

Liberty, Security, and Accountability: The Rise and Fall of Illiberal Democracies

Gabriele Gratton and Barton E. Lee*

July 18, 2022

Abstract

We study a model of the rise and fall of illiberal democracies. Voters value both liberty and economic security. In times of crisis, voters may prefer to elect an illiberal government that, by violating constitutional constraints, offers greater economic security but less liberty. However, violating these constraints allows the government to manipulate information, in turn reducing electoral accountability. We show how elements of liberal constitutions induce voters to elect illiberal governments that remain in power for inefficiently long—including forever. We derive insights into what makes constitutions stable against the rise of illiberal governments. We extend the model to allow for illiberal governments to overcome checks and balances and become autocracies. We show that stronger checks and balances are a double-edged sword: they slow down autocratization but may make it more likely. We discuss the empirical relevance of our theoretical framework and its connection to real world examples.

*Gratton: School of Economics, UNSW Business School, UNSW Sydney, g.gratton@unsw.edu.au, Lee: Magdalen College, University of Oxford, barton.e.lee@gmail.com. We thank the editor and three anonymous referees for their constructive advice. We are also grateful to Dan Bernhardt, Steve Callander, Juan Carlos Carbajal, Pauline Grosjean, Richard Holden, Anton Kolotilin, Federico Masera, Massimo Morelli, Carlo Prato, Adam Szeidl, Galina Zudenkova, and participants at various seminars and conferences. Chris Burnitt and Jonathan Nathan provided useful research assistance. Gratton is the recipient of an Australian Research Council Future Fellowship (FT210100176).

1 Introduction

Around the world, voters are increasingly electing governments that operate beyond the liberal constraints of their constitutions (Foa and Mounk, 2017; Svobik, 2019). The governments of Jarosław Kaczyński in Poland, Viktor Orbán in Hungary, and Recep Tayyip Erdoğan in Turkey have infringed (albeit to different extents) fundamental liberal rights: freedom of speech, civil rights, and the rule of law (Huq and Ginsburg, 2018; Luo and Przeworski, 2019). These governments also routinely abused their constitutional powers to manipulate information. They engaged the state bureaucracy to produce favorable or misleading information, or harassed bureaucrats who produced unfavorable reports. They co-opted media shareholders into the government, or threatened media outlets with lawsuits or the withholding of state funding and advertising.¹ Yet, they all remained formally accountable to the electorate and seldom, if at all, resorted to coercive means of repression.² These *illiberal democracies* rose to power with significant popular support and some remained sufficiently popular to ensure their own reelection.³

Much attention has been given to the rise of illiberal democracies (recent examples include Berman, 2019; Grillo and Prato, 2021; Luo and Przeworski, 2019; Chiopris et al., 2021; Rosenbluth and Shapiro, 2018; Svobik, 2020). However, illiberal democracies also frequently fall. Why do some illiberal democracies prove more resilient than others? When will voters who support illiberal leaders turn against them? When will voters' illiberal tendencies ultimately lead to autocratization?

In this paper we offer a theory of the rise and fall of illiberal democracies. We use this theory to draw insights into what makes some liberal constitutions stable against the lure of illiberalism. Our theory builds off two fundamental premises. Our first premise is that

¹Guriev and Treisman (2019, p. 118) review such soft censorship strategies in Lee Kuan Yew's Singapore, Orbán's Hungary, Putin's Russia, and Fujimori's Peru; Szeidl and Szucs (2021) focus on the Hungarian case. Using a measure of government transparency on economic indicators, Hollyer, Rosendorff and Vreeland (2014) document a sudden drop in transparency at the onset of illiberal democracies.

²These governments did not overthrow their democratic constitutions and violated liberal rights in ways that ostensibly observed the letter (if not the spirit) of their respective constitutions (Howell, Shepsle and Wolton, 2019; Huq and Ginsburg, 2018). However, commentators and scholars alike often worry that these illiberal democracies may be a step towards autocracy, perhaps following a similar pattern to that of fascist regimes in the 1930s. We return to this possible dynamic later in the paper.

³The term "illiberal democracy" is commonly attributed to Zakaria (1997). Guriev and Treisman (2020a) document long-lasting support for autocracies and illiberal democracies. In 2015, Kaczyński's party, Law and Justice, was the first party to win a majority of seats in the Polish parliament since the fall of communism and their vote share increased in the subsequent 2019 election. In 2010, Orbán's Fidesz won an outright majority of the vote share and continued to remain the first party in Hungary for two further general elections with a vote share of 49.27% in 2018. Singapore may represent an example of illiberal democracy that maintains popular support for several decades; as reported in Guriev and Treisman (2020a), "in Singapore in 2009, 98 percent of respondents told Gallup they thought the government of Prime Minister Lee Hsien Loong was doing a good job."

voters in democracies value both liberty and security.⁴ They value the fundamental tenets of the liberal state: individual freedoms and rights, the rule of law, and the protection of underrepresented minorities.⁵ All else equal, they prefer a more just and fair society. But voters also value security: being protected from negative shocks to their economic interests, physical security, and welfare.

Our second premise is that liberal constitutions place formal constraints on executives with the aim of guaranteeing liberty and fostering accountability. For example, the U.S. constitution features anti-majoritarian elements, such as the composition of the Senate or the Supreme Court, that constrain the actions of the administration. Around the world, democratic governments are forbidden from enforcing laws that courts have ruled infringe constitutionally protected rights. They are also often legally forbidden from using their powers to manipulate voters' information for electoral advantage. For example, governments may not recruit state employees for electoral campaigns,⁶ nor influence bureaucrats or media to censor unfavorable information. However, executives have *de facto* substantial leeway beyond these formal legal constraints—often limited only by their accountability to the voters.⁷

With these two premises in mind, we think of an illiberal democracy as a democratically elected government that, by operating beyond the liberal limits of the constitution, offers to voters something they may prefer to liberty: security. For example, in response to increasing terrorist attacks, illiberal governments can promise to voters a more incisive reaction, including extrajudicial treatment of suspects, or the suppression of dangerous media, movements, or even bad apples within the state. While limiting liberty, such policies may increase the economic security of the pivotal voter. In fact, illiberal leaders boast about their (presumed) results on economic security, quality public services, and protection against violent and economic threats from within and outside the country. In their rhetoric, their disrespect for formal liberties is exactly what enables them to deliver on

⁴The adage, commonly attributed to Benjamin Franklin, that “those who give up liberty for security deserve neither” is popularly used to remark that a true supporter of liberal democracy should not be willing to tradeoff liberty for greater individual benefits—though Franklin’s original meaning may have been misinterpreted (Wittes, 2011).

⁵Graham and Svobik (2020) and Svobik (2020) show experimentally that voters value democratic norms that protect minorities but are willing to trade them off for more favorable economic policies—in the language we use in this paper, for more security. See also Alsan, Braghieri, Eichmeyer, Kim, Stantcheva and Yang (2020).

⁶To this effect, in the U.S., the Hatch Act of 1939 precludes most federal employees from taking part in political campaigns.

⁷Posner and Vermeule (2011) argue that elections, parties, and political culture “constrain the executive far more than do legal rules created by Congress or the Courts; and although politics hardly guarantees that the executive will always act in the public interest, politics at least limits the scope for executive abuses.”

these promises.⁸ Therefore, when voters receive information that suggests a greater need for security, they may prefer an illiberal government to one that is committed to operate within the limits of the constitution. Yet, voters should also reason that governments that silence media and bad apples will inevitably also (ab)use their powers to manipulate information in their own favor. Such information manipulation is likely to reduce effective electoral accountability, perhaps explaining the enduring electoral success of some illiberal democracies.⁹

Analyzing this dynamic tradeoff requires a model with three key ingredients: (i) a long-lived principal (voter) chooses in each period whether to hire an agent (the illiberal government); (ii) the principal's per-period utility depends on an unknown period-specific binary state and whether she hired the agent or not; (iii) once an agent is hired, and until the principal chooses to dismiss him, the agent enjoys a rent and controls the flow of information the principal can observe about the state in future periods. While our focus is on the rise and fall of illiberal democracies, we believe this model may find useful applications in other contexts. For example, the choice to hire a professional manager exposes a business owner to a similar tradeoff. Hiring a professional manager is costly, but the manager can more effectively navigate internal reorganizations or expand the business to new markets. However, once hired, the manager may not disclose to the owner all the relevant information on whether there is any need for reorganizations or new market opportunity. The owner can choose at any time to retake control of the business, reduce management costs, and see all the available information with her own eyes. However, doing so carries the risk of exposing the business to shocks that would have been better addressed by a professional manager.¹⁰

In both our professional manager example and in illiberal democracies, the agent has an incentive to persuade the principal that the next period's state is such that it is better to retain the agent. However, if the information filtered by the agent is too opaque, the principal may prefer to dismiss the agent and see all the available information with her

⁸In a speech in 2014, citing Singapore, Russia, and Turkey as examples, Viktor Orbán famously remarked that “there is a race underway to find the method of community organization, the state, which is most capable of making a nation and a community internationally competitive”, adding that “systems that are not [...] liberal democracies and perhaps not even democracies, can nevertheless make their nations successful” and that “liberal democracy will probably be incapable of maintaining their global competitiveness in the upcoming decades” (Orbán, 2014).

⁹Hollyer, Rosendorff and Vreeland (2015) show empirically that autocratic regimes that provide less information to their citizens are more stable. Edmond (2013) argues that information manipulation can be an effective tool in protecting regimes against uprisings. Guriev and Treisman (2020b) offer a theory of how autocratic regimes can co-opt the media and elites into manipulating information observed by citizens.

¹⁰Our tradeoff is more likely to play an important role in contexts in which the principal cannot enforce more complex contracts that reward the agent conditional on performance.

own eyes. This dynamic tradeoff is what drives equilibrium behavior in our model. We show that the quality of the available information affects both the principal's value of dismissing the agent and the agent's ability to manipulate information in his favor.

In our benchmark model, an infinitely lived and forward looking voter chooses in each period whether to elect a liberal or an illiberal government. Liberal governments are committed to abide by the constraints of the constitution. Therefore, they offer the voter greater liberty, less security, and let the voter observe information from media and whistleblowers as much as guaranteed by the constitution. Illiberal governments do not abide by the constitution, so they offer less liberty, more security, and manipulate the information observed by the voter in their own favor. Depending on the features of the constitution and the observed information, in each period the voter may prefer either a liberal or an illiberal government, optimally trading off liberty and security.¹¹ However, the voter knows that electing an illiberal government comes with an endogenously determined effective *accountability cost*, because the voter knows that she may be induced to retain the illiberal government even if, had she access to unadulterated media, she would prefer not to. We model the illiberal government's problem as one of optimal Bayesian persuasion (Kamenica and Gentzkow, 2011). The voter knows the choice of information manipulation—which we call a *censorship policy*—but can only indirectly infer the information she would have otherwise observed had she chosen a liberal government.

In equilibrium, the parameters of the model determine which of four different regimes arises. At two extremes, the polity is either a *stable liberal democracy* with continuously liberal governments or a *stable illiberal democracy* with continuously illiberal governments. In both these cases, the choice of government is optimal for the voter given the uncensored information provided to her by institutions and media. However, the polity may also be an *inefficient stable illiberal democracy*. In this regime, the voter eventually resorts to electing an illiberal government that remains in power forever by censoring all relevant information for the voter. Absent censorship, the voter would actually find it optimal to cycle between liberal and illiberal governments. We show that an inefficient stable illiberal democracy arises only if the expected cost of security under full censorship is greater than the sum of the value of liberty and the discounted maximal accountability cost of illiberalism. When this condition fails, illiberal governments cannot ensure their own reelection by simply censoring all information. They need to provide the voter with

¹¹The voter would prefer a “benevolent” illiberal government that provides both liberty and security by operating beyond the constraints of the constitution only when it is useful for the voter and without manipulating information. However, illiberal governments also attain private benefits as a result of violating the constitution and manipulate information because they care about being reelected. Thus, absent a commitment mechanism, such benevolence is not sustainable.

sufficient information to persuade her to reelect them. In turn, this decreases the accountability cost of illiberalism, increasing the appeal of an illiberal government. This gives rise to a *cycling liberal-illiberal democracy* in which liberal and illiberal governments alternate, but illiberal governments remain in power for longer than would be optimal for the voter.

In our model, the voter elects illiberal governments during crises that increase her *expected cost of security*: the likelihood of a shock that can only be neutralized via illiberal means. However, because the illiberal government manipulates information, the voter may be persuaded to reelect the government even if there is no further crisis. We show how the probability that an illiberal government rises from a liberal democracy depends on the amount of constraints on the executive, the voter's value of liberty, and the transparency of institutions and media. An immediate implication is that excessive constraints on the executive, which render liberal governments ineffective, are more likely to induce voters to elect an illiberal government. This result resonates with the literature suggesting that constitutional reforms that strengthen the executive's power are needed to safeguard liberal democracies (e.g., [Howell and Moe, 2020](#)). We add to this that technological changes that increase transparency may exacerbate this perverse effect of excessive constraints on the government: when information leaks become more frequent, liberal stability demands that constitutions place fewer constraints on the executive.

We also characterize the optimal amount of censorship of illiberal governments and show how it depends on the voter's values of liberty and security, the constraints on the executive, and the transparency of liberal institutions and media. The latter affects censorship through a *persuasion* effect and an *accountability* effect that go in opposite directions and we discuss how the stringency of the constraints on liberal governments determines which of the two effects dominates. The same factors also affect the likelihood that an illiberal democracy falls, returning the country to liberalism, when the illiberal government's censorship fails to conceal to the voter that there is in fact no crisis. Again, the effect of more transparent liberal institutions depends on other features of the constitution, and can even be non-monotonic.

In Section 7 we study the welfare implications of the design of stable liberal constitutions and show that while more transparency is always favored by the median voter, it may harm every voter with a higher relative value of liberty—therefore potentially reducing aggregate welfare.

In Section 8 we extend our model to allow us to study the effects of checks and balances that prevent illiberal democracies from consolidating their power and becoming autocracies. We observe that illiberal leaders may exacerbate the voters' need for their economic security, for example by taking more aggressive diplomatic stances or refusing

negotiations with separatist groups.¹² We therefore allow the illiberal government to manipulate the voter’s value of security once in power. Similarly, illiberal governments may sometimes be able to build the ability to overthrow the democratic constitution and dispense of electoral accountability altogether. In both cases, the government becomes a *de facto* autocracy that remains in power forever. Both processes are more likely to be set in motion when *checks and balances* preventing total executive takeover are weaker. We show that in the long run this gives rise to only two possible regimes: stable liberal democracies and autocracies. When the probability that the government can build an autocracy is greater, non-stable illiberal democracies become autocracies faster. However, because the voter anticipates this risk, more constitutions induce stable liberal democracies. Our results suggest that stronger checks and balances aimed at preventing total executive takeover, while slowing down this process, may actually induce voters to elect an illiberal government with greater probability—ultimately choosing the path to autocracy.¹³ Finally, we discuss some realistic long-term implications of censorship on the probability of the rise and fall of illiberal democracies.

As we discuss in Section 9, our theoretical framework helps make sense of individual and aggregate phenomena in the real world. We show that within consolidated democracies, individuals’ fears of physical and economic threats correlate with less support for liberalism: more worried voters believe that the protection of civil rights is less important and that it would be better to have a strong executive who does not need to bother with parliament. We argue that the data are consistent with the view we propose in this paper that such individual preferences for illiberalism respond to the amount of constraints on the executive and may have consequences for the action of governments. We also discuss how our results, both theoretical and empirical, relate to the empirical literature on the causes of the rise of illiberal governments in Western countries in the past decade. We conclude (Section 10) by reviewing the key policy implications of our model.

2 Related literature.

Scholars have long wondered why and when illiberal leaders and governments are elected. The traditional view of democracy, from Montesquieu to Fearon, maintains that attempts to consolidate power from within a democratic constitution would be blocked by the

¹²For example, [Felshtinsky and Pribylovsky \(2008\)](#) and [Satter \(2003\)](#) argue that Vladimir Putin’s conduct of the Chechen wars, and likely the Moscow bombings attributed to Chechen separatists, were orchestrated to generate popular support for a strong leader with knowledge and experience of the secret services.

¹³[Aghion, Alesina and Trebbi \(2004\)](#) and [Gratton and Morelli \(2021\)](#) highlight different mechanisms through which excessive checks and balances may reduce voter welfare.

voters.¹⁴ [Besley and Persson \(2019\)](#) argue that there is a natural complementarity between democratic values and institutions, creating persistence. This view has recently been challenged by the rise of populist, illiberal candidates even in consolidated democracies ([Graham and Svulik, 2020](#); [Luo and Przeworski, 2019](#)). [Foa and Mounk \(2016\)](#) and [Mounk \(2018\)](#) argue that protracted dissatisfaction with the economic performance of liberal democracies has led voters to prefer more effective illiberal governments. Yet voters also value individual liberties, civil rights, and the rule of law ([Graham and Svulik, 2020](#); [Svulik, 2020](#)). As a result, prevalent theories on the rise of illiberal democracies rely on factors that may make liberty less salient than economic performance, resolving the voters' tradeoff in favor of more effective, yet less liberal governments. Such factors include political polarization ([Chiopris, Nalepa and Vanberg, 2021](#); [Svulik, 2020](#)) and the weakening of traditional mass parties ([Berman and Snegovaya, 2019](#); [Rosenbluth and Shapiro, 2018](#)).¹⁵ Our model captures this fundamental tradeoff between liberty and security, but by studying the choice of a forward-looking, infinitely-lived voter,¹⁶ we are also able to capture another fundamental and dynamic aspect of illiberal democracies: the loss of electoral accountability.¹⁷ This allows us to understand how various constitutional and technological features interact in determining the rise, and—in contrast with previous literature—the fall of illiberal democracies. In particular, our model allows us to understand the role played by the transparency of the state bureaucracy and the media, informing a growing debate on the role of government transparency in the rise of illiberal politicians ([Mounk, 2018](#); [Mudde and Rovira Kaltwasser, 2017](#); [Sgueo, 2018](#)).

We model the ruling of the illiberal government as a process of information manipulation *à la* [Kamenica and Gentzkow \(2011\)](#). We share this feature with [Gehlbach and Sonin \(2014\)](#), [Luo and Rozenas \(2019\)](#) and [Kolotilin, Mylovanov and Zapechelnyuk \(2019\)](#) (see also [Li, Raiha and Shotts, 2020](#); [Shadmehr and Bernhardt, 2015](#)). In a series of papers, [Guriev and Treisman \(2019, 2020b\)](#) document that many 21st century autocracies and illiberal democracies (what they call “informational autocracies”) maintain power through

¹⁴In *The Spirit of Law (1949 [1750])*, Montesquieu famously argues that if a power were to succeed in violating the constitution, there would be a revolution reestablishing the rule of law. [Fearon \(2011\)](#) shows that democracy may be self-enforcing because the convention of an electoral calendar provides a public signal for the voters to coordinate against a ruling faction that attempts to not hold or rig an election (see also [Bidner, François and Trebbi, 2014](#); [Little, Tucker and LaGatta, 2015](#)).

¹⁵[Mukand and Rodrik \(2020\)](#) study a model of the rise of illiberal democracies from autocracies. They show that such transitions are more likely when there exist non-income cleavages within a country or when the income cleavage is less deep.

¹⁶One feature that distinguishes our model from those of [Bernhardt, Krasa and Shadmehr \(2019\)](#) and [Sachs \(1989\)](#) is that our voter's choice for an illiberal government is not due to her shortsightedness.

¹⁷[Grillo and Prato \(2021\)](#) propose a different explanation for the dynamics of support for illiberalism that relies on voters' reference-dependent preferences.

subtle control of information, rather than through violent repression of the opposition. Importantly, they also convincingly argue that contemporary autocrats and illiberal leaders focus on economic performance rather than repression when addressing the general public.

Our theory of the rise and fall of illiberal democracies partially overlaps with theories that aim to explain the demand for populism. [Guiso, Herrera, Morelli and Sonno \(2019, 2020\)](#) document how economic insecurity fueled the demand for protection from globalization and the rise of populist leaders in Europe.¹⁸ [Guriev and Papaioannou \(2020\)](#) review the literature on the rise of populism focusing on the role of both economic and cultural causes. Although illiberalism and populism sometimes overlap, “only a minority of strongmen are populists and only a minority of populists is a strongman [...] the authoritarian characteristic of the strongman is not inherent to populism” ([Mudde and Rovira Kaltwasser, 2017](#)) (see also [Guriev and Papaioannou, 2020](#)).

In our model, illiberal governments offer protection to the median voter that a liberal government is unable to provide. While we do not mean to imply that illiberal governments are superior, we notice that this rhetoric is neither new nor confined to propaganda. Since at least [Samuelson \(1961\)](#), there has been a longstanding theory within economics and political science that some autocratic governments produce higher economic growth. Recently, [Collier \(2010\)](#) showed that, in some instances, a more liberal democracy can constrain economic growth. Furthermore, leaders facing less constraints on their power have been shown to have greater influence on a country’s economic performance and policies ([Jones and Olken, 2005](#)). [Cheibub, Hong and Przeworski \(2020\)](#) document how more solid democracies were significantly slower in limiting individual freedoms to stop the spread of the COVID-19 pandemic (see also [Stasavage, 2020a,b](#)). The offer of protection against the threat of socialism or elite capture was also at the basis of the rhetoric of fascist regimes in the 1930s ([Acemoglu, De Feo, De Luca and Russo, 2020](#)).

Our model studies a problem of dynamic delegation in which the illiberal government tries to persuade the voter to reelect it. This problem connects to the growing literature on dynamic Bayesian persuasion ([Bizzotto, Rüdiger and Vigier, 2021](#); [Ely, Frankel and Kamenica, 2015](#); [Ely, 2017](#); [Ely and Szydlowski, 2020](#); [Renault, Solan and Vieille, 2017](#)). In this literature (and using the principal-agent language from above) an agent controls the flow of information observed by a principal, and tries to persuade the principal to take an action either in each period or at a given deadline. [Che, Kim and Mierendorff \(2020\)](#) study a model in which, as in our framework, the agent can only commit to this period’s

¹⁸See also [Anelli, Colantone and Stanig \(2019\)](#); [Autor, Dorn, Hanson and Majlesi \(2020\)](#); [Colantone and Stanig \(2018, 2019\)](#); [Rodrik \(2018\)](#).

information structure, and the principal can choose when to take her action. Somewhat closer to us, [Orlov, Skrzypacz and Zryumov \(2020\)](#) study a model in which the agent wishes the principal to take a single irreversible action as late as possible. However, in contrast with our model, the incentive for the principal to take a decision earlier is the real cost of waiting and the agent’s optimal choice depends on a persistent state. As a result, these model are not suitable to capture the tradeoffs that arise when the principal can “fire” the agent to see (a more informative signal of) the truth for herself. That is, in our model, and in contrast to previous literature, the agent suffers from the fact that the principal dislikes being persuaded and the principal’s incentive to fire the agent increases when the agent is expected to engage in more persuasion in future periods. In addition, our model allows us to separate two concepts: the underlying state unobserved by the principal and the signal that the principal could observe were she to fire the agent. Disentangling these two concepts allows us to capture how transparency (the precision of uncensored information) affects the agent’s tradeoff through a (common to the literature) *persuasion* effect and a (novel) *accountability* effect, which operate in opposite directions.

3 The model

3.1 Summary

We study a model with a forward-looking and infinitely-lived voter. In each period, the voter chooses whether to elect a liberal or an illiberal government. A liberal government is a non-strategic actor that operates according to a constitution that constrains its action. An illiberal government does not abide by the constitution. We capture the stringency of the constitutional *constraints on the executive* with the parameter $\pi \in (0, 1)$.

The liberal constraints on the action of the executive have three effects.¹⁹ First, more stringent constraints guarantee greater liberties for the voter if she elects a liberal government. Therefore, more stringent constraints increase the *value of liberty*, $L = \mathcal{L}(\pi) > 0$, of electing a liberal government, where \mathcal{L} is strictly increasing and continuous. Second, more stringent constraints limit a liberal government’s ability to adequately respond to negative shocks. Therefore, more stringent constraints increase the chances of a shock that only an illiberal government would be able to adequately respond to. If such a shock

¹⁹Constitutions place differing types of constraints on executives. Here the focus is on the constraints that limit executive action to preserve individual freedoms. Another type of constraints limit the executive’s ability to seize powers from institutions of control such as parliaments, courts, and elections—we call such constraints “checks and balances” and we study their interaction with the constraint on the executive in [Section 8](#).

occurs and the voter has elected a liberal government, then the voter will suffer a cost.²⁰ After a normalization, electing a liberal government increases the expected cost of security for the voter by πS , where $S > 0$ is the voter's *value of security*. The parameter π is the probability of a shock that only an illiberal government would be able to adequately respond to.²¹ Third, more stringent constraints on the executive protect the flow of information from bureaucrats and mass media to the voter from governmental interference. Therefore, more stringent constraints increase the *transparency* of the information observed by the voter under a liberal government $q = Q(\pi)$, where Q is increasing and continuous.

Since it does not respect the constraints imposed by the constitution, an illiberal government does not guarantee the liberties that the voter would have under a liberal government and can freely set policies that intimidate bureaucrats from whistle-blowing and censor the media, so that the information observed by the voter is a garbled version of the one she would otherwise observe under a liberal government. The voter can observe the policies and intimidation chosen by the illiberal government but can only indirectly infer what information she may have received if she had chosen a liberal government instead. We call this information manipulation a *censorship policy*, but our model is more general in that it encompasses any mapping of the underlying signal of the state to a distribution of messages observed by the voter.

The objective of an illiberal government is to remain in power as long as possible. If the voter does not reelect an incumbent illiberal government, we assume that the fallen illiberal government ceases to exist; a new illiberal government is generated if and when the voter chooses to elect an illiberal government again.

Within our framework, a constitution is summarized by the constraints on the executive π which, by limiting the power of the government, guarantee an amount of liberties to individuals and minorities $L = \mathcal{L}(\pi)$ and of transparency and independence of the bureaucracy, the judiciary, and the media $q = Q(\pi)$. The functions \mathcal{L} and Q are the tech-

²⁰For sake of simplicity of exposition, we focus on shocks to the expected cost of security. However, an equivalent shock with the same effects could affect the value of more liberty.

²¹Our constraints on the executive measure the effective rate at which the constitution is constraining the action of governments that abide by it compared to governments that do not. Obviously, there are shocks that no government would be able to adequately respond to, as well as shocks that all governments would be able to neutralize. We normalize S to equal the expected cost of all shocks that can be neutralized. The probability of any negative shock is determined by factors outside the constitution, such as the hostility of neighboring countries or global macroeconomic conditions. However, with no constraints on the executive, the expected cost from a shock is the same whether the voter elects a liberal or illiberal government, so that electing a liberal government carries no additional cost of security: $\pi S = 0$. In contrast, with a constitution that totally immobilizes the executive, electing a liberal government exposes the voter to all shocks, while electing an illiberal government limits the shocks to those that cannot be neutralized. Therefore, electing a liberal government carries an additional cost of security $\pi S = S$.

nological possibility frontier on the design of constitutions. In the analysis of Sections 4 to 6, we describe comparative statics on π , L , and q separately. I.e., we study the effects of relaxing the technological constraint. We then return to the connection between the three elements in Section 7, where we derive normative conclusions about the design of stable liberal democracies.

3.2 Formal setup

A forward-looking voter lives for infinitely many periods $t \in \{1, 2, \dots\}$. In each period t the voter chooses the period- t government $g_t \in \{i, \ell\}$. The voter's payoff in period t , $v(g_t, \theta_t)$, depends on the period- t government, g_t , and a period-specific i.i.d. state, $\theta_t \in \{0, 1\}$:

$$v(g_t, \theta_t) = \begin{cases} L - S\theta_t & \text{if } g_t = \ell; \\ 0 & \text{if } g_t = i. \end{cases}$$

and π is the probability that $\theta_t = 1$, i.e., of the occurrence of a *shock* that the voter cannot be protected from unless the government is illiberal.

