (1-p)\frac{\partial}{\partial\gamma} \left(\frac{\gamma}{1-\delta(1-\gamma)} \right) &\leq (1-p+p\delta)\frac{\partial}{\partial\gamma} \left(\frac{\gamma}{1-\delta(1-\gamma)} \right) \\ &< \frac{\partial}{\partial\gamma} \left(\frac{\gamma}{1-\delta(1-\gamma)} \right). \end{aligned}$$

Hence, it must be true that $\frac{\partial}{\partial\gamma} \underline{U}(p) < \frac{\partial}{\partial\gamma} \left(\frac{\gamma}{1-\delta(1-\gamma)} \right)$. This inequality holds for all p , so that

$$\begin{aligned} \frac{\partial}{\partial\gamma} \left((1-\delta)x + \delta \int_{p^*}^{\bar{p}} \underline{U}(q)dF_{p^*}(q) \right) &= \delta \int_{p^*}^{\bar{p}} \frac{\partial}{\partial\gamma} \underline{U}(q)dF_{p^*}(q) \\ &< \frac{\partial}{\partial\gamma} \left(\frac{\gamma}{1-\delta(1-\gamma)} \right). \end{aligned}$$

In turn, by the definition of $x^*(\gamma, p^*)$, it must be strictly increasing in γ . ■

Lemma A.3 (effectiveness). *For any $p^* < \bar{p}$, there exists a unique $\gamma^*(p^*) \in \left[1 - \bar{p}, \frac{1-\bar{p}}{1-p^*}\right]$ such that*

$$\underline{V}(q) \leq V^*(p|p^*)$$

holds for all $q > p^* \geq p$ if and only if $\gamma \leq \gamma^*(p^*)$. Moreover, $\gamma^*(p^*)$ is strictly increasing in p^* and $\gamma^*(0) = 1 - \bar{p}$.

Proof. First,

$$\underline{V}(p) = p \left(1 - \delta + \delta \int_p^{\bar{p}} \max \{ \underline{V}(p), \underline{V}(q) \} dF_p(q) \right)$$

and the operator on the right hand side maps every increasing function to a strictly increasing function. Hence, $\underline{V}(p)$ is strictly increasing in p , so that

$$\underline{V}(p) = p \left(1 - \delta + \delta \int_p^{\bar{p}} \underline{V}(q)dF_p(q) \right)$$

and

$$\underline{V}(p) \leq \underline{V}(\bar{p}) = \bar{p}(1 - \delta + \delta \underline{V}(\bar{p})) = \frac{(1 - \delta)\bar{p}}{1 - \delta\bar{p}}$$

holds for all p . Similarly,

$$V^*(p|p^*) = (p + (1 - p)(1 - \gamma)) \left(1 - \delta + \delta \left(\int_p^{p^*} \max \{V^*(p|p^*), V^*(q|p^*)\} dF_p(q) + \int_{p^*}^{\bar{p}} \max \{V^*(p|p^*), \underline{V}(q)\} dF_p(q) \right) \right)$$

and the operator on the right hand side maps every function that is increasing on $[0, p^*]$ to a strictly increasing function on $[0, p^*]$. Hence, $V^*(p|p^*)$ is strictly increasing in p , so that

$$V^*(p|p^*) = (p + (1 - p)(1 - \gamma)) \left(1 - \delta + \delta \left(\int_p^{p^*} V^*(q|p^*) dF_p(q) + \int_{p^*}^{\bar{p}} \max \{V^*(p|p^*), \underline{V}(q)\} dF_p(q) \right) \right)$$

and

$$\begin{aligned} V^*(p|p^*) &\geq V^*(0|p^*) \\ &= (1 - \gamma) \left(1 - \delta + \delta \left(\int_0^{p^*} V^*(q|p^*) dF_p(q) + \int_{p^*}^{\bar{p}} \max \{V^*(p|p^*), \underline{V}(q)\} dF_p(q) \right) \right). \end{aligned}$$

holds for all $p \leq p^*$. Therefore,

$$\underline{V}(q) \leq V^*(p|p^*)$$

holds for all $q > p^* \geq p$ if and only if

$$V^*(0|p^*) \geq \frac{(1 - \delta)\bar{p}}{1 - \delta\bar{p}}.$$

Second, the operator on the right hand side in the definition of $V^*(p|p^*)$ maps every function

that is decreasing in γ to a function that is strictly decreasing in γ , which implies that $V^*(p|p^*)$ is strictly decreasing in γ . Moreover, at $\gamma = 1 - \bar{p}$, because $V^*(p|p^*)$ is strictly increasing in p ,

$$V^*(0|p^*) \geq (1 - \gamma) (1 - \delta + \delta V^*(0|p^*)) ,$$

so that

$$V^*(0|p^*) \geq \frac{(1 - \delta)(1 - \gamma)}{1 - \delta(1 - \gamma)} = \frac{(1 - \delta)\bar{p}}{1 - \delta\bar{p}} .$$