The government in period $t - 1$, $g_{t-1} \in \{i, \ell\}$ with $g_0 = \ell$, determines the information available to the voter at the beginning of period t . In particular, for a given stringency of the constraints π , there exists a signal $s(\theta_t) \in \{0, 1\}$ where $\Pr(s(\theta_t) = \theta_t) = q \in (1/2, 1)$. The government in period $t - 1$ chooses a *censorship policy* from the signal $s(\theta_t)$ into a message $m_t \in \{0, 1\}$ observed by the voter. More precisely, the government chooses the censorship policy $c_t : \{0, 1\} \rightarrow [0, 1]$, where $c_t(s)$ is the probability that the voter observes a period- t message $m_t = 1$ when $s(\theta_t) = s$.²² Without loss of generality, we focus on censorship policies c_t such that $c_t(1) \geq c_t(0)$. If $g_{t-1} = i$, an illiberal government strategically chooses the censorship policy. If $g_{t-1} = \ell$, then the liberal government is constrained by the constitution and cannot interfere with the level of transparency guaranteed by it. Formally, the liberal government chooses a censorship policy $c_t = c_\ell$ that fully reveals the signal $s(\theta_t)$, i.e., $c_\ell(s) = s$ for all $s \in \{0, 1\}$.

In each period t , the timing of the events is as follows. First, θ_t is realized, and the voter observes c_t and m_t . Second, the voter chooses whether to elect $g_t = i$ or $g_t = \ell$. Finally, the period- t government, g_t , chooses the censorship policy, c_{t+1} .

An illiberal government elected for the first time in period t receives a rent $R > 0$ for period t and each subsequent period until the first period $t' > t : g_{t'} = \ell$, at which point

²²The binary structure of the government's message is without loss of generality given the government's objectives (see below) and our focus on Markovian strategies.

a liberal government is elected and the previous illiberal government is replaced by a new illiberal government. The voter and the illiberal government maximize their present discounted value of their payoffs and discount future periods with factor $\delta \in (0, 1)$.²³

3.3 Welfare

We note here that social welfare in our model is *not* fully captured by $v(g_t, \theta_t)$ as this only represents the utility of the pivotal voter, who may trade off the liberty of a minority for a greater security of her own economic interests. In Section 7 we extend the model to allow for a continuum of voters with differing values of liberty and security. In this case, the pivotal voter is the one with the median relative value of liberty, L/S . Therefore, the single voter in our benchmark model should be interpreted as representing this median voter. In Section 7 we will return to this distinction and discuss how different members of a polity are differently affected by changes in the elements in its constitution.

3.4 Equilibrium concept

We characterize the perfect Bayesian equilibria (Fudenberg and Tirole, 1991) of our model in which the voter and the illiberal governments play pure Markovian strategies. Following the Bayesian persuasion literature (Kamenica and Gentzkow, 2011), we focus on equilibria in which the voter chooses an illiberal government whenever she is indifferent. Because θ_t is i.i.d., at time t , the voter's payoff relevant history is fully captured by c_t . Similarly, in period t , the illiberal government's payoff relevant history is fully captured by g_t . However, the illiberal government only takes action if $g_t = i$. Therefore, for any period t , the illiberal government's set of optimal censorship policies is time independent. For simplicity, we focus on equilibria where the illiberal government's choice of censorship policy is also time independent.²⁴

A Markovian pure strategy for the voter is a mapping g from the message observed by the voter, m_t , and the censorship policy, c_t , into the period- t government, g_t . A Markovian pure strategy for the period- t illiberal government is a choice c of censorship policy, c_{t+1} . The voter's belief that $\theta_t = 1$, denoted by μ_t , is a mapping from the message observed by the voter, m_t , and the censorship policy, c_t , into a probability. A Markovian assessment is therefore a triple $\sigma = (g, c, \{\mu_t\}_{t=1}^{\infty})$.

²³As we will show, the illiberal government's equilibrium strategy is independent of their own future payoff stream; hence, our results hold verbatim if the voter and illiberal government have discount factors $\delta_V = \delta$ and $\delta_i \in (0, 1)$, respectively.

²⁴This assumption does not affect the payoff-relevant properties of the equilibria, i.e., in each period t , the probability that an illiberal or liberal government is elected, conditional on the state, θ_t .

Let

$$V(g_t = \ell \mid \sigma) := \mathbb{E} \left[\sum_{\tilde{t}=t+1}^{\infty} \delta^{(\tilde{t}-t-1)} v(g_{\tilde{t}}, \theta_{\tilde{t}}) \mid g_t = \ell, \sigma \right]$$

and

$$V(g_t = i \mid \sigma) := \mathbb{E} \left[\sum_{\tilde{t}=t+1}^{\infty} \delta^{(\tilde{t}-t-1)} v(g_{\tilde{t}}, \theta_{\tilde{t}}) \mid g_t = i, \sigma \right]$$

denote the expected continuation payoffs from electing a liberal and illiberal government, respectively. A strategy g^* for the voter is sequentially rational, given σ , if in each period t , the voter's choice of government maximizes her expected payoff, i.e.,

$$g^*(m_t, c_t) = i \text{ if and only if } L - \mu_t(m_t, c_t)S + \delta V(g_t = \ell \mid \sigma) \leq \delta V(g_t = i \mid \sigma). \quad (1)$$

A strategy c^* for the illiberal government is sequentially rational, given σ , if in each period t in which $g_t = i$, the illiberal government's choice of censorship maximizes (the discounted value of) their expected time in power

$$\mathcal{X}(\{g_{t'}\}_{t+1}^{\tilde{t}}) = \begin{cases} 1 & \text{if } g_{t'} = i \text{ for all } t' \in \{t+1, \dots, \tilde{t}\}, \\ 0 & \text{otherwise.} \end{cases}$$

I.e., for all censorship policies c' ,

$$\mathbb{E} \left[\sum_{\tilde{t}=t+1}^{\infty} \delta^{\tilde{t}-t} R \mathcal{X}(\{g_{t'}\}_{t+1}^{\tilde{t}}) \mid c_{t+1} = c^*, \sigma \right] \geq \mathbb{E} \left[\sum_{\tilde{t}=t+1}^{\infty} \delta^{\tilde{t}-t} R \mathcal{X}(\{g_{t'}\}_{t+1}^{\tilde{t}}) \mid c_{t+1} = c', \sigma \right]. \quad (2)$$

Finally, in equilibrium, in each period t the voter's belief μ_t is derived using Bayes' rule,²⁵ i.e.,

$$\begin{aligned} \mu_t(1, c_t) &= \mu_t^*(1, c_t) := \Pr[\theta_t = 1 \mid m_t = 1, c_t] \\ &= \frac{\pi[c_t(1)q + c_t(0)(1-q)]}{\pi[c_t(1)q + c_t(0)(1-q)] + (1-\pi)[c_t(1)(1-q) + c_t(0)q]}, \end{aligned} \quad (3)$$

²⁵By Property 1 of Definition 3.1 in [Fudenberg and Tirole \(1991\)](#), under any censorship policy, c_t —whether on or off the equilibrium path—and any message, m_t , that occurs with positive probability under c_t , the voter's belief must be derived using Bayes' rule from the prior belief that $\theta_t = 1$, π , and the conditional probabilities, $c_t(0)$, $c_t(1)$, and q .

and

$$\begin{aligned} \mu_t(0, c_t) &= \mu_t^*(0, c_t) := \Pr[\theta_t = 1 \mid m_t = 0, c_t] \\ &= \frac{\pi[(1 - c_t(1))q + (1 - c_t(0))(1 - q)]}{\pi[(1 - c_t(1))q + (1 - c_t(0))(1 - q)] + (1 - \pi)[(1 - c_t(1))(1 - q) + (1 - c_t(0))q]}. \end{aligned} \quad (4)$$

Definition 1 (Equilibrium.) *An assessment $\sigma^* = (g^*, c^*, \{\mu_t^*\}_{t=1}^\infty)$ is an equilibrium if, for each period t , g^* satisfies (1) for $\sigma = \sigma^*$; c^* satisfies (2) for $\sigma = \sigma^*$; and $\mu_t^*(m_t, c_t)$ satisfies (3) and (4).*

In the following sections, we characterize the essentially unique equilibrium of the model.²⁶ All proofs appear in Appendix A.

4 The optimal voting and censorship strategies

We begin by characterizing the voter's sequentially optimal strategy. In each period t , the voter observes a message, m_t , about θ_t and forms a belief about whether $\theta_t = 1$. She then decides whether to elect an illiberal or a liberal government. In period t , electing a liberal government gives more liberty, L , but has an expected cost of security, $\mu_t(m_t, c_t)S$; electing an illiberal government always provides a payoff of zero. Therefore, absent dynamic considerations, the voter elects an illiberal government whenever the expected cost of security is greater than the value of liberty. However, electing an illiberal government comes at an additional dynamic cost for the voter. As an illiberal government can engage in censorship, electing an illiberal government today induces the voter to take a less-informed choice of government tomorrow. This means that the expected continuation payoff of electing an illiberal government, $V(g_t = i \mid \sigma^*)$, is smaller than the expected continuation payoff of electing a liberal government, $V(g_t = \ell \mid \sigma^*)$. We define the (equilibrium) *accountability cost of illiberalism* as

$$A(\pi, q, L, S, \delta \mid \sigma^*) := V(g_t = \ell \mid \sigma^*) - V(g_t = i \mid \sigma^*). \quad (5)$$

We will show that the accountability cost of illiberalism is decreasing in the informativeness of the censorship policy that the voter expects the illiberal government to choose. In

²⁶In Appendix B we prove that the equilibrium is *essentially unique*: for any given set of parameters, the probability that an illiberal government is elected (and hence the voter's and the illiberal government's expected payoff) are equal in all equilibria.

particular, the maximum accountability cost of illiberalism is²⁷

$$\bar{A}(\pi, q, L, S, \delta) := \max \left\{ 0, \frac{\Pr[s(\theta_t) = 0]}{1 - \delta \Pr[s(\theta_t) = 0]} (L - \Pr[\theta_t = 1 \mid s(\theta_t) = 0]S) \right\}. \quad (6)$$

This maximal cost may be achieved if an illiberal government chooses a *full censorship* policy $c_F := \{c : c(s) = \bar{c} \in [0, 1] \forall s \in \{0, 1\}\}$ and is able to ensure its own reelection.

Lemma 1 says that in equilibrium the voter elects an illiberal government if and only if the expected cost of security, $\mu_t(m_t, c_t)S$, is greater than the sum of the value of liberty, L , and the discounted accountability cost of illiberalism, $\delta A(\pi, q, L, S, \delta \mid \sigma^*)$.

Lemma 1 (The voter's optimal strategy.) *In equilibrium, the voter elects the illiberal government in period t if and only if*

$$\mu_t(m_t, c_t)S \geq L + \delta A(\pi, q, L, S, \delta \mid \sigma^*). \quad (7)$$

We now turn to the illiberal government's problem. Lemma 2 establishes that the illiberal government's problem can be reduced to the problem of choosing the censorship policy that maximizes their probability of being reelected in the following period.

Lemma 2 (The illiberal government's problem.) *In equilibrium, an illiberal government in period t chooses a censorship policy $c_{t+1} = c^*$ that maximizes its reelection probability in period $t + 1$.*

In what follows, we characterize the illiberal government's optimal strategy. Lemma 3 says that, in equilibrium, illiberal governments are either indifferent between all censorship policies, or choose one of two types of censorship policies.

Lemma 3 (Optimal censorship.) *In equilibrium, if an illiberal government is elected then it is either indifferent between all censorship policies or*

1. *if $\pi S \geq L + \delta \bar{A}(\pi, q, L, S, \delta)$, it chooses the full censorship policy, c_F , or any other censorship policy c^* satisfying*

$$\mu_t^*(m_t, c^*)S \geq L + \delta \bar{A}(\pi, q, L, S, \delta) \quad \forall m_t \in \text{Supp}(c^*),$$

where $\text{Supp}(c_t)$ is the support of c_t ;

²⁷The maximum accountability cost of illiberalism equals zero whenever $L \leq \Pr[\theta_t = 1 \mid s(\theta_t) = 0]S$ because, even absent censorship, the voter always prefers to elect an illiberal government. Otherwise, the maximum cost is the expected payoff of choosing a liberal government for as long as $s(\theta_t) = 0$ before then turning to illiberalism forever.

2. Otherwise, it chooses $c^*(1) = 1$ and $c^*(0)$ such that

$$\mu_t^*(1, c^*)S = L + \delta A(\pi, q, L, S, \delta \mid \sigma^*). \quad (8)$$

To understand Lemma 3, it is easier to focus on the censorship policies that always send signal $m_t = 1$ whenever $s(\theta_t) = 1$, i.e., $c(1) = 1$. Point 1 of Lemma 3 considers the case when the voter reelects the illiberal government absent any information about the state (i.e., the equilibrium expected cost of security is greater than the value of liberty plus the discounted maximum accountability cost). In this case, it is optimal for the illiberal government to choose $c(0) = 1$ —the full censorship policy.

Otherwise, the illiberal government engages in *partial censorship*: it chooses $c(0) < 1$. The precise probability is the one that induces a voter's belief $\mu_t^*(1, c)$ such that, upon observing message $m_t = 1$, the voter is indifferent between electing a liberal or an illiberal government for the next period. The intuition behind the optimal partial censorship is akin to that in the Bayesian persuasion literature.²⁸ Intuitively, greater censorship of unfavorable signals (a greater $c(0)$) increases the chances of reelection for the illiberal government up to the point at which the voter would correctly infer that, even upon observing message $m_t = 1$, a liberal government is preferable.

However, there is a key difference between the role of censorship in our model and the one it plays in the Bayesian persuasion literature. In our model, greater censorship affects the voter's optimal choice through *two* distinct channels. First, as mentioned above, more censorship directly increases the probability that the voter believes that today she would prefer illiberal policies to protect her from negative shocks. However, a second effect arises when the voter *expects* more censorship from illiberal governments. More (expected) censorship dynamically increases the accountability cost of illiberalism, in-

²⁸A censorship policy induces beliefs $\mu_t^*(1, c) \geq \pi \geq \mu_t^*(0, c)$. Notice that the probability of reelection is given by the probability that the voter observes a message m_t inducing a belief that makes her at least indifferent between electing an illiberal and a liberal government. Thus, any policy c' such that $\mu_t^*(1, c') < \mu_t^*(1, c^*)$ yields a probability of reelection equal to zero. Suppose that there existed an optimal partial censorship c' inducing belief $\mu_t^*(1, c') > \mu_t^*(1, c^*)$. I.e., upon observing message $m_t = 1$ the voter strictly prefers to reelect the illiberal government. By Bayes' rule, this implies $c'(0) < c^*(0)$. The probability of reelection under this policy is

$$c'(1)(q\pi + (1 - q)(1 - \pi)) + c'(0)((1 - q)\pi + q(1 - \pi))$$

which is less than the probability of reelection under the optimal censorship c^* :

$$(q\pi + (1 - q)(1 - \pi)) + c^*(0)((1 - q)\pi + q(1 - \pi)).$$

Egorov, Guriev and Sonin (2009) study a different reason for why strong leaders may allow for some amount of media freedom.

creasing the voter's value of a liberal system in which she can make a better-informed choice of government for the future. The tradeoff between these two effects drives the results in the following sections.

5 Regime stability and cycles

We now show that, depending on the constitution parameters π , L , and q , the equilibrium behavior of our model gives rise to four different regimes.

Efficient stable liberal democracy. In this regime, the voter always elects a liberal government, and this choice is optimal given the uncensored information offered by the constitution, $s(\theta_t)$. In equilibrium, an illiberal government chooses any censorship mapping c^* ; the voter elects an illiberal government if and only if $\mu_t^*(m_t, c_t)S \geq L$; and $g_t = \ell$ for all $t \geq 1$.

Efficient stable illiberal democracy. In this regime, the voter always elects an illiberal government. Yet, her choice of government is optimal given the uncensored information offered by the constitution, $s(\theta_t)$. In equilibrium, an illiberal government chooses any censorship mapping c^* ; the voter elects an illiberal government if and only if $\mu_t^*(m_t, c_t)S \geq L$; and $g_t = i$ for all $t \geq 1$.

Inefficient stable illiberal democracy. In this regime, the voter eventually (but not necessarily at $t = 1$) elects an illiberal government that remains in power forever. Crucially, absent censorship, the voter would find it optimal to cycle between liberal and illiberal governments. However, once she elects an illiberal government, the illiberal government chooses a full censorship policy and ensures its own reelection forever. To completely characterize the equilibrium, under a liberal government, the voter elects an illiberal government if and only if $m_t = 1$, which occurs with probability $\pi q + (1 - \pi)(1 - q)$. Once an illiberal government has been elected, it chooses the full censorship policy, c_F , or any other censorship policy c^* satisfying

$$\mu_t^*(m_t, c^*)S \geq L + \delta \bar{A}(\pi, q, L, S, \delta) \quad \forall m_t \in \text{Supp}(c^*);$$

and the voter reelects the illiberal government with probability one.

Cycling liberal-illiberal democracy. In this regime, the voter cycles between liberal and illiberal governments. In equilibrium, in any period t , the voter elects the illiberal government if and only if $m_t = 1$, and the illiberal government engages in partial censorship, c^* . Therefore, if $g_{t-1} = \ell$, the voter elects the illiberal government with probability $\pi q + (1 - \pi)(1 - q)$, and if $g_{t-1} = i$, the illiberal government falls with probability $(1 - c^*(0))(\pi(1 - q) + (1 - \pi)q) \in (0, 1)$. However, this cycling liberal-illiberal regime is still inefficient: absent censorship, the voter would optimally revert back to a liberal government with strictly higher probability.

Proposition 1 describes when each of these regimes arises in equilibrium.

Proposition 1 (Regime stability and cycles.) *In the essentially unique equilibrium,*

1. $L \leq \mu_t^*(0, c_\ell)S$ induces an efficient stable illiberal democracy;
2. $\mu_t^*(0, c_\ell)S < L \leq \pi S - \delta \bar{A}(\pi, q, L, S, \delta)$ induces an inefficient stable illiberal democracy;
3. $\pi S - \delta \bar{A}(\pi, q, L, S, \delta) < L \leq \mu_t^*(1, c_\ell)S$ induces a cycling liberal-illiberal democracy; and
4. $\mu_t^*(1, c_\ell)S < L$ induces an efficient stable liberal democracy.

Proposition 1 says that if the value of liberty, L , is sufficiently large, while the cost of security, S , the constraints on the executive, π , as well as the transparency of the institutions, q , are sufficiently small, then a stable liberal democracy arises. Intuitively, under this constitution, the liberal government offers valuable liberty and sufficient security to the voter, so that the voter always prefers to elect a liberal government.²⁹ Conversely, if the value of liberty, L , is sufficiently small, while the cost of security, S , and the constraints on the executive, π , are sufficiently large, but the transparency of the institutions, q , is sufficiently small, then a stable illiberal democracy arises. Intuitively, under this constitution, the liberal government does not offer enough liberty nor security, and so the voter always prefers to elect an illiberal government.³⁰

In between these two extreme cases lie the two intermediate regimes. First, for some parameters the liberal government offers valuable liberty and security but, whenever the voter observes the period- t message $m_t = 1$, her expected cost of security is so high that she prefers to elect an illiberal government. She correctly anticipates that the decision to elect an illiberal government is forever: because of censorship she will continue to

²⁹In this case, the accountability cost of illiberalism is zero because censorship cannot affect the voter's ability to choose the best government—the liberal government.

³⁰In this case, the accountability cost of illiberalism is again zero because censorship cannot affect the voter's ability to choose the best government—the illiberal government.

elect an illiberal government even though, absent censorship, she would optimally revert back to a liberal government with probability $\pi(1 - q) + (1 - \pi)q$ in each period. In fact, the voter correctly attributes to an illiberal government the maximal accountability cost, \bar{A} . Therefore, an inefficient stable illiberal democracy arises only if the expected cost of security under full censorship is greater than the sum of the value of liberty and the discounted maximal accountability cost:

$$\pi S \geq L + \delta \bar{A}(\pi, q, L, S, \delta).$$

When this condition is not satisfied, an illiberal government cannot ensure its own reelection. However, if anything, this increases the voter's propensity to elect an illiberal government in the first place because illiberalism now comes with a smaller accountability cost. This gives rise to cycles of liberal and illiberal governments. Under this constitution, whenever the voter observes the period- t message $m_t = 1$, her expected cost of security is so high that she prefers to elect an illiberal government. However, because the voter values liberty sufficiently, the illiberal government is unable to ensure its reelection and the full censorship policy is suboptimal. Instead, the illiberal government optimally chooses a partial censorship policy, c^* , that induces the voter to reelect the illiberal government only if she observes the message $m_t = 1$. Therefore, the illiberal government falls with positive probability equal to $(1 - c^*(0))(\pi(1 - q) + (1 - \pi)q)$.

For any given value of security S , Proposition 1 maps constitutions (parameters π , L , and q) into regimes. It is useful to normalize the value of liberty as the voter's *relative value of liberty*, $\bar{L} := L/S$. Corollary 1 says that, as the voter's relative value of liberty increases, the regime transitions from an efficient stable illiberal democracy to an inefficient stable illiberal democracy to a cycling liberal-illiberal democracy, and then finally to an efficient stable liberal democracy.

Corollary 1 (A typology of regimes.) *There exist cutoffs $0 < \underline{\kappa}(\pi, q) < \kappa(\pi, q; \delta) < \bar{\kappa}(\pi, q) < 1$ such that, in equilibrium, a constitution with*

1. $\bar{L} \leq \underline{\kappa}(\pi, q)$ induces an efficient stable illiberal democracy;
2. $\bar{L} \in (\underline{\kappa}(\pi, q), \kappa(\pi, q; \delta)]$ induces an inefficient stable illiberal democracy;
3. $\bar{L} \in (\kappa(\pi, q; \delta), \bar{\kappa}(\pi, q)]$ induces a cycling liberal-illiberal democracy; and
4. $\bar{L} > \bar{\kappa}(\pi, q)$ induces an efficient stable liberal democracy.

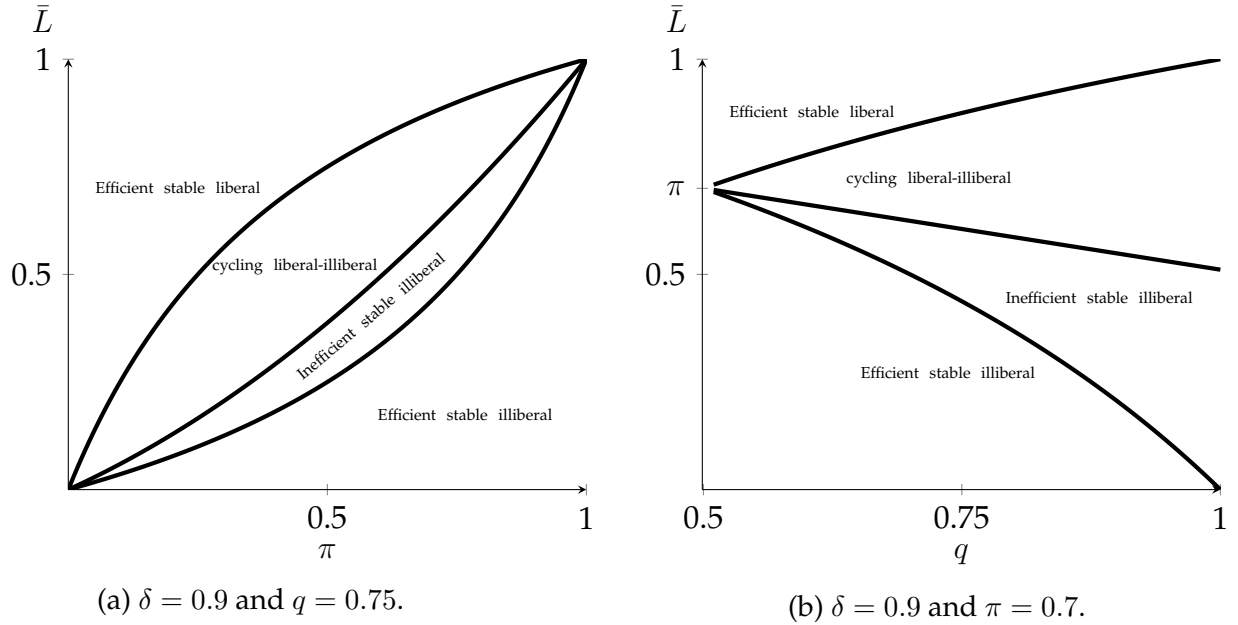


Figure 1: Typology of regimes.

The closed form of the cutoffs is given in Proposition A.1 in Appendix A. Panels (a) and (b) in Figure 1 illustrate this regime typology in the (π, \bar{L}) and (q, \bar{L}) space, respectively. In each figure, the plotted curves from bottom to top correspond to $\underline{\kappa}(\pi, q)$, $\kappa(\pi, q, \delta)$, and $\bar{\kappa}(\pi, q)$, respectively. As Panel (b) shows, for a given π an increase in q can change the regime from a stable one to an unstable one. Intuitively, more precise information for the voter implies that the voter is more likely to change which government she prefers upon observing a new signal. This key intuition drives many of the results in Section 6, which focuses on the stability of constitutions by studying how the different parameters affect the rise and fall of illiberal democracies.

5.1 Exploring the boundaries of the model

We now briefly pause our analysis of the rise and fall of illiberal democracies to return to the general problem we have identified in a principal's choice to hire an agent to manage a project. Doing so allows us to clarify the implications of some of our assumptions.

We have thus far studied one possible version of the problem. In our benchmark model, the agent has the ability to commit to a censorship policy, but only for one period. In the context of our main application, we believe it is natural to think that the illiberal government sets, at least for some time, policies that intimidate media and bureaucrats, but cannot commit to further relaxations or restrictions in a more distant future. However,

in other contexts it may be more realistic to assume that the agent can commit to a multi-period policy of information design. In Appendix C we study a variation of the model in which the government commits to long-term censorship policies and we show that the key insights and the characterization of the regimes are very similar to those in Section 5.

The Bayesian persuasion assumption that the agent commits to an information design rather than choosing what message to send is analytically convenient. In our context, this assumption also has important elements of realism, as in practice governments establish policies and intimidate media and whistleblowers with the objective of limiting information leaks; however, they are limited in how much they can conceal information once it has been leaked. In other contexts, this assumption may be less realistic, for example if the principal only receives information directly from the agent. For completeness we study in Appendix D a non-commitment version of our model. The results mimic very closely those we present in Section 5. However, in contrast with our benchmark model, when in equilibrium the voter cycles between liberal and illiberal governments, she does so inefficiently because she reelects the illiberal government too infrequently (instead of too frequently). Intuitively, when the agent’s communication is cheap talk, the agent is unable to convey any decision-relevant information to the principal. Therefore, in equilibrium, the principal expects the agent to—in effect—censor all information. Thus, she either always retain the agent because it is optimal absent any period-specific information, or she fires him after just one period. This is inefficient: if the principal could access all the available information, she would retain the agent (on average) whenever $s(\theta_t) = 1$.

As we discussed, it is essential to our tradeoff of interest that the voter (principal) may prefer to return to a liberal government (fire the agent) so to observe more precise information. For completeness, we study in Appendix E a version of the model in which the ability to manipulate information is independent of the ability to provide security and both liberal and illiberal governments can censor strategically. This partially mutes our key tradeoff. Therefore, in equilibrium we either have stable liberal democracies or stable illiberal democracies. Said otherwise, since the principal has no way to see things with her own eyes, once she decides to hire an agent, there is no value for her in changing her mind again.³¹

Our benchmark model assumes that the underlying state θ_t is independently distributed across periods. Essential to our tradeoff of interest is that the state is not fully persistent. Otherwise, our model would be a stopping game of learning for the principal.

³¹In fact, if both governments are sufficiently limited in how much they can censor information, the equilibrium can feature cycles between liberal and illiberal democracies in which both types of government may last longer than optimal for the voter.

Realistically, shocks to countries and firms alike are partially persistent. In this case, our main tradeoff of interest still drives behavior; however, its effects are compounded by other dynamics. For example, under a liberal government, upon observing the realization of a negative shock at time t , the voter's belief regarding the probability of a negative shock at time $t + 1$ would increase, therefore increasing her likelihood to elect an illiberal government for any message observed at time t . Furthermore, the illiberal government's ability to censor would also depend on how much time it has been in power and the whole history of messages observed by the voter under its regime.

6 The rise and fall of illiberal democracies

This section studies how different parameters of our model affect the dynamics of evolution of democracies. In particular, we ask what favors the rise of illiberal democracies, how different constitutions affect the type of censorship illiberal governments engage in, and, finally, what induces illiberal democracies to fall sooner, if at all.

6.1 The rise

We begin by studying what causes the rise of illiberal democracies from a liberal government. An immediate implication of Proposition 1 is that, unless the constitution induces an efficient stable liberal or efficient stable illiberal democracy, an illiberal government is elected whenever $s(\theta_t) = 1$. When this happens, the probability that the period-specific shock is one that the voter cannot be protected from by a liberal government (and therefore the expected cost of security) is high. This can be thought of as a crisis that increases the voter's perception that their welfare is under threat from terrorists, foreign powers, recessions, or a pandemic. The probability that such a crisis arises, as well as whether the voter is willing to elect an illiberal government during a crisis, depends on the different elements of the constitution.

Proposition 2 shows how the different elements of a constitution π , L , and q affect the probability that an illiberal government is elected following a liberal one. First, by Corollary 1, the probability that an illiberal government is elected equals 1 when the relative value of liberty is sufficiently small, and zero when it is sufficiently large. Otherwise, in inefficient stable illiberal democracies and in cycling democracies, an increase in the relative value of liberty induces voters to prefer a liberal government more, as the value of liberalism is greater. So the probability that an illiberal government follows a liberal one (weakly) decreases in the relative value of liberty. Second, more constraints on the execu-

tive (a greater π) increases the expected cost of security of choosing a liberal government for all levels of transparency, q . Therefore, more constraints on the executive induce the voter to prefer an illiberal government more. So the probability that an illiberal government follows a liberal one (weakly) increases with the amount of constraints placed on the executive by the constitution. Finally, the transparency of the liberal constitution also plays a role in the voter's choice. Proposition 2 says that whether greater transparency increases or decreases the probability of an illiberal government depends on the other parameters of the constitution as well as on the amount of transparency itself. These non-monotonicities are caused by the fact that, as shown in Panel (b) of Figure 1, an increase in q can cause a switch from a stable to an unstable regime. For example, Proposition 2 says that when $\pi < \bar{L}$, increasing q has no effect on the probability of a rise of an illiberal democracy for sufficiently small q , because the regime is a stable liberal democracy (the probability of a transition is zero). However, after q passes a threshold, the regime becomes a cycling liberal-illiberal democracy, and in the case in which $\pi < 1/2$, the probability of a liberal-illiberal transition is positive and maximal precisely at this threshold and then decreases for larger q , so that the total effect of q on the probability of a rise of an illiberal democracy is single-peaked.

Proposition 2 (The rise of an illiberal democracy.) *Suppose that in period $t - 1$ the government is liberal. The probability that the period- t government is illiberal is weakly increasing in the executive constraints, π ; weakly decreasing in the relative value of liberty, \bar{L} ; and not necessarily monotonic in the transparency of liberal institutions, q :*

1. *if $\pi < 1/2$ and $\pi < \bar{L}$, then it is single-peaked in q . It equals 0 for q sufficiently small and strictly decreases with q otherwise;*
2. *if $\pi < 1/2$ and $\pi > \bar{L}$, then it is decreasing in q . It equals 1 for q sufficiently small and strictly decreases with q otherwise;*
3. *if $\pi > 1/2$ and $\pi < \bar{L}$, then it is increasing in q . It equals 0 for q sufficiently small and strictly increases with q otherwise;*
4. *if $\pi > 1/2$ and $\pi > \bar{L}$, then it is U-shaped in q . It equals 1 for q sufficiently small and strictly increases with q otherwise.*

Panels (a) and (b) in Figure 2 illustrate the relationship between the illiberal government's election probability and the transparency of liberal institutions, q , under two parametric assumptions about π . This figure captures all the cases discussed in Proposition 2.

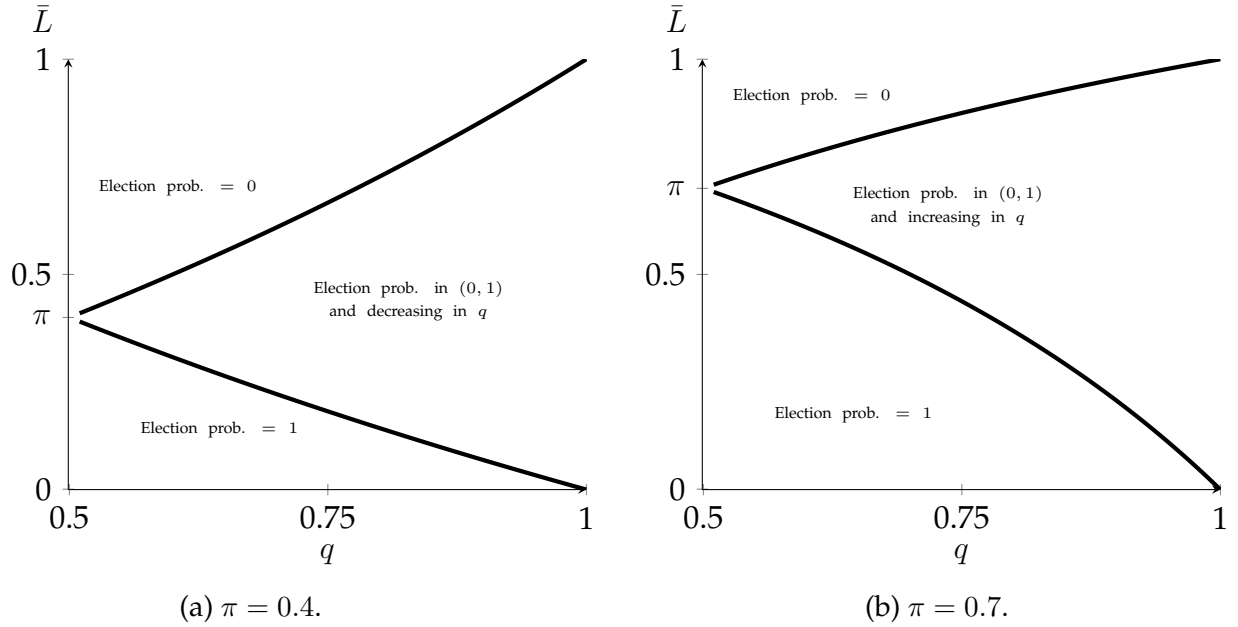


Figure 2: Illiberal government's election probability.

Intuitively, Point 1 of Proposition 2 refers to a constitution in which there are few constraints on the executive and the voter values liberty very much. Then, absent any information about the period-specific shock, the voter would always elect a liberal government (efficient stable liberal regime). The voter would resort to electing an illiberal government only if she were to observe sufficiently precise information about the current period's shock. Therefore, if the signal s is sufficiently precise (large q), the regime transitions to a cycling democracy and she elects an illiberal government whenever she observes $m_t = 1$. Because $\pi < 1/2$, a more precise signal sends message $m_t = 1$ with lower probability, and therefore the probability of electing an illiberal government is strictly decreasing in q .

Point 2 (Point 3) refers to a situation in which there are few (many) constraints on the executive and the voter's value of liberty is smaller (greater) than the expected cost of security under full censorship, πS . In this case, the voter always prefers to elect an illiberal (liberal) government when her information is sufficiently imprecise. However, the voter would rather elect the illiberal government when the current period's shock requires illiberal means of protection ($\theta_t = 1$) and otherwise to elect the liberal government. Therefore, if the signal s is sufficiently precise, the regime transitions to one in which the voter elects an illiberal government if and only if she observes $m_t = 1$. Because $\pi < 1/2$ ($\pi > 1/2$), a more precise signal sends message $m_t = 1$ with lower (higher) probability, and therefore the probability of electing an illiberal government is strictly decreasing (increasing) in q .

Point 4 is a mirror image of Point 1: the voter would elect a liberal government only if she were to observe sufficiently precise information that the current period's shock does not require illiberal means of protection. For more informative signals, because $\pi > 1/2$, the probability of electing an illiberal government is strictly increasing in q .

6.2 The reign: Optimal censorship

We now turn to studying how different elements affect an illiberal government's censorship policy. Recall from Lemmas 2 and 3 that, in stable illiberal democracies, the illiberal government is indifferent among many censorship policies, including the full censorship policy, c_F . In this section, we focus on the equilibrium where the illiberal government chooses the full censorship policy whenever it is optimal. Without loss of generality, we identify the full censorship policy, c_F , with the censorship policy that always sends $m_t = 1$.

Proposition 3 says that greater constraints on the executive (greater π) increases the expected cost of security for all censorship policies, and therefore allows an illiberal government to engage in more censorship (greater $c^*(0)$). In contrast, a greater relative value of liberty (greater \bar{L}) forces an illiberal government to engage in less censorship. However, the effect of the transparency of liberal institutions, q , on the optimal amount of censorship depends on the other parameters of the constitution.

Proposition 3 (The reign of an illiberal democracy.) *Whenever, in equilibrium, an illiberal government is elected with positive probability, the amount of censorship, $c^*(0)$, is weakly increasing in the constraints on the executive, π ; weakly decreasing in the relative value of liberty, \bar{L} ; and (1) if $\pi < \bar{L}$, it is increasing in the transparency of liberal institutions, q ; (2) if $\pi > \bar{L}$, it is decreasing in the transparency of liberal institutions, q .*

As the transparency of liberal institutions increases (greater q) there are two effects: a *persuasion* effect and an *accountability* effect. Intuitively, more transparent institutions increase an illiberal government's ability to persuade the voter to reelect the government. More precisely, when an illiberal government engages in the optimal partial censorship, upon observing message $m_t = 1$, the voter is indifferent between electing a liberal or an illiberal government for the next period. For a given censorship policy, a more transparent liberal institution (greater q) provides the voter with greater incentive to elect an illiberal government upon observing message $m_t = 1$. Therefore, this persuasion effect induces an illiberal government to increase the amount of censorship when the liberal institutions are more transparent. This effect is common in the Bayesian persuasion literature ([Kamenica and Gentzkow, 2011](#); [Kolotilin et al., 2019](#)).

In our model, more transparent institutions also affect an illiberal government's censorship through an accountability effect. Recall from Lemma 1 that the voter's choice to reelect the illiberal government depends whether the expected cost of security exceeds the sum of the value of liberty and the accountability cost of illiberalism. Naturally, when liberal institutions are more transparent, the accountability cost of illiberalism is greater because returning to a liberal government allows the voter to make even more informed decisions in the future. To keep the voter indifferent, an illiberal government must provide the voter with greater information to compensate them for the greater accountability cost. Therefore, this accountability effect induces an illiberal government to decrease the amount of censorship when the liberal institutions are more transparent.

Proposition 3 says that the persuasion effect dominates when the constitution has few constraints on the executive and the voter values liberty so much that, absent any information about the period-specific shock, the voter would always elect a liberal government. Otherwise, the accountability effect dominates.

6.3 The fall

We now turn to studying what causes illiberal governments to fall if they do at all. Because of censorship, the voter is unable to cause the fall of an illiberal government as soon as there is no crisis, i.e., as soon as $s(\theta_t) = 0$. Instead, the only thing that may trigger a fall of an illiberal democracy is the illiberal government's failure to conceal that there is no crisis: $m_t = 0$. Therefore, in cycling democracies and in inefficient stable illiberal democracies, illiberal governments remain in power for longer than the voter would optimally choose in the absence of censorship.

We now show what factors increase the probability that an illiberal government stays in power. Proposition 4 says that an illiberal government's reelection probability is increasing in the constraints on the executive, π ; decreasing in the relative value of liberty, \bar{L} ; but not necessarily monotonic in the transparency of liberal institutions, q .

Proposition 4 (The fall of an illiberal democracy.) *Whenever, in equilibrium, an illiberal government is elected with positive probability, it is reelected with probability weakly increasing in the constraints on the executive, π ; weakly decreasing in the relative value of liberty, \bar{L} ; and not necessarily monotonic in the transparency of liberal institutions, q . However: (1) if $1/2 < \pi < \bar{L}$, it is increasing in q ; (2) if $\bar{L} < \pi < 1/2$, it is decreasing in q .*

To understand how the transparency of liberal institutions affects the ability of an illiberal government to be reelected, it is useful to write the explicit formula for the derivative of the illiberal government's reelection probability with respect to q . Recall that,

in equilibrium, the voter reelects an illiberal government whenever she observes $m_t = 1$; therefore, this formula is

$$\frac{\partial \Pr[m_t = 1 \mid c^*]}{\partial q} = \underbrace{\left(\frac{\partial \Pr[s(\theta_t) = 1]}{\partial q} + c^*(0) \frac{\partial \Pr[s(\theta_t) = 0]}{\partial q} \right)}_{\text{Mechanical}} + \underbrace{\Pr[s(\theta_t) = 0] \frac{\partial c^*(0)}{\partial q}}_{\text{Strategic}}. \quad (9)$$

This formula highlights that there are two effects of transparency on the illiberal government's reelection. First, there is a *mechanical* component whereby, for a given censorship policy, the transparency of the liberal institutions affects the probability that the voter observes message $m_t = 1$ because the institutions determine the likelihood of an underlying signal $s(\theta_t) = 1$. The direction of the mechanical effect depends on the constraints on the executive, π . When $\pi < 1/2$ ($\pi > 1/2$), a more transparent institution produces an underlying signal $s(\theta_t) = 1$ with lower (higher) probability and, all else equal, the illiberal government's reelection probability decreases (increases) with q . Second, there is a *strategic* component whereby the transparency of the liberal institutions affects the optimal censorship policy of an illiberal government. By Proposition 3, when $\pi < \bar{L}$ ($\pi > \bar{L}$) the persuasion (accountability) effect determines the sign of the strategic component, and the illiberal government's reelection probability increases (decreases) with q .

Point 1 (Point 2) of Proposition 4 refers to a constitution where the persuasion (accountability) effect determines the sign of the strategic component and the mechanical component operates in the same direction; therefore, the illiberal government's reelection probability monotonically increases (decreases) with q . In the remaining cases, the two components have opposite signs. As illustrated by Figure 3, for low values of q , the strategic component dominates but, for high values, the marginal effect of q on the optimal censorship policy vanishes and, therefore, the mechanical component dominates.

7 Constitutional design and liberal stability

We now turn to the problem of the optimal design of a constitution. As we noted, welfare in our model is not necessarily captured by the utility of the pivotal voter. However, we can study what constitution the pivotal voter would choose if she had to pick one within the possibility frontier given by the functions \mathcal{L} and \mathcal{Q} . We focus on the long-run payoff of

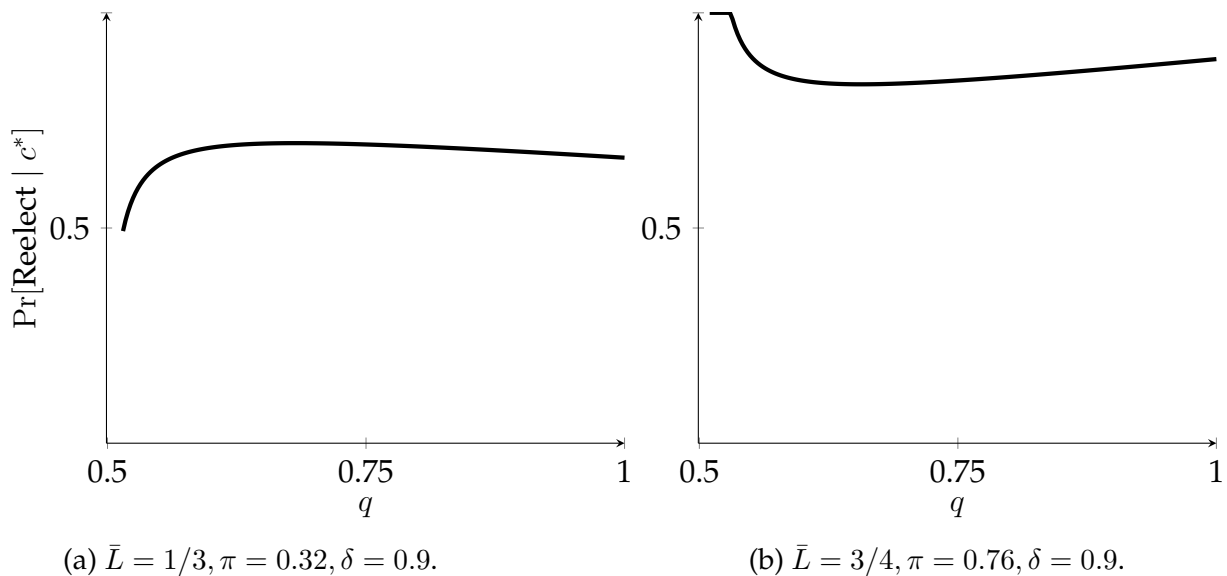


Figure 3: The illiberal government’s reelection probability.

the pivotal voter.³² By choosing a constitution, the voter also selects which of the four regimes in Proposition 1 will be sustained. Proposition 5 says that, unless the technological constraints only afford the voter constitutions that sustain stable illiberal democracies,³³ she always prefers constitutions that sustain either a stable liberal democracy or a cycling democracy.³⁴

Proposition 5 (The voter’s optimal constitution.) *Suppose \mathcal{L} and \mathcal{Q} are such that there exists $\pi \in (0, 1)$ for which the regime is not a stable (efficient or inefficient) illiberal democracy. Then the optimal constitution, π^* , for the voter induces either a liberal democracy or a cycling liberal-illiberal democracy.*

Beyond the pivotal voter’s welfare, the focus of our paper is on which liberal democratic constitutions are stable. One immediate implication of our analysis is that excessive constraints on the executive (high π) necessarily lead to illiberalism. Intuitively, excessive

³²This assumption, formally defined in Definition A.1 in Appendix A, prevents the analysis from being affected by the initial condition $g_0 = \ell$.

³³Essentially, Proposition 5 excludes constitutional possibility frontiers for which, as the constraints on the executive π become small, the extra liberty guaranteed by a liberal government goes to zero faster than the constraints themselves. For all other possibility frontiers, by choosing π the designer can induce regimes of different types.

³⁴Intuitively, whether a stable liberal democracy or a cycling democracy is optimal depends on the exact shapes of \mathcal{L} and \mathcal{Q} . When L tends to grow with π faster than q , then the value of liberal democracy is greater; otherwise, as information becomes more precise, the optimal constitution is one that induces the voter to choose different governments for different signals: a cycling democracy.

constraints render liberal governments ineffective and induce voters to elect illiberal governments that promise greater economic security, efficient public services, and protection from domestic and international violence. This result resonates with the literature that roots the rise of illiberal governments in the ineffectiveness of liberal democratic institutions (Berman, 2019; Matthijs and Blyth, 2018; see also Rodrik, 2000). Howell and Moe (2020) suggest that constitutional reforms that give greater power to the U.S. executive (and the president, in particular) are needed to safeguard the future of American liberal democracy. However, reducing constraints on the executive is not without costs: more constraints on the executive are needed to guarantee greater individual freedoms. This is captured by the constitutional possibility constraint $L = \mathcal{L}(\pi)$. Therefore, our framework allows us to derive implications for the design of the optimal stable liberal constitution: a constitution that maximizes liberty, $\mathcal{L}(\pi)$, without causing instability, i.e., subject to

$$\mathcal{L}(\pi) > \mu_t^*(1, c_\ell)S. \quad (10)$$

Proposition 6 (The optimal stable liberal constitution.) *Suppose \mathcal{L} and \mathcal{Q} are such that there exists $\pi \in (0, 1)$ for which the regime is a stable liberal democracy. The optimal stable liberal constitution, π^* ,*

1. equals 1 if $\mathcal{L}(1) > S$;
2. and otherwise solves

$$\mathcal{L}(\pi^*) = \frac{\pi^* \mathcal{Q}(\pi^*)}{\pi^* \mathcal{Q}(\pi^*) + (1 - \pi^*)(1 - \mathcal{Q}(\pi^*))} S. \quad (11)$$

Intuitively, Proposition 6 says that the optimal stable liberal constitution needs to limit how much liberty it affords to the voter. In fact, more liberty has both a direct and an indirect detrimental effect on liberal stability. More liberty directly requires more constraints on the executive, thus increasing the expected cost of security of electing a liberal government, increasing the voter's propensity to elect an illiberal government. Moreover, because more liberty requires greater constraints on the executive, it also indirectly implies that media and state bureaucracy are more transparent (because \mathcal{Q} is increasing). This increases the precision of the voter's information, increasing her propensity to elect an illiberal government when she observes information that suggests that the liberal government may be unable to adequately respond to this period's shocks.

Point 2 in Proposition 6 also provides some insights into the effects of technological changes that shift the constitutional possibility frontier. For example, technological

changes that affect the quality of information observed by voters and increase $Q(\pi)$ for all π . The right hand side of (11) is increasing in the transparency of liberal institutions, $q = Q(\pi)$. Therefore, the optimal constraints on the executive, π^* , will crucially depend on Q . Proposition 7 says that the arrival of technologies that increase the transparency of liberal democratic institutions must be accompanied by reductions in the constraints on the action of the executive. Otherwise, voters will demand stronger illiberal governments.³⁵

Proposition 7 (The paradox of liberal transparency.) *Suppose \mathcal{L} and Q are such that there exists $\pi \in (0, 1)$ for which the regime is a stable liberal democracy. Let Q' be such that $Q(\pi) < Q'(\pi)$ for all π . Then either there exists no $\pi \in (0, 1)$ for which the regime is a stable liberal democracy under Q' or the optimal stable liberal constitution $\pi_{Q'}^*$ under Q' features weaker executive constraints and less liberty than the optimal stable liberal constitution π_Q^* under Q : $\pi_{Q'}^* \leq \pi_Q^*$ and $\mathcal{L}(\pi_{Q'}^*) \leq \mathcal{L}(\pi_Q^*)$.*

The possibly detrimental effects of greater transparency are most evident when considering the welfare implications of different constitutions. To make ideas precise, suppose there is a continuum of voters, j , indexed by their relative value of liberty, \bar{L}_j , and decisions are made via majority rule. Then the pivotal voter is the voter with median relative value of liberty. Therefore, the single voter in our model corresponds to this median voter. Any increase in transparency, including a marginal increase that moves the constitution from a stable liberal democracy to a cycling liberal-illiberal democracy, benefits the median voter. However, it can be shown that this same marginal increase in transparency harms every voter with a higher relative value of liberty and, therefore, may reduce aggregate welfare.

8 Paths to autocracy and democratic stability

A major concern expressed by political leaders, scholars, and commentators is that illiberal democracies may constitute a first step toward autocracy (see, e.g., [Lüthmann and Lindberg, 2019](#); [Rhodes, 2020](#)). In reality, once an illiberal government rises to power, its ability to operate beyond constitutional constraints may not be restricted to information manipulation. We see two possible paths through which illiberal governments may build an autocracy. A first path is through the manipulation of the liberty-security tradeoff. A second path is through the direct repression of electoral accountability.

³⁵Proposition 7 holds strictly whenever the optimal constitution under Q satisfies (11).

Manipulating security. We extend our model to allow the illiberal government to manipulate the voter's value of security. In particular, in each period that the illiberal government is in power, with probability ζ , it can choose to increase the value of security, S , to S' such that

$$L \leq \mu_t^*(0, c_\ell)S'. \quad (12)$$

For example, the illiberal government can engage in aggressive diplomacy with a foreign power or exacerbate internal conflicts with separatists. From the point of view of the voter, this increases the expected security cost of electing a liberal government which cannot employ illiberal measures to protect the voter from foreign invasions or domestic terrorism. Condition 12 says that the increased value of security is large enough to induce the voter to forever prefer an illiberal government, thus inducing a *de facto* autocracy (see Proposition 1).

Repressing accountability. We extend our model to allow the illiberal government to directly repress electoral accountability, thus seizing power and establish an autocracy. Formally, in each period that the illiberal government is in power, with probability γ , it can choose to overthrow the constitution and seize power. If the government chooses to overthrow the constitution, from that period onward, there is no further election and the illiberal government is in power forever.

In equilibrium, illiberal governments obviously always choose to seize a chance to manipulate security or repress accountability. Therefore, both extensions imply that whenever the voter elects an illiberal government, there is a positive probability that this will induce an autocracy. In the long run, this means that a constitution either induces a stable liberal democracy, or it will result in an autocracy.³⁶ The greater the probability that an illiberal government can build an autocracy (greater ζ or γ), the sooner any non-stable liberal democracy becomes an autocracy. However, Proposition 8 says that a greater probability that an illiberal government can build an autocracy allows for a greater set of constitutions to induce a stable liberal democracy.

Proposition 8 (Long-run democratic stability.) *For $\zeta > 0$ or $\gamma > 0$, as t grows, all constitutions induce either a stable liberal democracy or an autocracy. A constitution induces a stable liberal democracy in the long run if the relative value of liberty \bar{L} is greater than a threshold $\bar{\kappa}'(\pi, q, \delta; \zeta, \gamma)$ which decreases with ζ and γ .*

³⁶However, this may take time and even repeated cycles between liberal and illiberal governments.

Intuitively, when electing an illiberal government comes with a greater risk of a total loss of accountability, the voter is more wary of the potential consequences of such a decision. Therefore, she elects an illiberal government only if the expected cost of security is very large. Thus, the set of constitutions that allow for a stable liberal democracy to emerge is greater.

We can interpret the parameters ζ and γ as measuring the weakness of checks and balances preventing total executive takeover. Thus, our result says that stronger checks and balances, while reducing the speed at which illiberal democracies may turn into autocracies, may actually induce democracies to turn illiberal with greater probability, and in the long run lead to autocracy.

Importantly, two types of long-run regimes may arise in equilibrium. First, some stable liberal democracies may be inefficient. For some constitutions, absent the risk that illiberal governments can build an autocracy, the voter would prefer to cycle between liberal and illiberal governments. However, because illiberal democracies may turn into autocracies, the voter refrains from electing an illiberal government and instead prefers to endure the full cost of security of liberalism. This happens when the liberal constraints on the executive are intermediate.

Second, some long-run autocracies are also inefficient for the median voter. For some constitutions, the voter prefers to cycle between liberal and illiberal governments. However, she chooses illiberalism conscious that this may lead with positive probability to a long-run autocracy, which on average offers too much security and too little liberty.

Informational legacy of illiberalism. Another realistic dynamic effect of illiberalism is that information manipulation may have long-lasting effects on the quality of some institutions of control. For example, a censored bureaucracy may find it hard to rebuild capabilities and experience after an illiberal spell. Similarly, intimidated media may not recover immediately after the fall of an illiberal government. In our model, this amounts to a permanent fall in q following an illiberal government. Notice that this means that the accountability cost of illiberalism is no longer constant: it is large before the voter ever elects an illiberal government, but it is small (in some cases even equal to zero) once the illiberal government rises to power. As a consequence, the voter is more wary of electing an illiberal government, therefore reducing the probability of a rise of an illiberal democracy, but she also is less likely to return to liberalism once she has elected an illiberal government, reducing the chances of a fall of an illiberal democracy.

9 Empirical relevance

The focus of our theory is at the voter’s level. In particular, we offer a framework to think about variation in support for liberalism even in consolidated democracies, which in turn may lead to the rise of an illiberal democracy. Our key idea is that a greater belief that individual welfare is under threat should come with lower support for liberal aspects of democracy but not necessarily with an outright rejection of democracy. We now argue that this premise is consistent with real-world data.