Note that by the definition of $V^*(p|p^*)$, all the equalities hold if and only if $p^* = 0$. Consider $\gamma = \frac{1 - \bar{p}}{1 - p^*}$, assume that $V^*(0|p^*) \geq \frac{(1 - \delta)\bar{p}}{1 - \delta\bar{p}}$, so that $\underline{V}(q) \leq V^*(p|p^*)$ for all $q > p^* \geq p$. Then, because $V^*(p|p^*)$ is strictly increasing in p ,

$$\begin{aligned} V^*(0|p^*) &\leq V^*(p^*|p^*) \\ &= (p^* + (1 - p^*)(1 - \gamma)) (1 - \delta + \delta V^*(p^*|p^*)) \\ &= \frac{(1 - \delta) (1 - (1 - p^*)\gamma)}{1 - \delta (1 - (1 - p^*)\gamma)} \\ &= \frac{(1 - \delta)\bar{p}}{1 - \delta\bar{p}} . \end{aligned}$$

Note that because $V^*(p|p^*)$ is strictly increasing in p , all the equalities hold if and only if $p^* = 0$. Therefore, there exists a unique $\gamma^*(p^*) \in \left[1 - \bar{p}, \frac{1 - \bar{p}}{1 - p^*}\right]$ such that

$$V^*(0|p^*) = \frac{(1 - \delta)\bar{p}}{1 - \delta\bar{p}}$$

at $\gamma = \gamma^*(p^*)$ and $\gamma^*(0) = 1 - \bar{p}$. Because $V^*(0|p^*)$ is strictly decreasing in γ , $V^*(0|p^*) \geq \frac{(1 - \delta)\bar{p}}{1 - \delta\bar{p}}$, so that $\underline{V}(q) \leq V^*(p|p^*)$ holds for all $q > p^* \geq p$ if and only if $\gamma \leq \gamma^*(p^*)$.

At last, because $\underline{V}$ is strictly increasing,

$$\begin{aligned}
\underline{V}(p^*) &= p^* \left(1 - \delta + \delta \int_p^{\bar{p}} \underline{V}(q) dF_{p^*}(q) \right) \\
&\leq p^* \left(1 - \delta + \delta \int_{p^*}^{\bar{p}} \max \{V^*(p^*|p^*), \underline{V}(q)\} dF_{p^*}(q) \right) \\
&< (p^* + (1 - p^*)(1 - \gamma)) \left(1 - \delta + \delta \int_{p^*}^{\bar{p}} \max \{V^*(p^*|p^*), \underline{V}(q)\} dF_{p^*}(q) \right) \\
&= V^*(p^*|p^*).
\end{aligned}$$

Hence, $V(p^*|p^*) > \max \{\underline{V}(p^*), V(p|p^*)\}$ for any $p < p^*$. This in turn implies that the operator on the right hand side in the definition of $V^*(p|p^*)$ maps every function that is increasing in p^* to a function that is strictly increasing in p^* . Therefore, $V^*(p|p^*)$ must be strictly increasing in p^* , in particular, $V^*(0|p^*)$. By the definition of $\gamma^*(p^*)$, this implies that $\gamma^*(p^*)$ is strictly increasing in p^* . ■

Proof of Proposition 3. It follows Lemma 2, Lemma A.2, and Lemma A.3. ■

Proof of Corollary 1. As shown in Proposition 1, if $x \geq \frac{\bar{p}\delta\gamma}{1-\delta(1-\gamma)}$, the unique equilibrium is populism under which the incumbent backslides. Hence, a necessary condition for sustainable democracy is $x < \frac{\bar{p}\delta\gamma}{1-\delta(1-\gamma)}$ or $p^\dagger := \frac{x}{\frac{\delta\gamma}{1-\delta(1-\gamma)}} < \bar{p}$. The rest of the proof presumes this condition.