As we noted earlier, [Graham and Svulik \(2020\)](#) and [Svulik \(2019\)](#) show experimentally that voters are willing to tradeoff liberty with security. In the context of the recent pandemic, [Alsan et al. \(2020\)](#) show that voters who are more exposed to the risk of the pandemic are more willing to impose limitations to individual freedoms to mitigate the spread of the virus. In a recent field experiment in Turkey, [Baysan \(2021\)](#) shows that conservative voters who are exposed to information about the recent increase in terrorist activity are more likely to vote in favor of lowering the accountability of the president.

We use data from Wave 6 of the World Values Survey (2010–2014), focusing our attention to respondents in fully consolidated democracies, i.e., countries scoring a 10 out of 10 in the Polity5 democracy score in 2010. We study how different fear factors correlate with support for liberal elements of democratic constitutions. We use two important variables capturing fear: the fear of terrorist attacks and the fear of losing employment (or not finding employment).³⁷ We measure (positive or negative) support for liberal aspects of democratic constitutions via questions regarding support for the protection of civil rights (positive), and for “having a strong leader who does not have to bother with parliament and elections” (negative).

Figure 4 shows that fears of terrorist attacks and losing employment strongly correlate with lower support for the protection of civil rights in democracy and greater support for a strong leader. However, the bottom panels of Figure 4 show that fears of terrorism or of losing employment do not have a similarly clear impact on support for democracy in general. All these results replicate in linear regressions with country fixed effects; all relations have the expected direction and are highly statistically significant, bar for the relation between fear of terrorism and support for democracy. The results also extend to including all democracies (Polity5 score of 6–10) but the effect of fear on civil rights protections is lost, perhaps because the importance of civil rights protection is perceived as more fundamental among citizens of democracies with very marginal scores.³⁸

³⁷The results are qualitatively similar if we use the fear of a “war involving my country.”

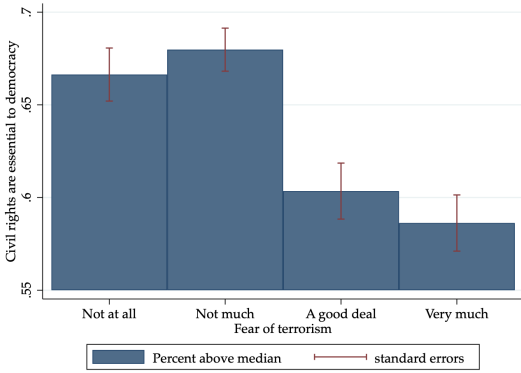
³⁸In addition, the connection between economic security and support for democracy in general is lost; if anything, fear of terrorism seem to foster the support for democracy in general in marginal democracies.

Albeit far from establishing any causal relation, the data support the view that individual fears are connected with lower support of key liberal aspects of democratic institutions. Even in Western consolidated democracies, economic insecurity maps into much stronger preferences for governments that do not need to bother with the procedural aspects of democratic decision-making. These results naturally map into the mechanics of our model. Upon observing information suggesting that individual welfare is under greater threat, a voter's expected cost of security rises, so that she may more likely prefer an illiberal government: one that provides less liberties, such as the protection of civil rights, but provides security through the effective action of a strong and less constrained leadership.

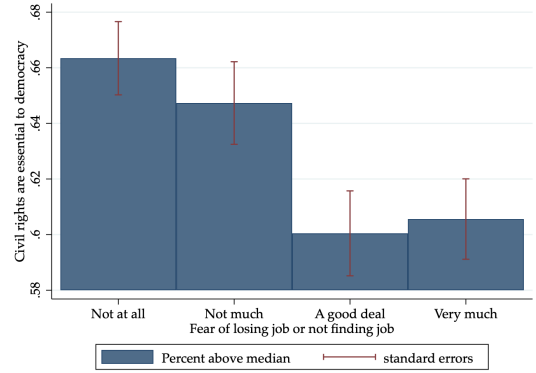
Our theory offers a mechanism through which these individual tendencies towards illiberalism become consequential: if the median voter prefers an illiberal government, she will elect one, leading the country to cycles of illiberalism, a stable illiberal democracy, or even to an autocracy. Survey data may not necessarily represent consequential preferences—perhaps only reflecting the respondents' frustration with the complications of democracy. However, even within our sample of consolidated democracies, country-average support for a strong leadership negatively correlate with the v-Dem measure of liberal democracy in 2010 (p-value: 0.023), suggesting that survey data, at least about the support for strong leaders, capture a substantial phenomenon that affects both voters' preferences and the action of governments.³⁹ Therefore, we use this variable as a measure of voters' support for illiberalism.

A distinctive feature of our model is that too many constraints on the executive aimed at guaranteeing liberty may induce worried voters to prefer an illiberal government. More specifically, our theory predicts that fears of terrorism, war, or economic insecurity should translate into preferences for illiberalism more in countries that have more stringent constraints on the executive. Following [Besley and Persson \(2016\)](#) and [Besley and Mueller \(2018\)](#), we measure executive constraints using a dummy variable which is equal to one when a country in 2010 receives the highest score (7) on this basis on the Polity5 data. In order to observe variance in this measure, we study the sample of all democracies (a score of 6 or more in the Polity5 democracy measure). We report in [Table 1](#) the results of regressing our preferred measure of support for illiberalism (respondents' support for a strong leader) on our measures of fear, the executive constraints dummy, and the interaction term between fear and executive constraints. Consistently with our theory, the effect of fear on preferences for illiberalism are larger—and in fact exist only—in countries that

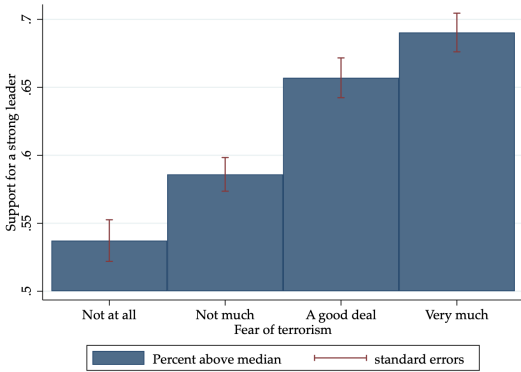
³⁹Similarly, [Skaaning and Krishnarajan \(2021\)](#) show that survey measures of support for free speech is positively correlated with the amount of freedom of expression enjoyed by nations in practice.



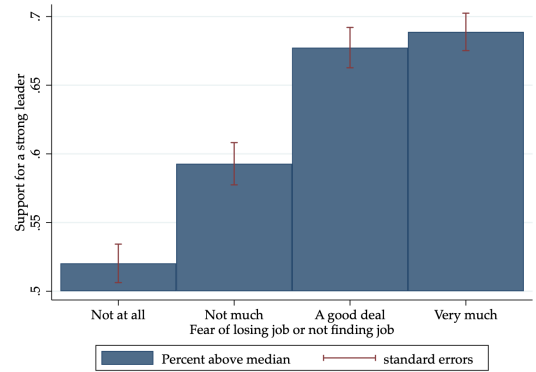
(a) Terrorism and civil rights



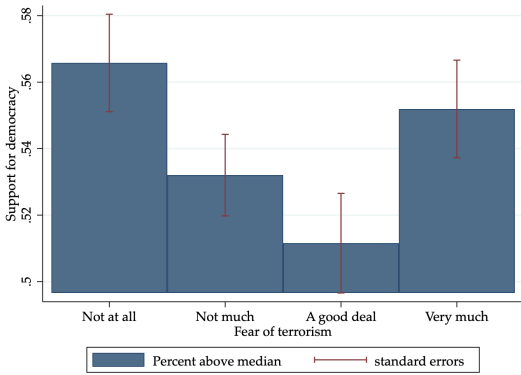
(b) Economic security and civil rights



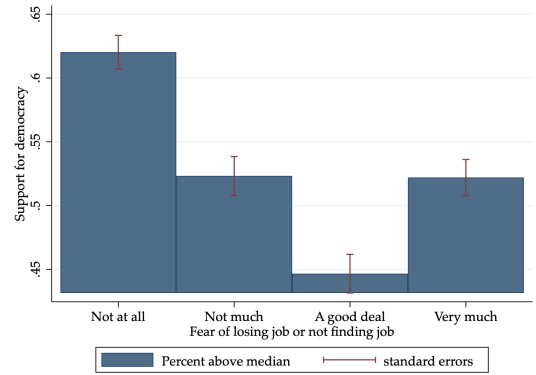
(c) Terrorism and strong leader



(d) Economic security and strong leader



(e) Terrorism and democracy



(f) Economic security and democracy

Percentage of respondents whose answer is weakly greater than the median answer to that question. Panels (a) and (b): Many things are desirable, but not all of them are essential characteristics of democracy: Civil rights protect people from state oppression (scale 1–10). Panels (c) and (d): Would you say it is a very good, fairly good, fairly bad or very bad way of governing this country: Having a strong leader who does not have to bother with parliament and elections? Panels (e) and (f): How important is it for you to live in a country that is governed democratically? (scale 1–10). X-axes: To what degree are you worried about the following situations?: A terrorist attack (left panels); losing my job or not finding a job (right panels).

Figure 4: World Values Survey (2010–2014)

	Support for a strong leader	
	(1)	(2)
Fear of terrorism	-0.0139 (0.0294)	
Fear of losing job		-0.0256 (0.0383)
Executive constraints	-0.630*** (0.169)	-0.678*** (0.167)
Fear of terrorism × Executive constraints	0.154*** (0.0381)	
Fear of losing job × Executive constraints		0.165*** (0.0459)
Observations	45335	43404

Clustered standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 1: Executive constraints and the connection between fear and illiberalism.

have stronger constraints on the executive.

Our key premise that insecurity—and especially economic insecurity—drives voters to support illiberal governments resonates with empirical data. [Mounk \(2018\)](#) argues that the stagnation of median household income since the 1980s is one of the three main causes of rising support for illiberal democracies. [Rothwell and Diego-Rosell \(2016\)](#) and [Kolko \(2016\)](#) document that voters’ support for Trump in the 2016 election did not reflect poor economic conditions; rather it reflected economic vulnerability to shocks in automation, immigration, globalization, and offshoring—in our words, voters who feared for their economic security. Much attention has been given to the recent rise of populist parties in Europe. Some of these populist parties (but not all, as pointed out by [Mudde and Rovira Kaltwasser, 2017](#)) share the key features of our illiberal governments: they promise economic security to the median voter by disrespecting constitutional (and supranational) constraints on the executive. [Guiso et al. \(2019, 2020\)](#) document that the rise of populist parties in Europe resulted from mainstream parties and status quo institutions failing to manage the shocks faced by their economies. In line with our framework that too many constraints on the executive may exacerbate voter insecurity and induce demand for illiberal governments, [Guiso et al. \(2019\)](#) find that “the effects of globalization and the

financial crisis on voting for populist parties in a European country crucially depends on whether or not such a country belongs to the Eurozone.” Belonging to the Eurozone, in fact, increases the institutional constraints on the fiscal action of governments, resulting in frustrated voters who turn to politicians that offer an alternative strategy to increase their economic security.

Our model also offers a framework to make empirical predictions related to the interaction between economic insecurity, information, and culture. In particular, economic insecurity is more (less) likely to spur demand for illiberalism when information about potential crises is more abundant if these crises are sufficiently recurrent (rare). Furthermore, economic insecurity is more likely to spur demand for illiberalism among voters who have a relatively low value of liberty, say because the cautionary tales of the hardship of dictatorship are in a distant past or because the voters inherit a local tradition of authoritarianism (a process of reactivation recently explored within the German context by [Cantoni, Hagemeister and Westcott, 2019](#)).

10 Conclusion

We proposed a theory of the rise and fall of illiberal democracies. We showed that, during times of crisis, voters may choose to elect illiberal governments that, by operating beyond the constraints of a liberal constitution, offer greater security and less liberty but may also manipulate information. We showed that the manipulation (or censoring) of information reduces effective electoral accountability. Therefore, the tradeoff that we highlight for the voters is one between liberty, security, and accountability.

By focusing on the ability of illiberal governments to manipulate information, we drew insights into what makes a liberal constitution stable. First, the constitution must give sufficient powers to the government to address emergencies. In this sense, an ideal liberal constitution that perfectly protects individual liberties and rights is ineffective: it inevitably leads to illiberal governments that systematically infringe those same rights. Importantly, the effect of too many and too stringent constraints on the executive may be exacerbated by excessive transparency. In other words, in a world in which information leaks are unavoidable, a stable liberal democracy requires greater powers in the hands of the executive.

We also derived the conditions under which voters may resort to illiberal governments even when they expect such governments to be able to completely overthrow the checks and balances of democracy and dispense with elections altogether. The key insight from this analysis is that cycles of illiberalism are short-term symptoms of a long-

run tendency to authoritarianism. Our results suggest that stronger checks and balances, while delaying the process of executive takeover, may nonetheless be counterproductive as they induce voters to be excessively optimistic about experimenting with short-term illiberal spells. If illiberal democracies are a step towards autocracy, then the only safety mechanism against them is to avoid them altogether by increasing the effectiveness of the government and possibly balancing executive constraints with transparency.

Our model also suggests a reason for why illiberal governments may be more popular in countries in which the constraints on the executive originate from supranational institutions. For example, voters in Poland, Greece, or Hungary may perceive that their governments are unduly constrained by decisions taken by the European Union or the International Monetary Fund. However, supranational institutions also guarantee stronger checks and balances. In our model, this combination of constrained sovereignty (high π) and strong checks and balances (low γ and ζ) is the perfect storm for the rise of illiberalism.

Our model focuses on the popular *demand* for illiberalism in that it assumes that viable illiberal candidates always exist. Whether such candidates exist may also depend on supply-driven factors such as the political and party system. [Berman and Snegovaya \(2019\)](#) and [Rosenbluth and Shapiro \(2018\)](#) argue the increase in demand for illiberal candidates is explained by recent trends in the design and ideology of political parties and that parties have a role in stemming the supply of illiberal candidates (see also [Benedetto, Hix and Mastrorocco, 2020](#)). [Grillo and Prato \(2021\)](#) show that even leaders who have no preferences for illiberal policies may partially adopt them to gain electoral advantages when voters have reference-dependent preferences. These supply-driven factors may exacerbate the vulnerability of liberal constitutions to changes in voters' demand for illiberalism.

Our results about the role played by transparency in liberal democracies may offer insights into how to interpret the relation between the expansion of internet, 3G networks, and social media and the rise of populism in Europe ([Campante, Durante and Sobbrío, 2018](#); [Gurieva, Melnikov and Zhuravskaya, 2020](#)). In our framework voters always hold consistent beliefs about the probability that their welfare is under threat. In reality voters' beliefs may overreact to information, overestimate small probabilities (for example, the probability of dying in a terrorist attack), and in general be on average inconsistent with reality ([Alesina, Miano and Stantcheva, 2020](#); [Alesina, Stantcheva and Teso, 2018](#)). Therefore, our results are likely to underestimate the probability that the median voter may resort to inefficient illiberal regimes, increasing the potential cost of constitutions that are more exposed to the menace of illiberalism.

References

- Acemoglu, Daron, Giuseppe De Feo, Giacomo De Luca, and Gianluca Russo**, "War, Socialism and the Rise of Fascism: An Empirical Exploration," *NBER Working Paper No. 27854*, 2020.
- Aghion, Philippe, Alberto Alesina, and Francesco Trebbi**, "Endogenous political institutions," *Quarterly Journal of Economics*, 2004, 119 (2), 565–611.
- Alesina, Alberto, Armando Miano, and Stefanie Stantcheva**, "The Polarization of Reality," *AEA Papers and Proceedings*, 2020, 110, 324–28.
- Alesina, Alberto, Stefanie Stantcheva, and Edoardo Teso**, "Intergenerational Mobility and Preferences for Redistribution," *American Economic Review*, 2018, 108 (2), 521–54.
- Alsan, Marcella, Luca Braghieri, Sarah Eichmeyer, Minjeong Joyce Kim, Stefanie Stantcheva, and David Y. Yang**, "Civil Liberties in Times of Crisis," *NBER Working Paper No. 27972*, 2020.
- Anelli, Massimo, Italo Colantone, and Piero Stanig**, "We Were the Robots: Automation and Voting Behavior in Western Europe," *IZA Discussion Paper No. 12485*, 2019.
- Autor, David, David Dorn, Gordon Hanson, and Kaveh Majlesi**, "Importing Political Polarization? The Electoral Consequences of Rising Trade Exposure," *American Economic Review*, 2020, 110 (10), 3139–3183.
- Baysan, Ceren**, "Persistent Polarizing Effects of Persuasion: Experimental Evidence from Turkey," *Working paper*, 2021.
- Benedetto, Giacomo, Simon Hix, and Nicola Mastrococco**, "The Rise and Fall of Social Democracy, 1918-2017," *American Political Science Review*, 2020, 114 (3), 928–939.
- Berman, Sheri**, "Populism is a Symptom Rather than a Cause: Democratic Disconnect, the Decline of the Center-Left, and the Rise of Populism in Western Europe," *Polity*, 2019, 51 (4), 654–667.
- Berman, Sheri and Maria Snegovaya**, "Populism and the Decline of Social Democracy," *Journal of Democracy*, 2019, 30 (3), 5–19.
- Bernhardt, Dan, Stefan Krasa, and Mehdi Shadmehr**, "Demagogues and the Fragility of Democracy," *UIUC Working Paper*, 2019.
- Besley, Timothy and Hannes Mueller**, "Cohesive institutions and the distribution of political rents: Theory and evidence," in "Institutions, Governance and the Control of Corruption," Springer, 2018, pp. 165–208.
- Besley, Timothy and Torsten Persson**, "Democratic Values and Institutions," *London School of Economics and Political Science Working Paper*, 2016.
- Besley, Timothy and Torsten Persson**, "Democratic Values and Institutions," *American Economic Review: Insights*, 2019, 1 (1), 59–76.
- Bidner, Chris, Patrick François, and Francesco Trebbi**, "A theory of minimalist democracy," *NBER Working Paper No. 20552*, 2014.
- Bizzotto, Jacopo, Jesper Rüdiger, and Adrien Vigier**, "Dynamic persuasion with outside information," *American Economic Journal: Microeconomics*, 2021, 13 (1), 179–94.
- Campante, Filipe, Ruben Durante, and Francesco Sobbrío**, "Politics 2.0: The multifaceted effect of broadband internet on political participation," *Journal of the European Economic Association*, 2018, 16 (4), 1094–1136.

- Cantoni, Davide, Felix Hagemeister, and Mark Westcott**, "Persistence and activation of right-wing political ideology," *LMU Munich Working Paper*, 2019.
- Che, Yeon-Koo, Kyungmin Kim, and Konrad Mierendorff**, "Keeping the listener engaged: a dynamic model of bayesian persuasion," *arXiv preprint arXiv:2003.07338*, 2020.
- Cheibub, José Antonio, Ji Yeon Jean Hong, and Adam Przeworski**, "Rights and Deaths: Government Reactions to the Pandemic," *SSRN Working Paper 3645410*, 2020.
- Chiopris, Caterina, Monika Nalepa, and Georg Vanberg**, "A Wolf in Sheep's Clothing: Citizen Uncertainty and Democratic Backsliding," *University of Chicago Working Paper*, 2021.
- Colantone, Italo and Piero Stanig**, "The Trade Origins of Economic Nationalism: Import Competition and Voting Behavior in Western Europe," *American Journal of Political Science*, 2018, 62 (4), 936–953.
- Colantone, Italo and Piero Stanig**, "The Surge of Economic Nationalism in Western Europe," *Journal of Economic Perspectives*, 2019, 33 (4), 128–51.
- Collier, Paul**, "The Political Economy of Natural Resources," *Social Research*, 2010, 77 (4), 1105–1132.
- Coppedge, Michael, John Gerring, Carl Henrik Knutsen, Staffan I. Lindberg, Jan Teorell, David Altman, Michael Bernhard, M. Steven Fish, Adam Glynn, Allen Hicken, Anna Luhrmann, Kyle L. Marquardt, Kelly McMann, Pamela Paxton, Daniel Pemstein, Brigitte Seim, Rachel Sigman, Svend-Erik Skaaning, Jeffrey Staton, Steven Wilson, Agnes Cornell, Nazifa Alizada, Lisa Gastaldi, Haakon Gjerløw, Garry Hindle, Nina Ilchenko, Laura Maxwell, Valeriya Mechkova, Juraj Medzihorsky, Johannes von Römer, Aksel Sundström, Eitan Tzelgov, Yi-ting Wang, Tore Wig, and Daniel Ziblatt**, *V-Dem [Country–Year/Country–Date] Dataset v10. Varieties of Democracy (V-Dem) Project*. <https://doi.org/10.23696/vdemds20>. 2020.
- Edmond, Chris**, "Information Manipulation, Coordination, and Regime Change," *Review of Economic Studies*, 2013, 80 (4), 1422–1458.
- Egorov, Georgy, Sergei Guriev, and Konstantin Sonin**, "Why resource-poor dictators allow freer media: A theory and evidence from panel data," *American political science Review*, 2009, 103 (4), 645–668.
- Ely, Jeffrey, Alexander Frankel, and Emir Kamenica**, "Suspense and surprise," *Journal of Political Economy*, 2015, 123 (1), 215–260.
- Ely, Jeffrey C**, "Beeps," *American Economic Review*, 2017, 107 (1), 31–53.
- Ely, Jeffrey C and Martin Szydlowski**, "Moving the goalposts," *Journal of Political Economy*, 2020, 128 (2), 468–506.
- Fearon, James D.**, "Self-enforcing Democracy," *Quarterly Journal of Economics*, 2011, 126 (4), 1661–1708.
- Felshitsky, Yuri and Vladimir Pribylovsky**, *The Age of Assassins: The Rise and Rise of Vladimir Putin*, Gibson Square Books, 2008.
- Foa, Roberto Stefan and Yascha Mounk**, "The Danger of Deconsolidation: The Democratic Disconnect," *Journal of Democracy*, 2016, 27 (3), 5–17.
- Foa, Roberto Stefan and Yascha Mounk**, "The Signs of Deconsolidation," *Journal of Democracy*, 2017, 28 (1), 5–15.

- Fudenberg, Drew and Jean Tirole**, "Perfect Bayesian Equilibrium and Sequential Equilibrium," *Journal of Economic Theory*, 1991, 53 (2), 236–260.
- Gehlbach, Scott and Konstantin Sonin**, "Government Control of the Media," *Journal of Public Economics*, 2014, 118, 163–171.
- Graham, Matthew H. and Milan W. Svobik**, "Democracy in America? Partisanship, Polarization, and the Robustness of Support for Democracy in the United States," *American Political Science Review*, 2020, 114 (2), 392–409.
- Gratton, Gabriele and Massimo Morelli**, "Optimal Checks and Balances Under Policy Uncertainty," *Forthcoming in International Economic Review*, 2021.
- Grillo, Edoardo and Carlo Prato**, "Reference Points and Democratic Backsliding," *Forthcoming in American Journal of Political Science*, 2021.
- Guiso, Luigi, Helios Herrera, Massimo Morelli, and Tommaso Sonno**, "Global Crises and Populism: The Role of Eurozone Institutions," *Economic Policy*, 2019, 34 (97), 95–139.
- Guiso, Luigi, Helios Herrera, Massimo Morelli, and Tommaso Sonno**, "Economic Insecurity and the Demand of Populism in Europe," *Einaudi Institute for Economics and Finance Working Paper*, 2020.
- Guriev, Sergei and Daniel Treisman**, "Informational Autocrats," *Journal of Economic Perspectives*, 2019, 33 (4), 100–127.
- Guriev, Sergei and Daniel Treisman**, "The Popularity of Authoritarian Leaders: A Cross-national Investigation," *World Politics*, 2020, 72 (4), 601–638.
- Guriev, Sergei and Daniel Treisman**, "A Theory of Informational Autocracy," *Journal of Public Economics*, 2020, 186, 104158.
- Guriev, Sergei and Elias Papaioannou**, "The Political Economy of Populism," *CEPR Discussion Paper No. DP14433*, 2020.
- Guriev, Sergei, Nikita Melnikov, and Ekaterina Zhuravskaya**, "3G Internet and Confidence in Government," *CEPR Discussion paper DP14022-8*, 2020.
- Hollyer, James R., B. Peter Rosendorff, and James Raymond Vreeland**, "Measuring Transparency," *Political Analysis*, 2014, 22 (4), 413–434.
- Hollyer, James R., B. Peter Rosendorff, and James Raymond Vreeland**, "Transparency, Protest, and Autocratic Instability," *American Political Science Review*, 2015, 109 (4), 764–784.
- Howell, William G. and Terry M. Moe**, *Presidents, Populism, and the Crisis of Democracy*, University of Chicago Press, 2020.
- Howell, William G., Kenneth Shepsle, and Stephane Wolton**, "Executive Absolutism: A Model," *SSRN Working Paper 3440604*, 2019.
- Huq, Aziz and Tom Ginsburg**, "How to Lose a Constitutional Democracy," *UCLA Law Review*, 2018, 65, 78.
- Inglehart, R., C. Haerpfer, A. Moreno, C. Welzel, K. Kizilova, J. Diez-Medrano, M. Lagos, P. Norris, E. Ponarin, B. Puranen et al., ed.**, *World Values Survey: Round Six - Country-Pooled Datafile 2010-2014. Version: <http://www.worldvaluessurvey.org/WVSDocumentationWV6.jsp>*, Madrid: JD Systems Institute, 2014.

- Jones, Benjamin F. and Benjamin A. Olken**, “Do Leaders Matter? National Leadership and Growth Since World War II,” *Quarterly Journal of Economics*, 2005, 120 (3), 835–864.
- Kamenica, Emir and Matthew Gentzkow**, “Bayesian Persuasion,” *American Economic Review*, 2011, 101 (6), 2590–2615.
- Kolko, Jed**, “Trump was stronger where the economy is weaker,” *FiveThirtyEight*, 2016, (November 10).
- Kolotilin, Anton, Tymofiy Mylovanov, and Andriy Zapechelnyuk**, “Censorship as Optimal Persuasion,” *Forthcoming in Theoretical Economics*, 2019.
- Li, Anqi, Davin Raiha, and Ken Shotts**, “Propaganda, Alternative Media, and Accountability in Fragile Democracies,” *Stanford GSB, Working paper No. 3832*, 2020.
- Little, Andrew T, Joshua A Tucker, and Tom LaGatta**, “Elections, protest, and alternation of power,” *Journal of Politics*, 2015, 77 (4), 1142–1156.
- Lührmann, Anna and Staffan I. Lindberg**, “A Third Wave of Autocratization is Here: What is New About It?,” *Democratization*, 2019, 26 (7), 1095–1113.
- Luo, Zhaotian and Adam Przeworski**, “Democracy and Its Vulnerabilities: Dynamics of Democratic Backsliding,” *SSRN Working Paper 3469373*, 2019.
- Luo, Zhaotian and Arturas Rozenas**, “Ruling the Ruling Coalition: Information Control and Authoritarian Power-Sharing,” *SSRN Working Paper 2978254*, 2019.
- Marshall, Monty G. and Ted Robert Gurr**, *Polity5: Political Regime Characteristics and Transitions, 1800–2018*, Polity5 Project, Center for Systemic Peace, 2020.
- Matthijs, Matthias and Mark Blyth**, “When is it rational to learn the wrong lessons? Technocratic authority, social learning, and euro fragility,” *Perspectives on Politics*, 2018, 16 (1), 110–126.
- Montesquieu, Charles de Secondat Baron de**, *The Spirit of Laws*, T. Nugent. New York: Hafner, 1949 [1750].
- Mouk, Yascha**, *The People vs. Democracy: Why Our Freedom Is in Danger and How to Save It*, Harvard University Press, 2018.
- Mudde, Cas and Cristóbal Rovira Kaltwasser**, *Populism: A Very Short Introduction*, Oxford University Press, 2017.
- Mukand, Sharun W and Dani Rodrik**, “The political economy of liberal democracy,” *Economic Journal*, 2020, 130 (627), 765–792.
- Orbán, Viktor**, “Prime Minister Viktor Orbán’s Speech at the 25th Bálványos Summer Free University and Student Camp,” Prime Minister of Hungary Office, <https://www.kormany.hu/en/the-prime-minister/the-prime-minister-s-speeches/prime-minister-viktor-orban-s-speech-at-the-25th-balvanyos-summer-free-university-and-student-camp>, Retrieved 13 September 2020. July 2014.
- Orlov, Dmitry, Andrzej Skrzypacz, and Pavel Zryumov**, “Persuading the principal to wait,” *Journal of Political Economy*, 2020, 128 (7), 2542–2578.
- Pishro-Nik, Hossein**, *Introduction to probability, statistics, and random processes*, Kappa Research LLC (available at <https://www.probabilitycourse.com>), 2014.

- Posner, Eric A. and Adrian Vermeule**, *The Executive Unbound: After the Madisonian Republic*, Oxford Scholarship Online, 2011.
- Renault, Jérôme, Eilon Solan, and Nicolas Vieille**, "Optimal dynamic information provision," *Games and Economic Behavior*, 2017, 104, 329–349.
- Rhodes, Ben**, "The Path to Autocracy," *The Atlantic*, 2020, (June 15).
- Rodrik, Dani**, "How Far Will International Economic Integration Go?," *Journal of Economic Perspectives*, 2000, 14 (1), 177–186.
- Rodrik, Dani**, "Is populism necessarily bad economics?," in "AEA Papers and proceedings," Vol. 108 2018, pp. 196–99.
- Rosenbluth, Frances and Ian Shapiro**, *Responsible Parties: Saving Democracy from Itself*, Yale University Press, 2018.
- Rothwell, Jonathan T and Pablo Diego-Rosell**, "Explaining nationalist political views: The case of Donald Trump," *Available at SSRN 2822059*, 2016.
- Sachs, Jeffrey D.**, "Social Conflict and Populist Policies in Latin America," Technical Report, National Bureau of Economic Research 1989.
- Samuelson, Paul A.**, *Economics: An Introductory Analysis*, 5th ed., McGraw Hill: New York, 1961.
- Satter, David**, *Darkness at Dawn: The Rise of the Criminal Russian State*, New Haven: Yale Univ. Press, 2003.
- Sgueo, Gianluca**, "The Influence of Transparency on Addressing Populism in the EU," *SSRN Working Paper 3137099*, 2018.
- Shadmehr, Mehdi and Dan Bernhardt**, "State censorship," *American Economic Journal: Microeconomics*, 2015, 7 (2), 280–307.
- Skaaning, Svend-Erik and Suthan Krishnarajan**, "Who Cares about Free Speech? Findings from a Global Survey of Support for Free Speech," *Justitia Report*, 2021.
- Stasavage, David**, *The Decline and Rise of Democracy: A Global History from Antiquity to Today*, Vol. 96, Princeton University Press, 2020.
- Stasavage, David**, "Democracy, Autocracy, and Emergency Threats: Lessons for COVID-19 from the Last Thousand Years," *International Organization*, 2020, pp. 1–17.
- Svolik, Milan W.**, "Polarization versus Democracy," *Journal of Democracy*, 2019, 30 (3), 20–32.
- Svolik, Milan W.**, "When Polarization Trumps Civic Virtue: Partisan Conflict and the Subversion of Democracy by Incumbents," *Quarterly Journal of Political Science*, 2020, 15 (1), 3–31.
- Szeidl, Adam and Ferenc Szucs**, "Media Capture Through Favor Exchange," *Econometrica*, 2021, 89 (1), 281–310.
- Wittes, Benjamin**, "Against a Crude Balance: Platform Security and the Hostile Symbiosis Between Liberty and Security," *The Brookings Institution and Harvard Law School: Project on Law and Security*, 2011.
- Zakaria, Fareed**, "The rise of illiberal democracy," *Foreign Affairs*, 1997, 76 (6), 22–43.

Appendix: For online publication

A Omitted proofs

Proof of Lemma 1. By Definition 1, in any equilibrium, for any period t , message m_t , censorship policy c_t , and continuation strategies σ^* , the voter elects an illiberal government if and only if

$$\begin{aligned} & \mu_t(m_t, c_t)S \\ & \geq L + \delta \left[\mathbb{E} \left[\sum_{\tilde{t}=t+1}^{\infty} \delta^{(\tilde{t}-t-1)} v(g_{\tilde{t}}, \theta_{\tilde{t}}) \mid g_t = \ell, \sigma^* \right] - \mathbb{E} \left[\sum_{\tilde{t}=t+1}^{\infty} \delta^{(\tilde{t}-t-1)} v(g_{\tilde{t}}, \theta_{\tilde{t}}) \mid g_t = i, \sigma^* \right] \right]. \end{aligned} \quad (\text{A.1})$$

Substituting from (5) yields $\mu_t(m_t, c_t)S \geq L + \delta A(\pi, q, L, S, \delta \mid \sigma^*)$. ■

Proof of Lemma 2. By Definition 1, in any equilibrium, in period t an illiberal government's optimal censorship policy $c_{t+1} = c^*$ is such that: for all censorship policies c' and continuation strategies σ^* (using the law of iterated expectations and noting that period- t deviations do not affect continuation payoffs after period $t + 1$)

$$\begin{aligned} & \delta R \Pr[g_{t+1} = i \mid c_{t+1} = c^*] + \mathbb{E} \left[\sum_{\tilde{t}=t+2}^{\infty} \delta^{\tilde{t}-t} R \mathcal{X}(\{g_{\tilde{t}}\}_{\tilde{t}+1}) \mid \sigma^* \right] \\ & \geq \delta R \Pr[g_{t+1} = i \mid c_{t+1} = c'] + \mathbb{E} \left[\sum_{\tilde{t}=t+2}^{\infty} \delta^{\tilde{t}-t} R \mathcal{X}(\{g_{\tilde{t}}\}_{\tilde{t}+1}) \mid \sigma^* \right] \\ & \iff \Pr[g_{t+1} = i \mid c_{t+1} = c^*] \geq \Pr[g_{t+1} = i \mid c_{t+1} = c']. \end{aligned}$$

■

Proof of Lemma 3. We begin by proving that $\bar{A}(\pi, q, L, S, \delta)$ (as defined in (6)) is indeed the maximum accountability cost of illiberalism.

Lemma A.1 *In equilibrium, $0 \leq A(\pi, q, L, S, \delta \mid \sigma^*)$, and either the illiberal government is indifferent between all censorship policies or*

$$A(\pi, q, L, S, \delta \mid \sigma^*) \leq \bar{A}(\pi, q, L, S, \delta). \quad (\text{A.2})$$

Proof. We begin by proving that $0 \leq A(\pi, q, L, S, \delta \mid \sigma^*)$. By Definition 1, for any censor-

ship policy c_t and equilibrium assessment σ^* , we have that

$$\begin{aligned}
V(g_t = i \mid \sigma^*) &= \sum_{m'=0,1} \Pr[m_t = m' \mid c_t] \max\{L - \mu_t^*(m', c_t)S + \delta V(g_t = \ell \mid \sigma^*), \delta V(g_t = i \mid \sigma^*)\} \\
&= \sum_{m'=0,1} \sum_{s'=0,1} \Pr[m_t = m', s(\theta_t) = s' \mid c_t] \max\{L - \mu_t^*(m', c_t)S + \delta V(g_t = \ell \mid \sigma^*), \delta V(g_t = i \mid \sigma^*)\} \\
&\leq \sum_{m'=0,1} \sum_{s'=0,1} \Pr[m_t = m', s(\theta_t) = s' \mid c_t] \max\{L - \mu_t^*(s', c_\ell)S + \delta V(g_t = \ell \mid \sigma^*), \delta V(g_t = i \mid \sigma^*)\} \\
&= \sum_{s'=0,1} \Pr[s(\theta_t) = s' \mid c_t] \max\{L - \mu_t^*(s', c_\ell)S + \delta V(g_t = \ell \mid \sigma^*), \delta V(g_t = i \mid \sigma^*)\} \\
&= V(g_t = \ell \mid \sigma^*),
\end{aligned}$$

where the inequality follows because any censorship policy is (weakly) less informative than the fully revealing censorship policy, and the final equality follows from the law of iterated expectations.

We now prove the upper bound $A(\pi, q, L, S, \delta \mid \sigma^*) \leq \bar{A}(\pi, q, L, S, \delta)$ when the illiberal government is not indifferent between all censorship policies. By Lemma 2, there exist a censorship policy $c' \neq c^*$ such that

$$\Pr[g_{t+1} = i \mid c_{t+1} = c^*] > \Pr[g_{t+1} = i \mid c_{t+1} = c'].$$

Therefore, there exists message m' that sent with positive probability by c' such that the voter elects a liberal government: $\mu_t^*(m', c')S < L + \delta A(\pi, q, L, S, \delta \mid \sigma^*)$. Similarly, there exists message m^* sent with positive probability by c^* such that the voter elects an illiberal government: $\mu_t^*(m^*, c^*)S \geq L + \delta A(\pi, q, L, S, \delta \mid \sigma^*)$. Note that for all censorship policies, c , and messages, m ,

$$\mu_t^*(0, c_\ell) \leq \mu_t^*(m, c) \leq \mu_t^*(1, c_\ell), \tag{A.3}$$

therefore

$$\mu_t^*(0, c_\ell)S < L + \delta A(\pi, q, L, S, \delta \mid \sigma^*) \leq \mu_t^*(1, c_\ell)S,$$

i.e., whenever $g_{t-1} = \ell$ (and hence $c_t = c_\ell$), the voter elects the illiberal government if and

only if $m_t = 1$. It follows that

$$\begin{aligned}
V(g_t = \ell \mid \sigma^*) &= \Pr[m_t = 0 \mid c_\ell] \left(L - \mu_t^*(0, c_\ell)S + \delta V(g_{t+1} = \ell \mid \sigma^*) \right) \\
&\quad + \Pr[m_t = 1 \mid c_\ell] \delta V(g_{t+1} = i \mid \sigma^*) \\
&= \Pr[s(\theta_t) = 0] \left(L - \Pr[\theta_t = 1 \mid s(\theta_t) = 0]S + \delta V(g_{t+1} = \ell \mid \sigma^*) \right) \\
&\quad + \Pr[s(\theta_t) = 1] \delta V(g_{t+1} = i \mid \sigma^*). \tag{A.4}
\end{aligned}$$

Because $V(g_t = i \mid \sigma^*) = V(g_{t+1} = i \mid \sigma^*)$ and $V(g_t = \ell \mid \sigma^*) = V(g_{t+1} = \ell \mid \sigma^*)$, we can recursively substitute (A.4) into (5):

$$\begin{aligned}
A(\pi, q, L, S, \delta \mid \sigma^*) &= \Pr[s(\theta_t) = 0] \left(L - \Pr[\theta_t = 1 \mid s(\theta_t) = 0]S + \delta V(g_t = \ell \mid \sigma^*) \right) \\
&\quad - V(g_t = i \mid \sigma^*) \left[1 - \delta \Pr[s(\theta_t) = 1] \right] \\
&= \Pr[s(\theta_t) = 0] \left(L - \Pr[\theta_t = 1 \mid s(\theta_t) = 0]S \right) \sum_{\tilde{t}=t}^{\infty} \delta^{\tilde{t}-t} \Pr[s(\theta_{\tilde{t}}) = 0]^{\tilde{t}-t} \\
&\quad - V(g_t = i \mid \sigma^*) \left[1 - \frac{\delta \Pr[s(\theta_t) = 1]}{1 - \delta \Pr[s(\theta_t) = 0]} \right] \\
&= \Pr[s(\theta_t) = 0] \left(L - \Pr[\theta_t = 1 \mid s(\theta_t) = 0]S \right) \frac{1}{1 - \delta \Pr[s(\theta_t) = 0]} \\
&\quad - V(g_t = i \mid \sigma^*) \left[\frac{1 - \delta}{1 - \delta \Pr[s(\theta_t) = 0]} \right] \tag{A.5}
\end{aligned}$$

$$\leq \bar{A}(\pi, q, L, S, \delta) - V(g_t = i \mid \sigma^*) \left[\frac{1 - \delta}{1 - \delta \Pr[s(\theta_t) = 0]} \right]. \tag{A.6}$$

Therefore, if $V(g_t = i \mid \sigma^*) \geq 0$, then the proof is complete. To see this end, we consider two exhaustive cases. First, suppose that, in equilibrium, the illiberal government is able to guarantee its reelection. In this case, it is immediate that $V(g_t = i \mid \sigma^*) = 0$. Second, suppose otherwise. In this case,

$$\begin{aligned}
V(g_t = i \mid \sigma^*) &= \Pr[m_t = 0 \mid c^*] \left(L - \mu_t^*(0, c^*)S + \delta V(g_{t+1} = \ell \mid \sigma^*) \right) \\
&\quad + \Pr[m_t = 1 \mid c^*] \delta V(g_{t+1} = i \mid \sigma^*).
\end{aligned}$$

However, in equilibrium the voter elects the liberal government if and only if

$$L - \mu_t^*(0, c^*)S + \delta V(g_{t+1} = \ell \mid \sigma^*) > \delta V(g_{t+1} = i \mid \sigma^*).$$

Therefore,

$$\begin{aligned}
V(g_t = i \mid \sigma^*) &> \Pr[m_t = 0 \mid c^*] \delta V(g_{t+1} = i \mid \sigma^*) + \Pr[m_t = 1 \mid c^*] \delta V(g_{t+1} = i \mid \sigma^*) \\
&= \delta V(g_{t+1} = i \mid \sigma^*) \\
&= \delta V(g_t = i \mid \sigma^*)
\end{aligned}$$

$$\iff V(g_t = i \mid \sigma^*) > 0. \quad \blacksquare$$

We now prove Point 1 of Lemma 3. Suppose that, in equilibrium, the illiberal government is not indifferent between all censorship policies, and $\pi S \geq L + \delta \bar{A}(\pi, q, L, S, \delta) \geq L + \delta A(\pi, q, L, S, \delta \mid \sigma^*)$, where the last inequality follows from Lemma A.1. Therefore, any censorship policy such that

$$\mu_t^*(m_t, c^*) S \geq L + \delta \bar{A}(\pi, q, L, S, \delta) \quad \forall m_t \quad (\text{A.7})$$

guarantees reelection and is optimal for the illiberal government; because

$$\mu_t^*(m_t, c_F) \equiv \Pr[\theta_t = 1 \mid m_t, c_F] = \pi \quad \forall m_t,$$

this includes the full censorship policy, c_F . We now show that any censorship policy that violates (A.7) is not optimal for the illiberal government. By contradiction, suppose that there is a censorship policy c' that does not satisfy (A.7) but guarantees the illiberal government's reelection. Then there exists m such that

$$L + \delta \bar{A}(\pi, q, L, S, \delta) > \mu_t^*(m, c') S \geq L + \delta A(\pi, q, L, S, \delta \mid \sigma^*), \quad (\text{A.8})$$

and $V(g_t = i \mid \sigma^*) = 0$. Furthermore, by (A.6)

$$A(\pi, q, L, S, \delta \mid \sigma^*) = \Pr[s(\theta_t) = 0] \left(L - \Pr[\theta_t = 1 \mid s(\theta_t) = 0] S \right) \frac{1}{1 - \delta \Pr[s(\theta_t) = 0]}.$$

Since, by Lemma A.1, $A(\pi, q, L, S, \delta \mid \sigma^*) \geq 0$ (and therefore $L - \Pr[\theta_t = 1 \mid s(\theta_t) = 0] S < 0$), then $A(\pi, q, L, S, \delta \mid \sigma^*) = \bar{A}(\pi, q, L, S, \delta)$, which contradicts (A.8).

We now prove Point 2 of Lemma 3. Suppose that, in equilibrium, the illiberal government is not indifferent between all censorship policies, and

$$\pi S < L + \delta \bar{A}(\pi, q, L, S, \delta). \quad (\text{A.9})$$

As was shown in the proof of Lemma A.1, this implies that if $c_t = c_\ell$, then the illiberal

government is elected if and only if $m_t = 1$. Hence, in equilibrium, the illiberal government must be reelected with positive probability that is no less than $\Pr[m_t = 1 \mid c_\ell] = \Pr[s(\theta_t) = 1]$.

We now show that the illiberal government cannot guarantee its reelection with any censorship policy. By contradiction, suppose that the illiberal government can guarantee its reelection by choosing some censorship policy c^* . Then, as shown in proving Point 1, $A(\pi, q, L, S, \delta \mid \sigma^*) = \bar{A}(\pi, q, L, S, \delta)$. Therefore, it must be that $\mu_t^*(m_t, c^*)S \geq L + \delta\bar{A}(\pi, q, L, S, \delta)$ for all m_t , and, hence, in equilibrium,

$$\begin{aligned} \sum_{m=0,1} \Pr[m_t = m \mid c^*] \mu_t^*(m, c^*)S &\geq \sum_{m=0,1} \Pr[m_t = m \mid c^*] \left(L + \delta\bar{A}(\pi, q, L, S, \delta) \right) \\ \iff \Pr[\theta_t = 1]S = \pi S &\geq L + \delta\bar{A}(\pi, q, L, S, \delta), \end{aligned}$$

which contradicts (A.9).

So far we have shown that the illiberal government is reelected with probability strictly between 0 and 1, which also implies that the full censorship policy, c_F , is not chosen in equilibrium: $c^*(0) < c^*(1)$. Therefore,

$$\mu_t^*(0, c^*) < \pi < \mu_t^*(1, c^*).$$

By (A.2) and (A.9), the voter does not elect the illiberal government upon observing $m_t = 0$. Therefore, the illiberal government's problem is

$$\begin{aligned} \max_{c : c(0) < c(1)} \{ &c(1) \Pr[s(\theta_t) = 1] + c(0) \Pr[s(\theta_t) = 0] \} \\ \text{s.t.} \quad &\mu_t^*(1, c)S \geq L + \delta A(\pi, q, L, S, \delta \mid \sigma^*) \end{aligned}$$

Notice that $\mu_t^*(1, c)$ is strictly increasing in $c(1)$ and strictly decreasing in $c(0)$. Therefore, (i) $c^*(1) = 1$ and (ii) the constraint is binding. ■

For a given set of parameters, our model allows for a multiplicity of equilibria. However, the equilibrium is essentially unique in the following sense. Two equilibria are *essentially equivalent* if, for any t and any g_{t-1} , the probability that an illiberal government is elected is equal in each equilibrium (and hence the voter's and illiberal government's expected payoffs are also equal). For any given set of parameters, we say that the equilibrium is *essentially unique* if all equilibria are essentially equivalent.

In Appendix B we prove the following lemma.

Lemma A.2 (Essentially unique equilibrium.) *For any given set of parameters, the equilibrium is essentially unique.*

We can now turn to the proof of our main result.

Proof of Proposition 1.

Part 1. Suppose $L \leq \mu_t^*(0, c_\ell)S$. Because the equilibrium is essentially unique (Lemma A.2), it suffices to show that $\sigma^* = (g^*, c^*, \{\mu_t^*\}_{t=1}^\infty)$ such that

$$g^*(m_t, c_t) = i \iff \mu_t^*(m_t, c_t)S \geq L, \quad (\text{A.10})$$

and c^* is any censorship policy is an equilibrium and that it induces $g_t = i$ for all $t \geq 1$.

That σ^* induces $g_t = i$ for all $t \geq 1$ follows by Lemma 1 because, for all censorship policies c_t and any message m'_t that is sent with positive probability under c_t ,

$$\mu_t^*(m'_t, c_t)S \geq \mu_t^*(0, c_\ell)S \geq L. \quad (\text{A.11})$$

We now show that σ^* is an equilibrium. Because $g_t = i$ for all t , $A(\pi, q, L, S, \delta \mid \sigma^*) = 0$, then Lemma 1 implies that g^* is the voter's equilibrium strategy. Sequential rationality for the illiberal government is trivially satisfied because the illiberal government is indifferent between all censorship policies (see (A.11)). Finally, the voter's beliefs were constructed via Bayes' rule.

Part 2. Suppose

$$\mu_t^*(0, c_\ell)S < L \quad \text{and} \quad L \leq \pi S - \delta \bar{A}(\pi, q, L, S, \delta). \quad (\text{A.12})$$

Because the equilibrium is essentially unique (Lemma A.2), it suffices to show that $\sigma^* = (g^*, c^*, \{\mu_t^*\}_{t=1}^\infty)$ such that

$$g^*(m_t, c_t) = i \iff \mu_t^*(m_t, c_t)S \geq L + \delta \bar{A}(\pi, q, L, S, \delta), \quad (\text{A.13})$$

and

$$\mu_t^*(m_t, c^*)S \geq L + \delta \bar{A}(\pi, q, L, S, \delta) \quad \forall m_t, \quad (\text{A.14})$$

is an equilibrium and

1. under a liberal government ($g_{t-1} = \ell$ and hence $c_t = c_\ell$), the voter elects an illiberal government if and only if $m_t = 1$; and

2. once an illiberal government has been elected ($g_{t-1} = i$) and chooses c^* , the voter reelects the illiberal government with probability one.

Starting with Point 1, by (A.12) and because $\bar{A}(\pi, q, L, S, \delta) \geq 0$ by (6), we have

$$\Pr[\theta_t = 1 \mid m_t = 0, c_\ell]S < L + \delta \bar{A}(\pi, q, L, S, \delta) \leq \pi S.$$

Furthermore, $\pi < \Pr[\theta_t = 1 \mid m_t = 1, c_\ell]$. Therefore, g^* implies that Point 1 holds.

Turning to Point 2. Combining (A.13) and (A.14), it is immediate that the illiberal government is reelected with probability one in equilibrium.

We now show that σ^* is an equilibrium. Because of (A.12) and Point 1 and 2, the accountability cost of illiberalism, $A(\pi, q, L, S, \delta \mid \sigma^*)$, is maximal:

$$\begin{aligned} & \sum_{\tilde{t}=t+1}^{\infty} \left(\delta \Pr[s(\theta_{\tilde{t}}) = 0] \right)^{\tilde{t}-(t+1)} \Pr[s(\theta_{\tilde{t}}) = 0] (L - \Pr[\theta_{\tilde{t}} = 1 \mid s(\theta_{\tilde{t}}) = 0]S) \\ &= \frac{\Pr[s(\theta_t) = 0]}{1 - \delta \Pr[s(\theta_t) = 0]} (L - \Pr[\theta_t = 1 \mid s(\theta_t) = 0]S) \\ &= \bar{A}(\pi, q, L, S, \delta). \end{aligned}$$

Then Lemma 1 implies that g^* is the voter's equilibrium strategy. If there exists a censorship policy satisfying (A.14), then sequential rationality for the illiberal government is trivially satisfied because it implies that the illiberal government is reelected with probability 1. It is immediate that such a censorship policy exists—namely, the full censorship policy c_F . This is because c_F always induces belief $\mu_t^*(m_t, c_F) = \Pr[s(\theta_t) = 1] = \pi$ and (A.12) holds. Finally, the voter's beliefs were constructed via Bayes' rule.

Part 3. Suppose

$$\pi S - \delta \bar{A}(\pi, q, L, S, \delta) < L \quad \text{and} \quad L \leq \mu_t^*(1, c_\ell)S. \quad (\text{A.15})$$

Because the equilibrium is essentially unique (Lemma A.2), it suffices to show that $\sigma^* = (g^*, c^*, \{\mu_t^*\}_{t=1}^\infty)$ such that

$$g^*(m_t, c_t) = i \iff \mu_t^*(m_t, c_t)S \geq L + \delta \Pr[s(\theta_t) = 1] \frac{\mu_t^*(1, c_\ell)S - L}{1 + \delta \Pr[s(\theta_t) = 1]}, \quad (\text{A.16})$$

$c^*(0) < 1 = c^*(1)$, and

$$\mu_t^*(1, c^*)S = L + \delta \Pr[s(\theta_t) = 1] \frac{\mu_t^*(1, c_\ell)S - L}{1 + \delta \Pr[s(\theta_t) = 1]}, \quad (\text{A.17})$$

is an equilibrium, and it induces the voter to elect an illiberal government if and only if $m_t = 1$ for any g_{t-1} .

That σ^* induces $g_t = i$ if and only if $m_t = 1$ follows by (A.16) and (A.17) and because

$$\mu_t^*(0, c_\ell) < \mu_t^*(0, c^*) < \mu_t^*(1, c^*) < \mu_t^*(1, c_\ell).$$

We now show that σ^* is an equilibrium. First, we show that

$$A(\pi, q, L, S, \delta \mid \sigma^*) = \Pr[s(\theta_t) = 1] \frac{\mu_t^*(1, c_\ell)S - L}{1 + \delta \Pr[s(\theta_t) = 1]}. \quad (\text{A.18})$$

Because σ^* induces $g_t = i$ if and only if $m_t = 1$, the voter's continuation payoffs are such that

$$\begin{aligned} V(g_t = \ell \mid \sigma^*) &= \Pr[s(\theta_t) = 1] \delta V(g_t = i \mid \sigma^*) \\ &\quad + \Pr[s(\theta_t) = 0] \left(L - \Pr[\theta_t = 1 \mid s(\theta_t) = 0] S + \delta V(g_t = \ell \mid \sigma^*) \right), \end{aligned} \quad (\text{A.19})$$

and

$$\begin{aligned} V(g_t = i \mid \sigma^*) &= \left(\Pr[s(\theta_t) = 1] + c^*(0) \Pr[s(\theta_t) = 0] \right) \delta V(g_t = i \mid \sigma^*) \\ &\quad + (1 - c^*(0)) \Pr[s(\theta_t) = 0] \left(L - \mu_t^*(0, c^*) S + \delta V(g_t = \ell \mid \sigma^*) \right). \end{aligned} \quad (\text{A.20})$$

The value of $c^*(0)$ is implicitly defined in (A.17), i.e., $c^*(0)$ satisfies

$$\begin{aligned} \frac{\pi q + c^*(0) \pi (1 - q)}{\Pr[s(\theta_t) = 1] + c^*(0) \Pr[s(\theta_t) = 0]} S &= L + \delta \Pr[s(\theta_t) = 1] \frac{\mu_t^*(1, c_\ell) S - L}{1 + \delta \Pr[s(\theta_t) = 1]} \\ \iff c^*(0) &= \frac{\pi q S - L \Pr[s(\theta_t) = 1]}{L \Pr[s(\theta_t) = 0] - \pi (1 - q - \delta (1 - \pi) (2q - 1)) S}. \end{aligned} \quad (\text{A.21})$$

Subtracting (A.19) from (A.20) and simplifying yields the accountability cost of illiberalism:

$$A(\pi, q, L, S, \delta \mid \sigma^*) = \frac{c^*(0) \Pr[s(\theta_t) = 0] \left(L - \Pr[\theta_t = 1 \mid s(\theta_t) = 0] S \right)}{1 - c^*(0) \delta \Pr[s(\theta_t) = 0]} \quad (\text{A.22})$$

which, after substituting (A.21), yields (A.18). Then Lemma 1 implies that g^* is the voter's equilibrium strategy.

Now consider the illiberal government's optimal choice. By Lemma 3 and (A.18), the censorship policy c^* described in (A.17) must be optimal for the illiberal government.