First, democracy is sustainable if and only if there exists an equilibrium that sustains democracy with the threshold of $p^\dagger$. Sufficiency is obvious. To prove necessity, suppose $p^\dagger$ cannot be the threshold in any equilibrium that sustains democracy. Then, because $p^\dagger = \frac{x}{\frac{\delta\gamma}{1-\delta(1-\gamma)}}$, either $x > x^*(\gamma, p^\dagger)$, so that $p^\dagger$ is not incentive compatible, or $\gamma > \gamma^*(p^\dagger)$, so that $p^\dagger$ is not effective. Suppose $x > x^*(\gamma, p^\dagger)$, then because $x^*(\gamma, p^*)$ is strictly increasing in p^* , $x > x^*(\gamma, p^\dagger) \geq x^*(\gamma, p^*)$ for all $p^* \leq p^\dagger = \frac{x}{\frac{\delta\gamma}{1-\delta(1-\gamma)}}$. As a result, there exists no $p^* \leq \frac{x}{\frac{\delta\gamma}{1-\delta(1-\gamma)}}$ that is incentive compatible, so that democracy cannot be sustainable. Now

suppose $\gamma > \gamma^*(p^\dagger)$, then because $\gamma^*(p^*)$ is strictly increasing in p^* , $\gamma > \gamma^*(p^\dagger) \geq \gamma^*(p^*)$ for all $p^* \leq p^\dagger = \frac{x}{\frac{\delta\gamma}{1-\delta(1-\gamma)}}$. As a result, there exists no $p^* \leq \frac{x}{\frac{\delta\gamma}{1-\delta(1-\gamma)}}$ that is effective, so that democracy cannot be sustainable. This proves necessity.

Second, note that $p^\dagger = \frac{x}{\frac{\delta\gamma}{1-\delta(1-\gamma)}}$ is always incentive compatible, as

$$x = \frac{p^\dagger \delta \gamma}{1 - \delta(1 - \gamma)} < x^*(\gamma, p^\dagger).$$

Hence, $p^\dagger$ can be a threshold in an equilibrium that sustains democracy if and only if it is effective, that is, if and only if

$$(A.5) \quad \gamma \leq \gamma^*(p^\dagger) = \gamma^* \left(\frac{x}{\frac{\delta\gamma}{1-\delta(1-\gamma)}} \right).$$

Because $\gamma^*(p^*) \geq \gamma^*(0) = 1 - \bar{p}$ for all p^* , the above condition holds for all $x \in (0, 1)$ if $\gamma \leq 1 - \bar{p}$. Because $\gamma^*(p^*) < \gamma^*(\bar{p}) =: \bar{\gamma} \in (1 - \bar{p}, 1)$ for all $p^* < \bar{p}$, the above condition fails for all $x \in (0, 1)$ if $\gamma \geq \bar{\gamma}$. Now suppose $1 - \bar{p} < \gamma < \bar{\gamma}$. Then, because $\gamma^*(p^*)$ is strictly increasing in p^* , $\gamma^* \left(\frac{x}{\frac{\delta\gamma}{1-\delta(1-\gamma)}} \right)$ is strictly increasing in x . Moreover,

$$\begin{aligned} \gamma^* \left(\frac{0}{\frac{\delta\gamma}{1-\delta(1-\gamma)}} \right) &= \gamma^*(0) = 1 - \bar{p} < \gamma \\ \gamma^* \left(\frac{\frac{\bar{p}\delta\gamma}{1-\delta(1-\gamma)}}{\frac{\delta\gamma}{1-\delta(1-\gamma)}} \right) &= \gamma^*(\bar{p}) = \bar{\gamma} > \gamma. \end{aligned}$$

Hence, there exists a unique $\underline{x}(\gamma) \in \left(0, \frac{\bar{p}\delta\gamma}{1-\delta(1-\gamma)}\right)$ such that $\gamma^* \left(\frac{\underline{x}(\gamma)}{\frac{\delta\gamma}{1-\delta(1-\gamma)}} \right) = \gamma$ and $\gamma \leq \gamma^* \left(\frac{x}{\frac{\delta\gamma}{1-\delta(1-\gamma)}} \right)$ if and only if $x \geq \underline{x}(\gamma)$. Therefore, $p^\dagger$ is effective and thus democracy is sustainable if and only if $\gamma \leq 1 - \bar{p}$ or $1 - \bar{p} < \gamma < \bar{\gamma}$ and $x \geq \underline{x}(\gamma)$.

At last, because $\gamma^*(p^*)$ is strictly increasing in p^* , $\gamma^* \left(\frac{x}{\frac{\delta\gamma}{1-\delta(1-\gamma)}} \right)$ is strictly decreasing in γ . By the definition of $\underline{x}(\gamma)$, this implies that $\underline{x}(\gamma)$ is strictly increasing in γ . ■