However, it remains to prove that there exists such a censorship policy, i.e., we must show that $c^*(0) \in [0, 1)$. First, note that the left hand side of (A.17) is continuous in $c^*(0)$; strictly decreasing with $c^*(0)$; and takes maximum and minimum values of $\Pr[\theta_t = 1 \mid s(\theta_t) = 1]S$ and πS at $c^*(0) = 0$ and $c^*(0) = 1$, respectively. Second, note that the right hand side is constant in $c^*(0)$. Therefore, by the Intermediate Value Theorem, it suffices to show that the right hand side of (A.17) takes a value in the interval

$$\left(\pi S, \Pr[\theta_t = 1 \mid s(\theta_t) = 1]S \right).$$

To see this, we rearrange the right hand side of (A.17) as

$$\frac{L + \delta \Pr[s(\theta_t) = 1] \mu_t^*(1, c_\ell) S}{1 + \delta \Pr[s(\theta_t) = 1]}. \quad (\text{A.23})$$

We show that (A.23) is bound between πS and $\mu_t^*(1, c_\ell) S$. Because (A.15) places an upper bound on L , the value (A.23) is no larger than $\mu_t^*(1, c_\ell) S$. Condition (A.15) also places a lower bound on L . Substituting the explicit form of $\bar{A}(\pi, q, L, S, \delta)$ into (A.15) yields

$$\pi S(1 - \delta \Pr[s(\theta_t) = 0]) + \delta \Pr[s(\theta_t) = 0] \Pr[\theta_t = 1 \mid s(\theta_t) = 0] S < L. \quad (\text{A.24})$$

Finally, substituting (A.24) into (A.23) proves the lower bound:

$$\frac{L + \delta \Pr[s(\theta_t) = 1] \mu_t^*(1, c_\ell) S}{1 + \delta \Pr[s(\theta_t) = 1]} > \frac{\pi S(1 - \delta \Pr[s(\theta_t) = 0]) + \delta \pi S}{1 + \delta \Pr[s(\theta_t) = 1]} = \pi S.$$

Last, we note that the voter's beliefs were constructed via Bayes' rule.

Part 4. Suppose $L > \mu_t^*(1, c_\ell) S$. Because the equilibrium is essentially unique (Lemma A.2), it suffices to show that $\sigma^* = (g^*, c^*, \{\mu_t^*\}_{t=1}^\infty)$ such that

$$g^*(m_t, c_t) = i \iff \mu_t^*(m_t, c_t) S \geq L, \quad (\text{A.25})$$

and c^* is any censorship policy is an equilibrium and that it induces $g_t = \ell$ for all $t \geq 1$.

That σ^* induces $g_t = \ell$ for all $t \geq 1$ follows by Lemma 1 because, for all censorship policies c_t and any message m'_t that is sent with positive probability under c_t ,

$$\mu_t^*(m'_t, c_t) S \leq \mu_t^*(1, c_\ell) S < L. \quad (\text{A.26})$$

We now show that σ^* is an equilibrium. Because $g_t = \ell$ for all t , $A(\pi, q, L, S, \delta \mid \sigma^*) = 0$, then Lemma 1 implies that g^* is the voter's equilibrium strategy. Sequential rational-

ity for the illiberal government is trivially satisfied because the illiberal government is indifferent between all censorship policies (see (A.26)). Finally, the voter's beliefs were constructed via Bayes' rule. ■

Proof of Corollary 1. Follows from Proposition A.1. ■

Proposition A.1 *There exist cutoffs*

$$\underline{\kappa}(\pi, q) := \mu_t^*(0, c_\ell), \quad (\text{A.27})$$

$$\kappa(\pi, q; \delta) := \pi(1 - \delta(1 - \pi)(2q - 1)), \quad (\text{A.28})$$

$$\bar{\kappa}(\pi, q) := \mu_t^*(1, c_\ell), \quad (\text{A.29})$$

such that $0 < \underline{\kappa}(\pi, q) < \kappa(\pi, q; \delta) < \bar{\kappa}(\pi, q) < 1$ and, in equilibrium, a constitution with

1. $\bar{L} \leq \underline{\kappa}(\pi, q)$ induces an efficient stable illiberal democracy;
2. $\bar{L} \in (\underline{\kappa}(\pi, q), \kappa(\pi, q; \delta)]$ induces an inefficient stable illiberal democracy;
3. $\bar{L} \in (\kappa(\pi, q; \delta), \bar{\kappa}(\pi, q)]$ induces a cycling liberal-illiberal democracy; and
4. $\bar{L} > \bar{\kappa}(\pi, q)$ induces an efficient stable liberal democracy.

Proof of Proposition A.1. From Proposition 1, an efficient stable liberal democracy occurs if $\mu_t^*(1, c_\ell) < \bar{L}$; an efficient stable illiberal democracy occurs if $\bar{L} \leq \mu_t^*(0, c_\ell)$; an inefficient stable illiberal democracy occurs if

$$\mu_t^*(0, c_\ell) < \bar{L} \quad \text{and} \quad L \leq \pi S - \delta \bar{A}(\pi, q, L, S, \delta); \quad (\text{A.30})$$

a cycling liberal-illiberal democracy occurs if

$$\pi S - \delta \bar{A}(\pi, q, L, S, \delta) < L \quad \text{and} \quad \bar{L} \leq \mu_t^*(1, c_\ell). \quad (\text{A.31})$$

Substituting (6) into the second inequality of (A.30) and noticing that the first inequality in (A.30) implies that $\bar{A}(\pi, q, L, S, \delta) > 0$ yields $\bar{L} \leq \pi(1 - \delta(1 - \pi)(2q - 1))$. Also the first inequality in (A.31) simplifies to

$$\bar{L} > \begin{cases} \pi & \text{if } \bar{L} < \Pr[\theta_t = 1 \mid s(\theta_t) = 0], \\ \pi(1 - \delta(1 - \pi)(2q - 1)) & \text{otherwise.} \end{cases} \quad (\text{A.32})$$

However, the first case in (A.32) cannot occur because $\Pr[\theta_t = 1 \mid s(\theta_t) = 0] > \pi$. Therefore, (A.30) and (A.31) simplify to

$$\mu_t^*(0, c_\ell) < \bar{L} \leq \pi(1 - \delta(1 - \pi)(2q - 1))$$

and

$$\pi(1 - \delta(1 - \pi)(2q - 1)) < \bar{L} \leq \mu_t^*(1, c_\ell),$$

respectively. Defining $\underline{\kappa}(\pi, q)$, $\kappa(\pi, q; \delta)$, and $\bar{\kappa}(\pi, q)$ as in (A.27), (A.28), and (A.29) and noting that $0 < \underline{\kappa}(\pi, q) < \kappa(\pi, q; \delta) < \bar{\kappa}(\pi, q) < 1$, completes the proof. ■

Proof of Proposition 2. *For* π . If $\bar{L} \geq 1$, then an efficient stable liberal democracy occurs for all values of $\pi \in (0, 1)$, and the illiberal government is elected with probability 0 for all π . Thus, the proposition statement trivially holds. Now suppose that $\bar{L} \in (0, 1)$. Recall from Proposition A.1 that

$$0 < \underline{\kappa}(\pi, q) < \kappa(\pi, q; \delta) < \bar{\kappa}(\pi, q) < 1 \quad \forall \pi \in (0, 1).$$

These thresholds are all continuous and strictly increasing with π . Furthermore, in the limiting cases $\pi = 0$ and $\pi = 1$, we have

$$0 = \underline{\kappa}(0, q) = \kappa(0, q; \delta) = \bar{\kappa}(0, q) \quad \text{and} \quad 1 = \underline{\kappa}(1, q) = \kappa(1, q; \delta) = \bar{\kappa}(1, q).$$

Combing these observations, implies the following lemma.

Lemma A.3 *There exists three values $0 < \pi_1 < \pi_2 < \pi_3 < 1$ such that*

1. *if $\pi < \pi_1$, an efficient stable liberal democracy occurs, and if $g_{t-1} = \ell$, the illiberal government is elected with probability 0;*
2. *if $\pi_1 \leq \pi < \pi_2$ or $\pi_2 \leq \pi < \pi_3$, a cycling liberal-illiberal democracy or an inefficient stable illiberal democracy occurs. In either case, if $g_{t-1} = \ell$, the illiberal government is elected with probability $\Pr[s(\theta_t) = 1] = \pi q + (1 - \pi)(1 - q)$, which is strictly increasing with π ;*
3. *if $\pi_3 \leq \pi$, an efficient stable illiberal democracy occurs, and if $g_{t-1} = \ell$, an efficient stable illiberal democracy occurs, and the illiberal government is elected with probability 1.*

For \bar{L} . From Proposition A.1, as \bar{L} increases from 0, the regime transitions from an efficient stable illiberal, inefficient stable illiberal, cycling liberal-illiberal to an efficient stable liberal democracy. Therefore, when $g_{t-1} = \ell$, the illiberal government's election

probability transitions from 1 to $\Pr[s(\theta_t) = 1]$ and then to 0. Because $\Pr[s(\theta_t) = 1]$ is constant in \bar{L} , the illiberal government's election probability is weakly decreasing in \bar{L} .

For q . If $\bar{L} \geq 1$, then an efficient stable liberal democracy occurs for all values of $q \in (1/2, 1)$, and the illiberal government is elected with probability 0 for all π . Thus, the proposition statement holds.

Now suppose that $\bar{L} < \pi$. In this case, an efficient stable liberal democracy will never arise because $\pi < \bar{\kappa}(\pi, q)$ for all $q \in (1/2, 1)$ (Proposition A.1). We now show that there is a unique value \underline{q} such that $\bar{L} = \underline{\kappa}(\pi, q)$: for all $q \leq \underline{q}$, an efficient stable illiberal democracy occurs; for $q > \underline{q}$, either a cycling liberal-illiberal or an inefficient stable illiberal democracy will occur. This will complete Parts 2 and 4 of the proposition statement because, when $g_{t-1} = \ell$, the illiberal government's election probability will transition from 1 to

$$\Pr[s(\theta_t) = 1] = \pi q + (1 - \pi)(1 - q) = 1 - \pi + q(2\pi - 1),$$

which is strictly increasing with q if $\pi > 1/2$ and strictly decreasing with q if $\pi < 1/2$. To show the existence of \underline{q} notice that $\underline{\kappa}(\pi, q)$ is continuous and decreasing with $q \in (1/2, 1)$, and in the limiting cases of $q = 1/2$ and $q = 1$, takes maximum and minimum values of π and 0, respectively. Therefore, $\underline{\kappa}(\pi, q)$ ranges from 0 to π and $\bar{L} \in (0, \pi)$, and so there is a unique value of $\underline{q} \in (1/2, 1)$ such that $\bar{L} = \underline{\kappa}(\pi, \underline{q})$.

Now suppose that $\bar{L} > \pi$. In this case, neither an efficient stable illiberal democracy nor an inefficient stable illiberal democracy will occur because $\underline{\kappa}(\pi, q) < \kappa(\pi, q; \delta) < \pi$ for all $q \in (1/2, 1)$ (Proposition A.1). We now prove Lemma A.4.

Lemma A.4 *There exists a unique value \bar{q} such that $\bar{L} = \bar{\kappa}(\pi, \bar{q})$: for all $q < \bar{q}$, an efficient stable liberal democracy occurs; for $q \geq \bar{q}$, a cycling liberal-illiberal occurs.*

Proof. Notice that $\bar{\kappa}(\pi, q)$ is continuous and increasing with $q \in (1/2, 1)$ and, in the limiting cases of $q = 1/2$ and $q = 1$, takes maximum and minimum values of 1 and π , respectively. Therefore, $\bar{\kappa}(\pi, q)$ ranges from π and 1 and $\bar{L} \in (\pi, 1)$, and so there is a unique value of $\bar{q} \in (1/2, 1)$ such that $\bar{L} = \bar{\kappa}(\pi, \bar{q})$. ■

This completes Parts 1 and 3 of the proposition because, when $g_{t-1} = \ell$, the illiberal government's election probability will transition from 0 to

$$\Pr[s(\theta_t) = 1] = \pi q + (1 - \pi)(1 - q) = 1 - \pi + q(2\pi - 1),$$

which is strictly increasing with q if $\pi > 1/2$ and strictly decreasing with q if $\pi < 1/2$. ■

Proof of Proposition 3. If $\bar{L} > 1$, then a stable liberal democracy occurs and the illiberal

government is elected with probability 0. We use the cutoffs π_1, π_2, π_3 from Lemma A.3. If $\pi < \pi_1$, then the illiberal government is elected with probability 0. Therefore, the proposition only applies when $\pi \geq \pi_1$ and $\bar{L} \leq 1$.

For π . For $\pi \in [\pi_1, \pi_2)$, a cycling liberal-illiberal democracy occurs and, by (A.21), we have

$$\begin{aligned} c^*(0) &= \frac{\pi q S - L \Pr[s(\theta_t) = 1]}{L \Pr[s(\theta_t) = 0] - \pi(1 - q - \delta(1 - \pi)(2q - 1))S} \\ &= \frac{\pi q - \bar{L}(\pi q + (1 - \pi)(1 - q))}{\bar{L}(\pi(1 - q) + (1 - \pi)q) - \pi(1 - q - \delta(1 - \pi)(2q - 1))}. \end{aligned} \quad (\text{A.33})$$

The derivative of (A.33) with respect to π is

$$\frac{\partial c^*(0)}{\partial \pi} = - \frac{(2q - 1)(\bar{L}(-1 + \bar{L} - \delta(1 - \pi)^2) + \delta(\bar{L} - 2\bar{L}(1 - \pi)\pi - \pi^2)q)}{(\pi(1 + \delta(1 - \pi) - \bar{L}) - (1 + 2\delta(1 - \pi))\pi q + \bar{L}(2\pi - 1)q)^2}.$$

This derivative is non-negative if and only if

$$\delta(\bar{L} - 2\bar{L}(1 - \pi)\pi - \pi^2)q \leq \bar{L}(1 - \bar{L} + \delta(1 - \pi)^2). \quad (\text{A.34})$$

If the left hand side of (A.34) is non-positive, the inequality holds because $\bar{L} \leq 1$ and, hence, the right hand side is positive. Otherwise, the left hand side of (A.34) is positive. Therefore,

$$\delta(\bar{L} - 2\bar{L}(1 - \pi)\pi - \pi^2)q < \delta(\bar{L} - 2\bar{L}(1 - \pi)\pi - \pi^2) \quad (\text{A.35})$$

$$\leq \bar{L}(1 - \bar{L} + \delta(1 - \pi)^2) \quad (\text{A.36})$$

where the first inequality follows from $q < 1$ and the last follows if and only if $0 \leq \bar{L} - \bar{L}^2 + \pi^2(1 - \delta\bar{L})$, which is true for all $\bar{L} \leq 1$. We conclude that $\frac{\partial c^*(0)}{\partial \pi} \geq 0$.

For $\pi \geq \pi_2$, either an inefficient stable illiberal democracy occurs or an efficient stable illiberal democracy occurs; in either case, we have $c^*(0) = 1$.

For \bar{L} . For $\bar{L} \in (\kappa(\pi, q; \delta), \bar{\kappa}(\pi, q)]$, a cycling liberal-illiberal democracy occurs and $c^*(0)$ is given by (A.33). The derivative of (A.33) with respect to \bar{L} is

$$\frac{\partial c^*(0)}{\partial \bar{L}} = \frac{-(1 - \pi)\pi(2q - 1)(1 + \delta(\pi q + (1 - \pi)(1 - q)))}{(\pi(1 + \delta - \bar{L} - \delta\pi) - (1 + 2\delta(1 - \pi))\pi q + \bar{L}(2\pi - 1)q)^2} \leq 0.$$

For $\bar{L} \leq \kappa(\pi, q; \delta)$, either an efficient stable illiberal democracy occurs or an inefficient stable illiberal democracy occurs; in either case, $c^*(0) = 1$.

For q . For $\pi < \bar{L} \leq 1$, because

$$0 < \underline{\kappa}(\pi, q) < \kappa(\pi, q; \delta) < \pi < \bar{L} \quad \forall q \in (1/2, 1),$$

only an efficient stable liberal democracy or a cycling liberal-illiberal democracy can occur. Furthermore, using the threshold in Lemma A.4, if $q < \bar{q}$, the illiberal government is never elected and, therefore, the proposition does not apply. Otherwise, $q \geq \bar{q}$ and the amount of censorship, $c^*(0)$, is given by (A.33). Taking the first derivative with respect to q yields

$$\frac{\partial c^*(0)}{\partial q} = \frac{(\bar{L} - \pi)((1 - \bar{L} + \delta(1 - \pi))\pi + \bar{L}(1 - \pi))}{(\pi(1 - \bar{L} + \delta(1 - \pi)) - (1 + 2\delta(1 - \pi))\pi q + \bar{L}(2\pi - 1)q)^2}, \quad (\text{A.37})$$

which is positive because $\pi < \bar{L} \leq 1$.

For $\bar{L} < \pi$, because $\bar{L} < \pi < \bar{\kappa}(\pi, q)$ for all $q \in (1/2, 1)$, an efficient stable liberal democracy will never occur. We prove the following lemma.

Lemma A.5 *There exists a unique value \hat{q} such that $\bar{L} = \kappa(\pi, \hat{q}; \delta)$: if $q \leq \hat{q}$, an efficient or inefficient stable illiberal democracy occurs; if $q > \hat{q}$, a cycling liberal-illiberal democracy occurs.*

Proof. Notice that

$$\underline{\kappa}(\pi, q), \kappa(\pi, q; \delta) : 0 < \underline{\kappa}(\pi, q) < \kappa(\pi, q; \delta) < \pi \quad \forall q \in (1/2, 1)$$

are continuous and decreasing with q . Furthermore, in the limiting cases of $q = 1/2$, $\kappa(\pi, q; \delta) = \pi(1 - \delta(1 - \pi)(2q - 1))$ takes a maximum value of π . Hence, as q increases from $1/2$, $\kappa(\pi, q; \delta)$ decreases continuously and at linear rate from π and $\bar{L} \in (0, \pi)$. Therefore, there is a unique value \hat{q} (possibly exceeding 1) such that $\bar{L} = \kappa(\pi, \hat{q}; \delta)$. ■

Lemma A.5 suffices to prove Part 2 of the proposition because, if $q \leq \hat{q}$, then $c^*(0)$ is equal to 1; and if $q > \hat{q}$, then $c^*(0)$ is given by (A.33), which is decreasing with q for $\bar{L} < \pi$ (see (A.37)). ■

Proof of Proposition 4. If $\bar{L} > 1$, then a stable liberal democracy occurs and the illiberal government is elected with probability 0. We use the cutoffs π_1, π_2, π_3 from Lemma A.3. If $\pi < \pi_1$, then the illiberal government is elected with probability 0. Therefore, the proposition only applies when $\pi \geq \pi_1$ and $\bar{L} \leq 1$.

For π . For $\pi \in [\pi_1, \pi_2)$, a cycling liberal-illiberal democracy occurs and the illiberal

government's reelection probability is

$$\Pr[m_t = 1 \mid c^*] = \Pr[s(\theta_t) = 1] + c^*(0) \Pr[s(\theta_t) = 0], \quad (\text{A.38})$$

where $c^*(0)$ is given by (A.21). Taking the derivative (A.38) with respect to π

$$\frac{\partial \Pr[m_t = 1 \mid c^*]}{\partial \pi} = \left(\frac{\partial \Pr[s(\theta_t) = 1]}{\partial \pi} + c^*(0) \frac{\partial \Pr[s(\theta_t) = 0]}{\partial \pi} \right) + \Pr[s(\theta_t) = 0] \frac{\partial c^*(0)}{\partial \pi}.$$

Because

$$\frac{\partial \Pr[s(\theta_t) = 1]}{\partial \pi} = - \frac{\partial \Pr[s(\theta_t) = 0]}{\partial \pi} = 2q - 1 > 0,$$

the derivative simplifies to

$$\frac{\partial \Pr[m_t = 1 \mid c^*]}{\partial \pi} = (2q - 1)(1 - c^*(0)) + \Pr[s(\theta_t) = 0] \frac{\partial c^*(0)}{\partial \pi} > 0$$

where the inequality follows because, by Part 3 of Proposition 1, $c^*(0) < 1$ and, by Proposition 3, $\frac{\partial c^*(0)}{\partial \pi} \geq 0$.

For $\pi \geq \pi_2$, either an inefficient stable illiberal democracy occurs or an efficient stable illiberal democracy occurs; in either case, the illiberal government is reelected with probability 1.

For \bar{L} . For $\bar{\kappa} < \bar{L}$, the illiberal government is elected with probability 0. Therefore, the proposition does not apply.

For $\bar{L} \in (\kappa(\pi, q; \delta), \bar{\kappa}(\pi, q)]$, a cycling liberal-illiberal democracy occurs and the reelection probability is given by (A.38). Taking the derivative of (A.38) with respect to \bar{L} yields

$$\frac{\partial \Pr[m_t = 1 \mid c^*]}{\partial \bar{L}} = \Pr[s(\theta_t) = 0] \frac{\partial c^*(0)}{\partial \bar{L}} \leq 0$$

where the inequality follows because, by Proposition 3, $\frac{\partial c^*(0)}{\partial \bar{L}} \leq 0$.

For $\bar{L} \leq \kappa(\pi, q; \delta)$, either an efficient stable illiberal democracy occurs or an inefficient stable illiberal democracy occurs; in either case, the illiberal government's reelection probability 1.

For q . Suppose $1/2 < \pi < \bar{L} \leq 1$. Using the threshold in Lemma A.4, the proposition only applies to $q \geq \bar{q}$, when a cycling liberal-illiberal democracy occurs and the illiberal government's reelection probability is given by (A.38). Taking the derivative of (A.38) with respect to q

$$\frac{\partial \Pr[m_t = 1 \mid c^*]}{\partial q} = \left(\frac{\partial \Pr[s(\theta_t) = 1]}{\partial q} + c^*(0) \frac{\partial \Pr[s(\theta_t) = 0]}{\partial q} \right) + \Pr[s(\theta_t) = 0] \frac{\partial c^*(0)}{\partial q}.$$

Because

$$\frac{\partial \Pr[s(\theta_t) = 1]}{\partial q} = -\frac{\partial \Pr[s(\theta_t) = 0]}{\partial q} = (2\pi - 1)$$

the derivative simplifies to

$$\frac{\partial \Pr[m_t = 1 \mid c^*]}{\partial q} = (2\pi - 1)(1 - c^*(0)) + \Pr[s(\theta_t) = 0] \frac{\partial c^*(0)}{\partial q} \geq 0, \quad (\text{A.39})$$

where the inequality follows because, Proposition 3, $\frac{\partial c^*(0)}{\partial q} \geq 0 \iff \pi < \bar{L}$.

Now suppose that $\bar{L} < \pi < 1/2$. Using the threshold in Lemma A.5, the reelection probability is equal to one for all $q \leq \hat{q}$ and equal to (A.38) otherwise. Thus, it suffices to show that (A.38) is decreasing with q when $\bar{L} < \pi < 1/2$ and $q > \hat{q}$ —this follows from (A.39) by noting that, by Proposition 3, $\frac{\partial c^*(0)}{\partial q} \geq 0 \iff \pi < \bar{L}$.

It remains to show that if neither $\bar{L} < \pi < 1/2$ nor $1/2 < \pi < \bar{L}$ hold, then the reelection probability can be non-monotonic in q . We provide an explicit example that mirrors Panel (a) of Figure 3. Let $\bar{L} = 1/3, \pi = 0.32, \delta = 0.9$. By Proposition A.1, a cycling liberal-illiberal democracies occurs if

$$\begin{aligned} \pi(1 - \delta(1 - \pi)(2q - 1)) < \bar{L} &\leq \Pr[\theta_t = 1 \mid s(\theta_t) = 1] \\ \iff \frac{1 - \pi}{1 + \pi} = \frac{17}{33} &\leq q < 1 \end{aligned} \quad (\text{A.40})$$

Therefore, when (A.40) holds, the derivative of illiberal government's reelection probability is equal to (A.39). For our parameter values, this is approximately

$$-\frac{8.16 \times 10^5 (2.60341 - 12.4286q + 12.6311q^2)}{(-3836 + 7797q)^2}.$$

In the limit case of $q = 1$, this value is negative and equal to -0.145933 . For $q = \frac{17}{33}$, this value is positive and equal to 11.1825 . ■

Proof of Proposition 5. We begin by giving a formal definition of the voter's long-run payoff. To do so we introduce some notation and auxiliary results. For any set of parameters, the equilibrium dynamics constitute a Markov process with 2 states that correspond to the type of government. Abusing notation slightly, let state ℓ (state i) be the event that $g_t = \ell$ ($g_t = i$). Formally, we represent the Markov process via a (Markovian) transition matrix:

$$M = \begin{pmatrix} p_{\ell,\ell} & p_{\ell,i} \\ p_{i,\ell} & p_{i,i} \end{pmatrix},$$

where the entry $p_{jk} \in [0, 1]$ is the probability of transitioning from state $j \in \{i, \ell\}$ to state $k \in \{i, \ell\}$ and $\sum_{k=i, \ell} p_{jk} = 1$ for all j .

Lemma A.6 *For any given set of parameters, the equilibrium dynamics induce a Markov process that has a unique stationary distribution, $M_\infty := (M_{\infty; \ell}, M_{\infty; i})$, and this stationary distribution is the limiting distribution. In particular, the stationary distribution equals*

1. $M_\infty(\text{stable liberal}) := \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$ for an efficient stable liberal democracy;
2. $M_\infty(\text{cycling}) := \begin{pmatrix} \frac{\Pr[m_t=0|c^*]}{\Pr[s(\theta_t)=1]+\Pr[m_t=0|c^*]} & \frac{\Pr[s(\theta_t)=1]}{\Pr[s(\theta_t)=1]+\Pr[m_t=0|c^*]} \\ \frac{\Pr[m_t=0|c^*]}{\Pr[s(\theta_t)=1]+\Pr[m_t=0|c^*]} & \frac{\Pr[s(\theta_t)=1]}{\Pr[s(\theta_t)=1]+\Pr[m_t=0|c^*]} \end{pmatrix}$ for a cycling liberal-illiberal democracy;
3. $M_\infty(\text{inefficient stable illiberal}) := \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}$ for an inefficient stable illiberal democracy;
4. $M_\infty(\text{efficient stable illiberal}) := \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}$ for an efficient stable illiberal democracy.

Proof. We begin by constructing the transition matrix induced by each equilibrium regime.

1. An efficient stable liberal democracy has transition matrix

$$M(\text{stable liberal}) := \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix};$$

2. A cycling liberal-illiberal democracy has transition matrix

$$M(\text{cycling}) := \begin{pmatrix} \Pr[s(\theta_t) = 0] & \Pr[s(\theta_t) = 1] \\ \Pr[m_t = 0 | c^*] & \Pr[m_t = 1 | c^*] \end{pmatrix};$$

3. An inefficient stable illiberal democracy has transition matrix

$$M(\text{inefficient stable illiberal}) := \begin{pmatrix} \Pr[s(\theta_t) = 0] & \Pr[s(\theta_t) = 1] \\ 0 & 1 \end{pmatrix};$$

4. An efficient stable illiberal democracy has transition matrix

$$M(\text{efficient stable illiberal}) := \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}.$$

The results then follow from the fact⁴⁰ that, for any binary state Markov process $\begin{pmatrix} 1-a & a \\ b & 1-b \end{pmatrix}$ with $0 < a + b < 2$, the unique stationary (and limiting) distribution is $(\frac{b}{a+b} \frac{a}{a+b})$. ■

We now define the voter's long-run payoff.

Definition A.1 For a given set of parameters, let $M_\infty = (M_{\infty;\ell}, M_{\infty;i})$ denote the stationary distribution induced by the equilibrium dynamics. The voter's long-run payoff is

$$U_\infty := M_{\infty;\ell}V(g_t = \ell \mid \sigma^*) + M_{\infty;i}V(g_t = i \mid \sigma^*). \quad (\text{A.41})$$

We now turn to proving the proposition. Formally, we study π^* that solves the problem

$$\begin{aligned} & \sup_{\pi \in [0,1]} U_\infty \\ \text{s.t.} \quad & L = \mathcal{L}(\pi), \quad q = \mathcal{Q}(\pi). \end{aligned}$$

We show that any π that induces a stable illiberal democracy yields a long-run payoff of zero. In contrast, any π that induces either a stable liberal or a cycling democracy yields strictly positive long-run payoff. It then follows that, under the assumption that there exists $\pi \in [0, 1]$ for which the regime is not a stable (efficient or inefficient) illiberal democracy, the optimal π^* will induce either a stable liberal or a cycling democracy.

It remains to calculate the voter's long-run payoff for each regime, using Lemma A.6. In particular,

1. An efficient or inefficient stable illiberal democracy provides the voter with long-run payoff $U_\infty = V(g_t = i \mid \sigma^*) = 0$.
2. A cycling liberal-illiberal democracy provides the voter with long-run payoff

$$U_\infty = M_{\infty;\ell}V(g_t = \ell \mid \sigma^*) + M_{\infty;i}V(g_t = i \mid \sigma^*).$$

Furthermore, under an efficient stable illiberal democracy, $V(g_t = \ell \mid \sigma^*) > 0$ and $V(g_t = i \mid \sigma^*) > 0$. For the second inequality, see the final line of Proof of Lemma A.1; the first inequality is then immediate from Lemma A.1 and (5). Therefore, $U_\infty > 0$.

⁴⁰See, e.g., Chapter 11.2.6 of Pishro-Nik (2014).

3. An efficient stable liberal democracy provides the voter with long-run payoff

$$U_\infty = V(g_t = \ell \mid \sigma^*) = \frac{L - \pi S}{1 - \delta}.$$

An efficient stable liberal democracy only arises when $L > \mu_t^*(1, c_\ell)S \implies L - \pi S > 0$ and, hence, $U_\infty > 0$.

■

We now turn to the proofs of Propositions 6 and 7. Formally, the optimal constitution solves the following problem:

$$\begin{aligned} & \sup_{\pi \in (0,1)} L && \text{(A.42)} \\ \text{s.t.} \quad & L > \mu_t^*(1, c_\ell)S, \quad L = \mathcal{L}(\pi), \quad q = \mathcal{Q}(\pi). \end{aligned}$$

Before proving the propositions, we present an auxiliary lemma.

Lemma A.7 *Suppose π^* solves (A.42), then*

$$\mathcal{L}(\pi^*) \geq \frac{\pi^* \mathcal{Q}(\pi^*)}{\pi^* \mathcal{Q}(\pi^*) + (1 - \pi^*)(1 - \mathcal{Q}(\pi^*))} S.$$

Proof. First notice that, when $L = \mathcal{L}(\pi^*)$ and $q = \mathcal{Q}(\pi^*)$,

$$\mu_t^*(1, c_\ell)S = \frac{\pi^* \mathcal{Q}(\pi^*)}{\pi^* \mathcal{Q}(\pi^*) + (1 - \pi^*)(1 - \mathcal{Q}(\pi^*))} S.$$

By definition of the supremum, for every $\varepsilon > 0$, there must exist $L_\varepsilon : L_\varepsilon > \mathcal{L}(\pi^*) - \varepsilon$, which can be obtained within the constraints of (A.42). For the sake of a contradiction, suppose $\mathcal{L}(\pi^*) < \frac{\pi^* \mathcal{Q}(\pi^*)}{\pi^* \mathcal{Q}(\pi^*) + (1 - \pi^*)(1 - \mathcal{Q}(\pi^*))} S$. Because \mathcal{L} is strictly increasing, there exists a unique $\pi_\varepsilon \in [0, 1] : \mathcal{L}(\pi_\varepsilon) = L_\varepsilon$. The continuity of \mathcal{L} ensures that, for $\varepsilon > 0$ sufficiently small, π_ε is arbitrarily close to π^* . But both sides of the constraint $L > \mu_t^*(1, c_\ell)S$, i.e.,

$$\mathcal{L}(\pi) > \frac{\pi \mathcal{Q}(\pi)}{\pi \mathcal{Q}(\pi) + (1 - \pi)(1 - \mathcal{Q}(\pi))} S,$$

are continuous in π . Thus, when $\varepsilon > 0$ is sufficiently small, π_ε is such that $\mathcal{L}(\pi_\varepsilon) < \frac{\pi_\varepsilon \mathcal{Q}(\pi_\varepsilon)}{\pi_\varepsilon \mathcal{Q}(\pi_\varepsilon) + (1 - \pi_\varepsilon)(1 - \mathcal{Q}(\pi_\varepsilon))} S$, which does not satisfy the constraints of the problem—a contradiction. ■

Proof of Proposition 6. Recall that, when $L = \mathcal{L}(\pi)$ and $q = \mathcal{Q}(\pi)$, the first constraint in (A.42) reduces to

$$\mathcal{L}(\pi) > \frac{\pi \mathcal{Q}(\pi)}{\pi \mathcal{Q}(\pi) + (1 - \pi)(1 - \mathcal{Q}(\pi))} S. \quad (\text{A.43})$$

Point 1 is immediate because $\pi = 1$ satisfies the constraints of (A.43) and \mathcal{L} is strictly increasing.

To prove point 2, for sake of a contradiction, suppose there is an optimal $\pi^* \in [0, 1]$ such $\mathcal{L}(\pi^*) \neq \frac{\pi^* \mathcal{Q}(\pi^*)}{\pi^* \mathcal{Q}(\pi^*) + (1 - \pi^*)(1 - \mathcal{Q}(\pi^*))} S$. By Lemma A.7, it must be that

$$\mathcal{L}(\pi^*) > \frac{\pi^* \mathcal{Q}(\pi^*)}{\pi^* \mathcal{Q}(\pi^*) + (1 - \pi^*)(1 - \mathcal{Q}(\pi^*))} S$$

and, furthermore, $\pi^* < 1$ since $\mathcal{L}(1) \leq S$. Because both the left- and right-hand side of (A.43) are continuous in π , there exists $\pi' > \pi^*$ such that (A.43) holds with $\pi = \pi'$. But \mathcal{L} is strictly increasing in π ; thus $\mathcal{L}(\pi') > \mathcal{L}(\pi^*)$, which contradicts the optimality of π^* . ■

Proof of Proposition 7. Suppose there exists $\pi \in (0, 1)$ for which the regime is a stable liberal democracy under \mathcal{Q}' . Because $\mathcal{Q}(\pi) < \mathcal{Q}'(\pi)$ for all π and the right-hand side of the constraint $L > \mu_t^*(1, c_\ell)S$ is increasing in $\mathcal{Q}(\pi)$, it is immediate that there exists $\pi \in (0, 1)$ for which the regime is a stable liberal democracy under \mathcal{Q} . Therefore, a solution to (A.42) exists under both \mathcal{Q} and \mathcal{Q}' .

Recall that the first constraint in (A.42) reduces to

$$\mathcal{L}(\pi) > \frac{\pi \mathcal{Q}(\pi)}{\pi \mathcal{Q}(\pi) + (1 - \pi)(1 - \mathcal{Q}(\pi))} S. \quad (\text{A.44})$$

The right-hand side of (A.44) is increasing in $\mathcal{Q}(\cdot)$. Thus, under \mathcal{Q}' , the constraints of the problem become more stringent for every $\pi \in (0, 1)$, i.e., the set of π that satisfy the constraints under \mathcal{Q}' is smaller in the set-inclusion ordering compared to the set of π that satisfy the constraints under \mathcal{Q} . Because \mathcal{L} is increasing, it follows that $\pi_{\mathcal{Q}'}^* \leq \pi_{\mathcal{Q}}^*$ and $\mathcal{L}(\pi_{\mathcal{Q}'}^*) \leq \mathcal{L}(\pi_{\mathcal{Q}}^*)$. ■

Proof of Proposition 8. Suppose $\zeta > 0$ or $\gamma > 0$. Let σ^* be an equilibrium and recall that $g_0 = \ell$ (and, hence, $c_1 = c_\ell$). Let ρ_0 be the probability that the voter chooses $g_t = i$ when $c_t = c_\ell$, and let $\rho_1 \geq \rho_0$ be the probability that the voter chooses $g_t = i$ when $c_t = c^*$.⁴¹ If $g_t = i$, the illiberal government builds an autocracy with probability $\eta := 1 - (1 - \zeta)(1 - \gamma) > 0$ and engages in censorship c^* otherwise. Therefore, unless $\rho_0 = 0$

⁴¹ $\rho_1 \geq \rho_0$ because, in equilibrium, it must be the case that c^* guarantees the illiberal government at least as high probability of election as c_ℓ ; otherwise, there would be a profitable deviation—namely, $c_t = c_\ell$.

and conditional on not already being an autocracy, for any period t and censorship policy c_t , the illiberal government builds an autocracy in period $t + 1$ with probability at least $\rho_0\eta > 0$, and the total probability that the illiberal government builds an autocracy by time t is at least

$$1 - (1 - \rho_0\eta)^t.$$

Therefore, either we have a stable liberal democracy if $\rho_0 = 0$, or the probability that the illiberal government builds an autocracy converges to 1 as $t \rightarrow \infty$.

In equilibrium, a stable liberal democracy occurs if and only if the voter never elects the illiberal government. Suppose σ^* induces a stable liberal democracy. Then the accountability cost of illiberalism, $A(\pi, q, L, S, \delta, \zeta, \gamma \mid \sigma^*)$, is

$$\frac{L - S\pi}{1 - \delta} - (1 - \gamma)(1 - \zeta)\frac{L - S\pi}{1 - \delta} = \frac{L - S\pi}{1 - \delta}(1 - (1 - \gamma)(1 - \zeta)), \quad (\text{A.45})$$

and the voter's optimal strategy is

$$g_t = i \quad \iff \quad \mu_t^*(m_t, c_t)S \geq L + \delta\frac{L - S\pi}{1 - \delta}(1 - (1 - \gamma)(1 - \zeta)). \quad (\text{A.46})$$

Therefore, a necessary condition for a stable liberal democracy is that (A.46) never holds for any m_t and c_t . Because $\mu_t^*(m_t, c_t) \leq \mu_t^*(1, c_t)$ for all c_t , this condition is equivalent to

$$\mu_t^*(1, c_t)S < L + \delta\frac{L - S\pi}{1 - \delta}(1 - (1 - \gamma)(1 - \zeta)) \quad (\text{A.47})$$

$$\begin{aligned} \iff L/S &> \frac{\mu_t^*(1, c_t)(1 - \delta) + \delta\pi(1 - (1 - \gamma)(1 - \zeta))}{1 - \delta + \delta(1 - (1 - \gamma)(1 - \zeta))} \\ &=: \bar{\kappa}'(\pi, q, \delta; \zeta, \gamma), \end{aligned} \quad (\text{A.48})$$

which is strictly decreasing with γ and ζ .

We now show that (A.48) is a sufficient condition. Suppose that (A.48) holds and consider the assessment $\sigma' = (g', c', \{\mu_t^*\}_{t=1}^\infty)$ such that the voter's strategy g' satisfies (A.46) and c' is any censorship policy. Because (A.47) holds and voter's strategy, g' , the voter never elects the illiberal government and, hence, we have a stable liberal democracy. Furthermore, the accountability cost of illiberalism is given by (A.45). Therefore, given σ' , the voter's strategy is optimal. Furthermore, the illiberal government's choice of c' is also optimal, because (A.47) holds, the illiberal government is indifferent between any choice of censorship policy because they all induce a reelection probability of 0. Finally, the voter's beliefs are derived using Bayes' rule. We conclude that σ' is an equilibrium, and (A.48) is a sufficient condition for a stable liberal democracy. ■

B Equilibrium uniqueness

We prove that, for any given set of parameters, our model has an *essentially unique* equilibrium (i.e., Lemma A.2). Formally, we say that two equilibria are *essentially equivalent* if, for any t and any g_{t-1} , the probability that an illiberal government is elected is equal in each equilibrium (and hence the voter's and illiberal government's expected payoffs are also equal).⁴² For any given set of parameters, we say that the equilibrium is *essentially unique* if all equilibria are essentially equivalent.

We will now prove Lemma A.2 which we restate for convenience.

Lemma A.2 (Essentially unique equilibrium.) *For any given set of parameters, there is an essentially unique equilibrium.*

We begin by identifying three cases that can arise under any censorship policy, c_t :

- (I) the voter chooses $g_t = i$ with probability zero;
- (II) the voter chooses $g_t = i$ with probability one; or
- (III) the voter chooses $g_t = i$ with positive but non-unit probability.

On the equilibrium path, the censorship policy is either $c_t = c_\ell$ (when $g_{t-1} = \ell$) or $c_t = c^*$ (when $g_{t-1} = i$). Therefore, there are at most 9 “types” of equilibria that are characterized by the cases above. We will refer to an equilibrium where Case I applies when $c_t = c_\ell$ and Case II applies when $c_t = c^*$ as an I-II equilibrium (and similarly for the other 8 equilibria). Lemma B.1 rules out many of these types of equilibria and shows that just four different types of equilibria can exist: an I-I, II-II, III-II, or III-III equilibrium.

Lemma B.1

1. An I-II and I-III equilibrium does not exist.
2. An II-I and II-III equilibrium does not exist.
3. An III-I equilibrium does not exist.

Proof. *Part 1.* Suppose that Case I applies whenever $c_t = c_\ell$. By Lemma 1, this implies

$$\mu_t^*(1, c_\ell)S < L + \delta A(\pi, q, L, S, \delta \mid \sigma^*). \quad (\text{B.1})$$

⁴²In addition, by Lemma 3, if an illiberal government chooses a partial censorship policy, then it is unique across all equilibria.

We now show that, under c^* , neither Case II nor III can occur. For any censorship policy c_t and any message m that is sent with positive probability under c_t

$$\mu_t^*(0, c_\ell) \leq \mu_t^*(m, c_t) \leq \mu_t^*(1, c_\ell), \quad (\text{B.2})$$

therefore, by (B.1), for any m_t that is sent with positive probability under c^*

$$\mu_t^*(m_t, c^*)S < L + \delta A(\pi, q, L, S, \delta \mid \sigma^*).$$

Thus, the voter chooses $g_t = i$ with probability zero.

Part 2. Suppose that Case II applies whenever $c_t = c_\ell$. By Lemma 1, this implies

$$\mu_t^*(0, c_\ell)S \geq L + \delta A(\pi, q, L, S, \delta \mid \sigma^*). \quad (\text{B.3})$$

We now show that, under c^* , neither Case I nor III can occur. By (B.2) and (B.3), for any m_t that is sent with positive probability under c^*

$$\mu_t^*(m_t, c^*)S \geq L + \delta A(\pi, q, L, S, \delta \mid \sigma^*).$$

Thus, the voter chooses $g_t = i$ with probability one.

Part 3. Suppose that Case III applies whenever $c_t = c_\ell$. By Lemma 1, this implies that

$$\mu_t^*(0, c_\ell)S < L + \delta A(\pi, q, L, S, \delta \mid \sigma^*) \leq \mu_t^*(1, c_\ell)S.$$

Under c^* , Case I cannot occur because if the illiberal government were elected with probability zero when $c_t = c^*$, then c^* must not be optimal—a contradiction. ■

Lemma B.2 says that any equilibrium must be essentially unique.

Lemma B.2 *For a given set of parameters,*

1. *if an I-I equilibrium exists, then it is essentially unique;*
2. *if an II-II equilibrium exists, then it is essentially unique;*
3. *if an III-II equilibrium exists, then it is essentially unique; and*
4. *if an III-III equilibrium exists, then it is essentially unique.*

Proof. Part 1. Let σ^* be a I-I equilibrium. Then $A(\pi, q, L, S, \delta \mid \sigma^*) = 0$ and, by Lemma 1

and (B.2), this implies that

$$\mu_t^*(1, c_\ell)S < L. \quad (\text{B.4})$$

Now let σ' be another equilibrium. Clearly, if σ' is an I-I equilibrium, then σ^* and σ' are essentially equivalent. For sake of a contradiction, suppose that σ' is not an I-I equilibrium. By Lemma B.1, this implies that σ' is either an II-II, III-II, or III-III equilibrium. In either case, by (B.2), the voter elects the illiberal government when $m_t = 1$ and $c_t = c_\ell$, i.e.,

$$\mu_t^*(1, c_\ell)S \geq L + \delta A(\pi, q, L, S, \delta \mid \sigma') \geq L, \quad (\text{B.5})$$

where the last inequality follows from (Lemma A.1). But (B.5) contradicts (B.4).

Part 2. Let σ^* be a II-II equilibrium. Then $A(\pi, q, L, S, \delta \mid \sigma^*) = 0$ and, by Lemma 1 and (B.2), this implies that

$$\mu_t^*(0, c_\ell)S \geq L. \quad (\text{B.6})$$

Now let σ' be another equilibrium. Clearly, if σ' is an II-II equilibrium, then σ^* and σ' are essentially equivalent. For sake of a contradiction, suppose that σ' is not an II-II equilibrium. By Lemma B.1, this implies that σ' is either a I-I, III-II, or III-III equilibrium. However, by Part 1 of this Lemma, σ' cannot be a I-I equilibrium. Therefore, σ' must be an III-II or III-III equilibrium.

First, suppose that σ' is an III-II equilibrium. In this case, the accountability cost of illiberalism is

$$A(\pi, q, L, S, \delta \mid \sigma') = \frac{\Pr[s(\theta_t) = 0]}{1 - \delta \Pr[s(\theta_t) = 0]} (L - \Pr[\theta_t = 1 \mid s(\theta_t) = 0]S). \quad (\text{B.7})$$

This follows because if the voter elects the illiberal government just once, then the illiberal government remains in power forever and, under $c_t = c_\ell$, the voter must choose $g_t = i$ if and only if $m_t = 1$. However, because $A(\pi, q, L, S, \delta \mid \sigma') \geq 0$ (Lemma A.1), (B.7) implies that $L \geq \Pr[\theta_t = 1 \mid s(\theta_t) = 0]S$. If this inequality is strict, then we achieve a contradiction because

$$L > \Pr[\theta_t = 1 \mid s(\theta_t) = 0]S \equiv \mu_t^*(0, c_\ell)S \geq L,$$

where the final inequality follows from (B.6). Otherwise, i.e., if $L = \Pr[\theta_t = 1 \mid s(\theta_t) = 0]S$, then $A(\pi, q, L, S, \delta \mid \sigma') = 0$ and we achieve a contradiction: under σ' and $c_t = c_\ell$, the voter

elects the liberal government when $m_t = 0$ and so

$$\mu_t^*(0, c_\ell)S < L + \delta A(\pi, q, L, S, \delta \mid \sigma') = L \leq \mu_t^*(0, c_\ell),$$

where the final inequality follows by (B.6).

Second, suppose that σ' is an III-III equilibrium. In this case, the accountability cost of illiberalism is

$$A(\pi, q, L, S, \delta \mid \sigma') = \frac{c'(0) \Pr[s(\theta_t) = 0] \left(L - \Pr[\theta_t = 1 \mid s(\theta_t) = 0]S \right)}{1 - c'(0)\delta \Pr[s(\theta_t) = 0]}. \quad (\text{B.8})$$

This follows because (i) if the voter elects the illiberal government with positive but non-unit probability when $c_t \in \{c_\ell, c'\}$, then she must choose $g_t = i$ if and only if $m_t = 1$ when $c_t \in \{c_\ell, c'\}$; and (ii) by Lemma 3, the illiberal government's equilibrium choice of censorship policy, c' , will be such that when $m_t = 1$ the voter is indifferent between choosing $g_t = i$ or $g_t = \ell$. However, because $A(\pi, q, L, S, \delta \mid \sigma') \geq 0$ (Lemma A.1), (B.8) implies that $L \geq \Pr[\theta_t = 1 \mid s(\theta_t) = 0]S$. As shown in the previous step (when assuming that σ' is an III-II equilibrium), this leads to a contradiction with (B.6).

Part 3. Let σ^* be a III-II equilibrium. Then $A(\pi, q, L, S, \delta \mid \sigma^*) = \bar{A}(\pi, q, L, S, \delta)$ (see Point 2 in Proof of Part 2 of Proposition 1 in Appendix A) and, by Lemma 1 and (B.2), this implies that for any message m_t that is sent with positive probability when $c_t = c^*$ we have

$$\mu_t^*(m_t, c^*)S \geq L + \delta \bar{A}(\pi, q, L, S, \delta). \quad (\text{B.9})$$

Now let σ' be another equilibrium. If σ' is a III-II equilibrium, then σ^* and σ' are essentially equivalent because, when $g_{t-1} = i$ (and hence $c_t = c^*$), the illiberal government is elected with probability one and, when $g_{t-1} = \ell$ (and hence $c_t = c_\ell$), the illiberal government is elected with probability $\Pr[s(\theta_t) = 1]$. For sake of a contradiction, suppose that σ' is not a III-II equilibrium. By Lemma B.1, and Parts 1 and 2 of this Lemma, σ' must be a III-III equilibrium. Therefore, by Lemma 3, the illiberal government's equilibrium choice of censorship policy c' is such that both messages $m_t = 0$ and $m_t = 1$ are sent with positive probability, and

$$\mu_t^*(0, c')S < L + \delta A(\pi, q, L, S, \delta \mid \sigma') = \mu_t^*(1, c').$$

Notice that the illiberal government cannot be indifferent between all censorship policies: if they were, the full censorship policy, c_F , would lead to a non-deterministic election outcome—an impossibility because c_F induces a single belief for the voter and the voter

breaks ties in favor of the illiberal government. Therefore, by Lemma A.1, $A(\pi, q, L, S, \delta | \sigma') \leq \bar{A}(\pi, q, L, S, \delta)$. But this is a contradiction: c' cannot be optimal for the illiberal government since they could guarantee their reelection by deviating to the policy $c_t = c^*$.

Part 4. Let σ^* be a III-III equilibrium, and let σ' be another equilibrium. By Lemma B.1 and Parts 1–3 of this Lemma, σ' must be a III-III equilibrium. Note that, by (B.2), in both σ^* and σ' the illiberal government's probability of election when $g_{t-1} = \ell$ (and hence $c_t = c_\ell$) is equal to $\Pr[s(\theta_t) = 1]$. Therefore, it suffices to prove that the illiberal government's period- t election probability when $g_{t-1} = i$ is equal across both equilibria. In particular, we prove this by showing that in both equilibria the illiberal government chooses the same censorship policy, i.e., $c^* = c'$. Following our argument in the Proof of Part 3 of Proposition 1 in Appendix A, in any III-III equilibrium, $\hat{\sigma}$, we have

$$A(\pi, q, L, S, \delta | \hat{\sigma}) = \frac{\hat{c}(0) \Pr[s(\theta_t) = 0] \left(L - \Pr[\theta_t = 1 | s(\theta_t) = 0] S \right)}{1 - \hat{c}(0) \delta \Pr[s(\theta_t) = 0]},$$

$\hat{c}(1) = 1$ and $\hat{c}(0)$ such that

$$\frac{\pi q + \hat{c}(0) \pi (1 - q)}{\Pr[s(\theta_t) = 1] + \hat{c}(0) \Pr[s(\theta_t) = 0]} S = L + \delta A(\pi, q, L, S, \delta | \hat{\sigma}). \quad (\text{B.10})$$

By assumption (σ^* and σ' exist), a solution $\hat{c}(0) \in [0, 1)$ exists. But then it must be unique because the left hand side of (B.10) is decreasing with $\hat{c}(0)$ and the right hand side is increasing with $\hat{c}(0)$. Therefore, $c^*(0) = c'(0)$ and, hence, $c^* = c'$. ■

Collectively, Lemmas B.1 and B.2 prove Lemma A.2.

C A model with long term commitment

We extend our analysis to a setting where the illiberal government has access to a technology that allows it to commit to a multi-period (infinite sequence) of censorship policies.

C.1 Setup and equilibrium concept

The voter's payoff and action set, the liberal government's behavior, and the illiberal government's payoffs are unchanged. In contrast to the benchmark model, once elected, an illiberal government publicly chooses (and commits to) an infinite sequence of censorship policies for the next period and each subsequent period that they are reelected.

Before formalizing the illiberal government's action, we introduce some notation. We

will denote an infinite sequence of censorship policies by $z = (z_1, z_2, \dots)$, i.e., for each positive integer r , $z_r : \{0, 1\} \rightarrow [0, 1]$ where $z_r(s)$ is the probability that the illiberal government sends message $m = 1$ to the voter when the signal realization equals $s \in \{0, 1\}$. Let \mathcal{Z} denote the set of all infinite sequence censorship policies. If an illiberal government comes to power in period t (i.e., $g_{t-1} = \ell$ and $g_t = i$), then it (publicly) chooses $z \in \mathcal{Z}$ and it is common knowledge that for as long as the voter continues to reelect the illiberal government, $c_{t+k} = z_k$.

As in the benchmark model, we characterize the perfect Bayesian equilibria of our model in which the voter and illiberal government play pure Markovian strategies. Following the Bayesian persuasion literature ([Kamenica and Gentzkow, 2011](#)), we focus on equilibria in which the voter chooses the illiberal government whenever she is indifferent.

A Markovian pure strategy for an illiberal government is an infinite sequence of censorship policies z that they choose whenever they are elected for the first time. The voter's belief that $\theta_t = 1$, denoted by μ_t , is a mapping from the message observed by the voter, m_t , and the censorship policy, c_t , into a probability.

As a consequence of the illiberal government's ability to commit to a sequence of censorship policies, the voter's period- t strategy will depend on not only the period- t message and period- t censorship policy but also on whether the incumbent government is an illiberal government and, if so, how long they have continually been elected. For any period t such that $g_t = i$, let $h_t \in \{1, 2, \dots\}$ denote the number of periods that the illiberal government has continuously maintained election, i.e., h_t is the largest positive integer h such that $g_{t'} = i$ for all $t' = t + 1 - h, \dots, t$. If $g_t = \ell$, we set $h_t = 0$.

Depending on $g_{t-1} \in \{i, \ell\}$, the voter's Markovian pure strategy, g , takes two distinct forms. If $g_{t-1} = \ell$ (equivalently, $h_{t-1} = 0$), then the voter's strategy is a mapping $g(\cdot \mid g_{t-1} = \ell)$ from the message observed by the voter, m_t , and the censorship policy, c_t , into the period- t government, g_t . If $g_{t-1} = i$ (equivalently, $h_{t-1} \geq 1$), then the voter's strategy is a mapping $g(\cdot \mid g_{t-1} = i)$ from the message observed by the voter, m_t , the censorship policy, $c_t = z_{h_{t-1}}$, the infinite sequence of censorship policies chosen by the in-power illiberal government, z , and the number of periods that the illiberal government has continuously maintained election, $h_{t-1} \geq 1$, into the period- t government, g_t . A Markovian assessment is therefore a triple $\sigma = (g, z, \{\mu_t\}_{t=1}^{\infty})$.

Definition C.1 (Equilibrium.) *An assessment $\sigma^* = (g^*, z^*, \{\mu_t^*\}_{t=1}^{\infty})$ is an equilibrium if, for each period t :*

(i) $g^*(m_t, c_t \mid g_{t-1} = \ell) = i$ if and only if

$$\begin{aligned} & L - \mu_t(m_t, c_t)S + \mathbb{E}\left[\sum_{\tilde{t}=t+1}^{\infty} \delta^{(\tilde{t}-t)} v(g_{\tilde{t}}, \theta_{\tilde{t}}) \mid g_t = \ell, \sigma^*\right] \\ & \leq \mathbb{E}\left[\sum_{\tilde{t}=t+1}^{\infty} \delta^{(\tilde{t}-t)} v(g_{\tilde{t}}, \theta_{\tilde{t}}) \mid g_t = i, \sigma^*, h_t = 1\right] \end{aligned}$$

and $g^*(m_t, c_t \mid g_{t-1} = i, \mathbf{z}, h_{t-1}) = i$ if and only if

$$\begin{aligned} & L - \mu_t(m_t, c_t)S + \mathbb{E}\left[\sum_{\tilde{t}=t+1}^{\infty} \delta^{(\tilde{t}-t)} v(g_{\tilde{t}}, \theta_{\tilde{t}}) \mid g_t = \ell, \sigma^*\right] \\ & \leq \mathbb{E}\left[\sum_{\tilde{t}=t+1}^{\infty} \delta^{(\tilde{t}-t)} v(g_{\tilde{t}}, \theta_{\tilde{t}}) \mid g_t = i, \mathbf{z}, \sigma^*, h_t = h_{t-1} + 1\right]; \end{aligned}$$

(ii) whenever $g_t = i$ and $h_t = 1$, then, for all sequences of censorship policies \mathbf{z}' ,

$$\mathbb{E}\left[\sum_{\tilde{t}=t+1}^{\infty} \delta^{\tilde{t}-t} R \mathcal{X}(\{g_{t'}\}_{t+1}^{\tilde{t}}) \mid \mathbf{z} = \mathbf{z}^*, \sigma^*\right] \geq \mathbb{E}\left[\sum_{\tilde{t}=t+1}^{\infty} \delta^{\tilde{t}-t} R \mathcal{X}(\{g_{t'}\}_{t+1}^{\tilde{t}}) \mid \mathbf{z} = \mathbf{z}', \sigma^*\right],$$

where

$$\mathcal{X}(\{g_{t'}\}_{t+1}^{\tilde{t}}) = \begin{cases} 1 & \text{if } g_{t'} = i \text{ for all } t' \in \{t+1, \dots, \tilde{t}\}, \\ 0 & \text{otherwise;} \end{cases}$$

(iii) the voter's belief, $\mu_t(m_t, c_t)$, is derived using Bayes' rule,⁴³ i.e.,

$$\begin{aligned} \mu_t(1, c_t) &= \mu_t^*(1, c_t) := \Pr[\theta_t = 1 \mid m_t = 1, c_t] \\ &= \frac{\pi[c_t(1)q + c_t(0)(1-q)]}{\pi[c_t(1)q + c_t(0)(1-q)] + (1-\pi)[c_t(1)(1-q) + c_t(0)q]}, \end{aligned}$$

and

$$\begin{aligned} \mu_t(0, c_t) &= \mu_t^*(0, c_t) := \Pr[\theta_t = 1 \mid m_t = 0, c_t] \\ &= \frac{\pi[(1-c_t(1))q + (1-c_t(0))(1-q)]}{\pi[(1-c_t(1))q + (1-c_t(0))(1-q)] + (1-\pi)[(1-c_t(1))(1-q) + (1-c_t(0))q]}. \end{aligned}$$

⁴³By Property 1 of Definition 3.1 in [Fudenberg and Tirole \(1991\)](#), under any censorship policy, c_t —whether on or off the equilibrium path—and any message, m_t , that occurs with positive probability under c_t , the voter's belief must be derived using Bayes' rule from the prior belief that $\theta_t = 1$, π , and the conditional probabilities, $c_t(0)$, $c_t(1)$, and q .

C.2 Preliminaries

We begin by introducing notation for the voter's continuation payoffs. Notice that, in contrast to the benchmark model, the payoff from electing an illiberal government differs from the continuation payoff of reelecting an illiberal government. Furthermore, the continuation payoff of reelecting an illiberal government is conditional on the announced sequence of censorship policies, z .

Suppose $g_{t-1} = \ell$ (equivalently, $h_{t-1} = 0$), the voter's continuation payoffs from electing a liberal and illiberal government in period t are:

$$V(g_t = \ell \mid \sigma, h_{t-1} = 0) := \mathbb{E} \left[\sum_{\tilde{t}=t+1}^{\infty} \delta^{(\tilde{t}-t-1)} v(g_{\tilde{t}}, \theta_{\tilde{t}}) \mid g_t = \ell, \sigma \right]$$

and

$$V(g_t = i \mid \sigma, h_{t-1} = 0) := \mathbb{E} \left[\sum_{\tilde{t}=t+1}^{\infty} \delta^{(\tilde{t}-t-1)} v(g_{\tilde{t}}, \theta_{\tilde{t}}) \mid g_t = i, \sigma, h_t = 1 \right],$$

respectively.

Suppose $g_{t-1} = i$, $h_{t-1} \geq 1$ (and, hence, given $z \in \mathcal{Z}$), the voter's continuation payoffs from electing a liberal and illiberal government in period t are:

$$V(g_t = \ell \mid \sigma, h_{t-1}) := \mathbb{E} \left[\sum_{\tilde{t}=t+1}^{\infty} \delta^{(\tilde{t}-t-1)} v(g_{\tilde{t}}, \theta_{\tilde{t}}) \mid g_t = \ell, \sigma \right]$$

and

$$V(g_t = i \mid z, \sigma, h_{t-1}) := \mathbb{E} \left[\sum_{\tilde{t}=t+1}^{\infty} \delta^{(\tilde{t}-t-1)} v(g_{\tilde{t}}, \theta_{\tilde{t}}) \mid g_t = i, z, \sigma, h_t = h_{t-1} + 1 \right],$$

respectively.

It is immediate that, in equilibrium, the voter's continuation payoff from electing a liberal government is independent of if and how long the illiberal government has been in power, h_{t-1} . Similarly, keeping other parameters fixed, the (equilibrium) continuation payoffs are independent of t . Remark C.1 formally states these observations.

Remark C.1

1. *The continuation payoff from electing a liberal government is history independent:*

$$V(g_t = \ell \mid \sigma^*, h_{t-1} = 0) = V(g_t = \ell \mid \sigma^*, h_{t-1} = h) \quad \text{for all positive integers } h.$$

2. Holding all else equal, the continuation payoffs are time independent, i.e.,

$$V(g_t = \ell \mid \sigma^*, h_{t-1} = 0) = V(g_{t'} = \ell \mid \sigma^*, h_{t'-1} = 0) \quad \text{for all positive integers } t, t'$$

and

$$V(g_t = i \mid \mathbf{z}, \sigma^*, h_{t-1} = h) = V(g_{t'} = i \mid \mathbf{z}, \sigma^*, h_{t'-1} = h) \quad \text{for all positive integers } t, t', h \geq 0.$$

We define two version of the *accountability cost of illiberalism* that depend on the type of government that was previously elected: when $g_{t-1} = \ell$ (equivalently, $h_{t-1} = 0$),

$$A(\pi, q, L, S, \delta \mid \sigma^*, 0) := V(g_t = \ell \mid \sigma^*, h_{t-1} = 0) - V(g_t = i \mid \sigma^*, h_{t-1} = 0);$$

when $g_{t-1} = i, h_{t-1} = h \geq 1$ and given \mathbf{z} ,

$$A(\pi, q, L, S, \delta \mid \sigma^*, h, \mathbf{z}) := V(g_t = \ell \mid \sigma^*, h_{t-1} = h) - V(g_t = i \mid \sigma^*, h_{t-1} = h).$$

Notice that the notation for the accountability cost omits the time subscript (this is without loss due to Remark C.1).

The accountability cost of illiberalism is history dependent. As a result, the voter's equilibrium strategy is also (possibly) history dependent. Lemma C.1 characterizes the voter's equilibrium strategy. The proof of Lemma C.1 follows the proof of Lemma 1.

Lemma C.1 *In every equilibrium,*

1. if $h_{t-1} = 0$ ($g_{t-1} = \ell$), the voter elects the illiberal government in period t if and only if

$$\mu_t(m_t, c_\ell)S \geq L + \delta A(\pi, q, L, S, \delta \mid \sigma^*, 0);$$

2. if $h_{t-1} \geq 1$ ($g_{t-1} = i$) and given \mathbf{z} , the voter elects the illiberal government in period t if and only if

$$\mu_t(m_t, c_t)S \geq L + \delta A(\pi, q, L, S, \delta \mid \mathbf{z}, \sigma^*, h_{t-1}).$$

We now prove a series of lemmas that will be useful in characterizing the set of equilibria. We begin with an auxiliary lemma.

Lemma C.2 *In every equilibrium,*

- (i) $V(g_t = \ell \mid \sigma^*, h_{t-1} = 0) \geq \frac{L - \pi S}{1 - \delta}$;
- (ii) $V(g_t = i \mid \sigma^*, h_{t-1} = 0) \geq \frac{L - \pi S}{1 - \delta}$;
- (iii) $V(g_t = i \mid \sigma^*, h_{t-1} = 0) \geq 0$;
- (iv) $V(g_t = i \mid \mathbf{z}', \sigma^*, h_{t-1} = h') \geq 0$ for all positive integers h' and all sequences of censorship policies $\mathbf{z}' \in \mathcal{Z}$.

Proof. Part (i): The continuation payoff from the liberal government can be bounded as follows:

$$\begin{aligned}
& V(g_t = \ell \mid \sigma^*, h_{t-1} = 0) \\
&= \sum_{m' \in \text{Supp}(c_\ell)} \Pr[m_{t+1} = m' \mid c_{t+1} = c_\ell] \max\{L - \mu_{t+1}^*(m', c_\ell)S + \delta V(g_{t+1} = \ell \mid \sigma^*, h_t = 0), \\
&\quad \delta V(g_{t+1} = i \mid \mathbf{z}^*, \sigma^*, h_t = 0)\} \\
&\geq L - \pi S + \delta V(g_{t+1} = \ell \mid \sigma^*, h_t = 0) \\
&= L - \pi S + \delta V(g_t = \ell \mid \sigma^*, h_{t-1} = 0), \tag{C.1}
\end{aligned}$$

where the final equality follows from Remark C.1. Rearranging (C.1) gives

$$V(g_t = \ell \mid \sigma^*, h_{t-1} = 0) \geq \frac{L - \pi S}{1 - \delta}.$$

Part (ii): The continuation payoff from the illiberal government can be bounded as follows:

$$\begin{aligned}
& V(g_t = i \mid \sigma^*, h_{t-1} = 0) \\
&= \sum_{m' \in \text{Supp}(z_1^*)} \Pr[m_{t+1} = m' \mid c_{t+1} = z_1^*] \max\{L - \mu_{t+1}^*(m', z_1^*)S + \delta V(g_{t+1} = \ell \mid \sigma^*, h_t = 1), \\
&\quad \delta V(g_{t+1} = i \mid \mathbf{z}^*, \sigma^*, h_t = 1)\} \\
&\geq L - \pi S + \delta V(g_{t+1} = \ell \mid \sigma^*, h_t = 1) \\
&= L - \pi S + \delta V(g_t = \ell \mid \sigma^*, h_{t-1} = 0) \\
&\geq \frac{L - \pi S}{1 - \delta},
\end{aligned}$$

where the last equality follows from Remark C.1 and the final inequality follows from Part (i).

Part (iii): Notice the following relationship between $V(g_t = i \mid \sigma^*, h_{t-1} = 0)$ and $V(g_{t+1} = i \mid \mathbf{z}^*, \sigma^*, h_t = 1)$:

$$\begin{aligned}
& V(g_t = i \mid \sigma^*, h_{t-1} = 0) \\
&= \sum_{m' \in \text{Supp}(z_1^*)} \Pr[m_{t+1} = m' \mid c_{t+1} = z_1^*] \max\{L - \mu_{t+1}^*(m', z_1^*)S + \delta V(g_{t+1} = \ell \mid \sigma^*, h_t = 1), \\
&\quad \delta V(g_{t+1} = i \mid \mathbf{z}^*, \sigma^*, h_t = 1)\} \\
&\geq \delta V(g_{t+1} = i \mid \mathbf{z}^*, \sigma^*, h_t = 1).
\end{aligned}$$

Applying a similar argument to $V(g_{t+1} = i \mid \mathbf{z}^*, \sigma^*, h_t = h)$ for all h recursively gives

$$V(g_t = i \mid \sigma^*, h_{t-1} = 0) \geq \delta^{h'} V(g_{t+h'} = i \mid \mathbf{z}^*, \sigma^*, h_{t-1+h'} = h'), \quad (\text{C.2})$$

for all positive integers h' . However, by construction, the lowest possible per-period payoff is $-S$ and, therefore, $V(g_{t+h'} = i \mid \mathbf{z}^*, \sigma^*, h_{t-1+h'} = h')$ is bounded below by $-S/(1-\delta)$. Taking the limit of (C.2) as $h' \rightarrow \infty$, we conclude that

$$V(g_t = i \mid \sigma^*, h_{t-1} = 0) \geq 0.$$

Part (iv) follows from a similar argument as Part (iii). ■

Lemma C.3 *Suppose $L \leq \mu_t^*(1, c_\ell)S$. In every equilibrium, when $g_{t-1} = \ell$, the illiberal government is elected with positive probability.*

Proof. For sake of a contradiction, suppose there is an equilibrium σ^* such that, when $g_{t-1} = \ell$, the illiberal government is elected with probability zero. It is immediate that the voter's continuation payoff from electing the liberal government is

$$\begin{aligned}
& V(g_t = \ell \mid \sigma^*, h_{t-1} = 0) = L - \pi S + \delta V(g_{t+1} = \ell \mid \sigma^*, h_t = 0) \\
& \iff V(g_t = \ell \mid \sigma^*, h_{t-1} = 0) = \frac{L - \pi S}{1 - \delta},
\end{aligned} \quad (\text{C.3})$$

where the equivalence follows from Remark C.1 and by rearranging.

We now analyze the voter's equilibrium strategy when $g_{t-1} = \ell$. Because the illiberal government is never elected when $g_{t-1} = \ell$, it must be that

$$L - \mu_t^*(1, c_\ell)S + \delta V(g_t = \ell \mid \sigma^*, h_{t-1} = 0) > \delta V(g_t = i \mid \sigma^*, h_{t-1} = 0). \quad (\text{C.4})$$

Because $L - \mu_t^*(1, c_\ell)S \leq 0$, it follows from (C.4) that

$$V(g_t = \ell \mid \sigma^*, h_{t-1} = 0) > V(g_t = i \mid \sigma^*, h_{t-1} = 0). \quad (\text{C.5})$$

But, by Lemma C.2 Part (ii) and (C.3),

$$V(g_t = i \mid \sigma^*, h_{t-1} = 0) \geq \frac{L - \pi S}{1 - \delta} = V(g_t = \ell \mid \sigma^*, h_{t-1} = 0),$$

which contradicts (C.5). ■

Lemma C.4 *Suppose $L \leq \mu_t^*(0, c_\ell)S$. In every equilibrium, when $g_{t-1} = \ell$, the voter elects an illiberal government with probability one.*

Proof. For sake of a contradiction, suppose there is an equilibrium σ^* such that, when $g_{t-1} = \ell$, the voter elects the illiberal government with non-unit probability. Since $L \leq \mu_t^*(0, c_\ell)S$ implies that $L < \mu_t^*(1, c_\ell)S$, by Lemma C.3, the voter must elect the illiberal government with positive probability when $g_{t-1} = \ell$. Together, this implies that, upon observing message $m_t = 0$, the voter reelects the liberal government and, upon observing $m_t = 1$, the voter elects the illiberal government, i.e.,

$$\begin{aligned} L - \mu_t^*(1, c_\ell)S + \delta V(g_t = \ell \mid \sigma^*, h_{t-1} = 0) &\leq \delta V(g_t = i \mid \sigma^*, h_{t-1} = 0) \\ &< L - \mu_t^*(0, c_\ell)S + \delta V(g_t = \ell \mid \sigma^*, h_{t-1} = 0). \end{aligned} \quad (\text{C.6})$$

Since $L - \mu_t^*(0, c_\ell)S \leq 0$, we have

$$V(g_t = i \mid \sigma^*, h_{t-1} = 0) < V(g_t = \ell \mid \sigma^*, h_{t-1} = 0). \quad (\text{C.7})$$

Expanding the right-hand side gives

$$\begin{aligned} &V(g_t = \ell \mid \sigma^*, h_{t-1} = 0) \\ &= \Pr[s(\theta_t) = 0](L - \mu_t^*(0, c_\ell)S + \delta V(g_{t+1} = \ell \mid \sigma^*, h_t = 0)) + \Pr[s(\theta_t) = 1]\delta V(g_{t+1} = i \mid \sigma^*, h_t = 0) \\ &< L - \mu_t^*(0, c_\ell)S + \delta V(g_{t+1} = \ell \mid \sigma^*, h_t = 0), \end{aligned}$$

where the inequality follows by substituting (C.6). This implies that

$$V(g_t = \ell \mid \sigma^*, h_{t-1} = 0) < \frac{L - \mu_t^*(0, c_\ell)S}{1 - \delta} \leq 0,$$

because $L - \mu_t^*(0, c_\ell)S \leq 0$. Combined with (C.7) this gives

$$V(g_t = i \mid \sigma^*, h_{t-1} = 0) < V(g_t = \ell \mid \sigma^*, h_{t-1} = 0) < 0.$$

But, by Lemma C.2 Part (iii), $0 \leq V(g_t = i \mid \sigma^*, h_{t-1} = 0)$ —a contradiction. ■

Lemma C.5 *Suppose $L \leq \mu_t^*(0, c_\ell)S$. In every equilibrium, when $g_{t-1} = i$, the voter reelects the illiberal government with probability one.*

Proof. For sake of a contradiction, suppose there is an equilibrium σ^* such that, when $g_{t-1} = i$, the illiberal government is reelected with non-unit probability. This implies that there that for some $h' \geq 1$, when $g_{t-1} = i$ and $h_{t-1} = h'$, the voter elects the liberal government with positive probability, i.e., there exists message $m \in \text{Supp}(z_{h'}^*)$ such that

$$L - \mu_t^*(m, z_{h'}^*)S + \delta V(g_t = \ell \mid \sigma^*, h_{t-1} = h') > \delta V(g_t = i \mid z^*, \sigma^*, h_{t-1} = h'). \quad (\text{C.8})$$

Because $L \leq \mu_t^*(0, c_\ell)S$, we have that $L - \mu_t^*(m', c_t)S \leq 0$ for all censorship policies c_t and all messages $m' \in \text{Supp}(c_t)$. Thus, (C.8) implies that

$$V(g_t = \ell \mid \sigma^*, h_{t-1} = h') > V(g_t = i \mid z^*, \sigma^*, h_{t-1} = h').$$

Applying Part (iv) of Lemma C.2, then gives

$$V(g_t = \ell \mid \sigma^*, h_{t-1} = h') > 0.$$

A contradiction then ensues because the voter's continuation payoff $V(g_t = \ell \mid \sigma^*, h_{t-1} = 1)$ is bounded above by

$$\sum_{\tilde{t}=t+1}^{\infty} \delta^{(\tilde{t}-t-1)} \max\{0, L - \mu_t^*(0, c_\ell)S\} \leq 0.$$

■

Lemma C.6 *Suppose $L > \mu_t^*(1, c_\ell)S$. In every equilibrium the voter elects an illiberal government with probability zero in all periods.*

Proof. Because $g_0 = \ell$, it suffices to show that the voter never elects an illiberal government when $g_{t-1} = \ell$. For the sake of a contradiction, suppose σ^* is an equilibrium where

the illiberal government is elected with positive probability when $g_{t-1} = \ell$. Then it must be that the illiberal government is elected when $m_t = 1$:

$$L - \mu_t^*(1, c_\ell)S + \delta V(g_t = \ell \mid \sigma^*, h_{t-1} = 0) \leq \delta V(g_t = i \mid \sigma^*, h_{t-1} = 0).$$

Because $L - \mu_t^*(1, c_\ell)S > 0$, it follows that

$$V(g_t = \ell \mid \sigma^*, h_{t-1} = 0) < V(g_t = i \mid \sigma^*, h_{t-1} = 0). \quad (\text{C.9})$$

By Part (i) of Lemma C.2,

$$V(g_t = \ell \mid \sigma^*, h_{t-1} = 0) \geq \frac{L - \pi S}{1 - \delta}. \quad (\text{C.10})$$

But we also have the following bound on the continuation payoff from electing an illiberal government:

$$\begin{aligned} V(g_t = i \mid \sigma^*, h_{t-1} = 0) &\leq \sum_{\tilde{t}=t+1}^{\infty} \delta^{(\tilde{t}-t-1)} \mathbb{E}[\max\{0, L - \mu_{\tilde{t}}^*(m_{\tilde{t}}, c_{\tilde{t}})S\} \mid c_{\tilde{t}} = c_\ell] \\ &= \frac{L - \pi S}{1 - \delta}, \end{aligned} \quad (\text{C.11})$$

this bounds holds because $L > \mu_t^*(1, c_\ell)S > \mu_t^*(0, c_\ell)S$. Combining (C.9), (C.10), and (C.11) yields a contradiction:

$$\frac{L - \pi S}{1 - \delta} \geq V(g_t = i \mid \sigma^*, h_{t-1} = 0) > V(g_t = \ell \mid \sigma^*, h_{t-1} = 0) \geq \frac{L - \pi S}{1 - \delta}.$$

■

C.3 Equilibrium regimes

We begin by characterizing the efficient stable regimes.

Proposition C.1 (Efficient stable illiberal democracy.) *Suppose $L \leq \mu_t^*(0, c_\ell)S$. There exists an equilibrium where the voter elects an illiberal government if and only if $\mu_t^*(m_t, c_t)S \geq L$; and $g_t = i$ for all $t \geq 1$. This equilibrium is essentially unique: in every equilibrium, the illiberal government is elected with unit probability in every period.*

Proof. Equilibrium existence is immediate because, since $L \leq \mu_t^*(0, c_\ell)S$, the equilibrium of the benchmark model is also an equilibrium in the extended model. The illiberal gov-

ernment does not have a profitable deviation since they obtain the maximum payoff: once elected, they remain in power forever. Essential uniqueness follows from Lemmas C.4 and C.5. ■

Proposition C.2 (Efficient stable liberal democracy.) *Suppose $L > \mu_t^*(1, c_\ell)S$. There exists an equilibrium where the voter elects an illiberal government if and only if $\mu_t^*(m_t, c_t)S \geq L$; and $g_t = \ell$ for all $t \geq 1$. This equilibrium is essentially unique: in every equilibrium the voter elects an illiberal government with probability zero in all periods.*

Proof. We begin with existence. Consider the assessment, σ^* , where the illiberal government chooses the sequence of (non) censorship policies $\mathbf{z}^* = (c_\ell, c_\ell, \dots)$, $\mu_t^*(m_t, c_t)$ is determined by Bayes' rule, and the voter (re)elects the illiberal government if and only if $\mu_t^*(m_t, c_t)S \geq L$.

Because $L > \mu_t^*(1, c_\ell)S > \mu_t^*(0, c_\ell)S$ and given the voter's strategy, it is immediate that the voter never (re)elects the illiberal government and always (re)elects the liberal government. Therefore,

$$V(g_t = \ell \mid \sigma^*, h_{t-1} = 0) = V(g_t = i \mid \sigma^*, h_{t-1} = 0)$$

and

$$V(g_t = \ell \mid \sigma^*, h_{t-1} = h) = V(g_t = i \mid \sigma^*, h_{t-1} = h)$$

yielding

$$A(\pi, q, L, S, \delta \mid \sigma^*, 0) = A(\pi, q, L, S, \delta \mid \sigma^*, h, \mathbf{z}^*) = 0,$$

for all $h \geq 1$. It follows that the voter's strategy satisfies the equilibrium requirement per Definition C.1 and Lemma C.1.

Finally, it suffices to show that, once elected, the illiberal government does not have a profitable deviation $\mathbf{z}' = (z'_1, z'_2, \dots)$. For sake of a contradiction, suppose that \mathbf{z}' is a profitable deviation. Then it must be that, for some $m' \in \text{Supp}(z'_1)$,

$$\mu_t^*(m', z'_1)S \geq L;$$

this is impossible since among all censorship policies the maximum belief that can be induced is $\mu_t^*(1, c_\ell)$ and $\mu_t^*(1, c_\ell)S < L$. We conclude that the illiberal government does not have a profitable deviation and the proposed strategy profile is an equilibrium.

Essential uniqueness follows from Lemma C.6. ■

We now turn to the inefficient stable illiberal regime.

Proposition C.3 (Inefficient stable illiberal democracy.) *Suppose $\mu_t^*(0, c_\ell)S < L \leq \pi S - \delta \bar{A}(\pi, q, L, S, \delta)$. There exists an equilibrium in which, under a liberal government, the voter elects an illiberal government if and only if $m_t = 1$, which occurs with probability $\pi q + (1 - \pi)(1 - q)$. Once an illiberal government has been elected, it chooses a sequence of censorship policies such that the voter reelects the illiberal government with probability one in every period.*

Proof. The proof is immediate because, since $\mu_t^*(0, c_\ell)S < L \leq \pi S - \delta \bar{A}(\pi, q, L, S, \delta)$, the equilibrium of the benchmark model is also an equilibrium in the extended model. The illiberal government does not have a profitable deviation since they obtain the maximum payoff: once elected, they remain in power forever. ■

We now turn to the cycling regime.

Proposition C.4 (Cycling liberal-illiberal democracy.) *Suppose $\pi S - \delta \bar{A}(\pi, q, L, S, \delta) < L \leq \mu_t^*(1, c_\ell)S$. If an equilibrium exists, then it must feature liberal-illiberal cycles: when $g_t = \ell$, the illiberal government is elected with positive probability; when $g_t = i$, the illiberal government cannot guarantee its reelection forever.*

Proof. First notice that, by Lemma C.3, in every equilibrium, when $g_{t-1} = \ell$, the illiberal government is elected with positive probability. Now for sake of a contradiction, suppose there is an equilibrium σ^* such that once elected the illiberal government can guarantee its own reelection.

We begin by first determining the voter's equilibrium strategy when $g_{t-1} = \ell$. Because the illiberal government guarantees its own reelection forever, it must be that $V(g_t = i \mid \sigma^*, h_{t-1} = 0) = 0$. Furthermore,

$$\begin{aligned} V(g_t = \ell \mid \sigma^*, h_{t-1} = 0) &= \sum_{m=0}^1 \Pr[m_t = m \mid c_\ell] \max\{0, L - \mu_t^*(m, c_\ell)S + \delta V(g_{t+1} = \ell \mid \sigma^*, h_t = 0)\} \\ &\geq 0 \end{aligned}$$

and, hence, $A(\pi, q, L, S, \delta \mid \sigma^*, 0) \geq 0$. By Lemma C.1, when $g_{t-1} = \ell$, the voter elects the illiberal government in period t if and only if

$$\mu_t^*(m_t, c_\ell)S \geq L + \delta A(\pi, q, L, S, \delta \mid \sigma^*, 0).$$

By the argument within the proof of Proposition A.1, $\mu_t^*(0, c_\ell)S < L \leq \mu_t^*(1, c_\ell)S$. Thus, given that the illiberal government is elected with positive probability when $g_{t-1} = \ell$ and

because $A(\pi, q, L, S, \delta \mid \sigma^*, 0) \geq 0$, it must be that

$$\mu_t^*(0, c_\ell)S < L + \delta A(\pi, q, L, S, \delta \mid \sigma^*, 0) \leq \mu_t^*(1, c_\ell)S,$$

i.e., when $g_{t-1} = \ell$, the voter elects the illiberal government if and only if $m_t = 1$.

Further simplifying the voter's continuation payoff from electing the liberal government gives

$$\begin{aligned} V(g_t = \ell \mid \sigma^*, h_{t-1} = 0) &= \Pr[m_t = 0 \mid c_\ell](L - \mu_t^*(0, c_\ell)S + \delta V(g_{t+1} = \ell \mid \sigma^*, h_t = 0)) \\ \iff V(g_t = \ell \mid \sigma^*, h_{t-1} = 0) &= \frac{\Pr[s(\theta_t) = 0](L - \Pr[\theta_t = 1 \mid s(\theta_t) = 0]S)}{1 - \delta \Pr[s(\theta_t) = 0]} \\ \implies A(\pi, q, L, S, \delta \mid \sigma^*, 0) &= \bar{A}(\pi, q, L, S, \delta), \end{aligned}$$

where the second line follows from Remark C.1.

We now turn to the voter's equilibrium behavior when $g_{t-1} = i$. By assumption, the illiberal government is reelected with probability 1 when $g_{t-1} = i$; therefore,

$$\mu_t(m_t, z_{h_{t-1}}^*)S \geq L + \delta A(\pi, q, L, S, \delta \mid z^*, \sigma^*, h_{t-1}) \quad \text{for all } m_t \in \text{Supp}(z_{h_{t-1}}^*). \quad (\text{C.12})$$

Using the fact that $V(g_t = i \mid z^*, \sigma^*, h_{t-1}) = 0$ and that (using Remark C.1)

$$V(g_t = \ell \mid \sigma^*, h_{t-1}) = V(g_t = \ell \mid \sigma^*, h_{t-1} = 0) = \bar{A}(\pi, q, L, S, \delta);$$

we conclude that

$$A(\pi, q, L, S, \delta \mid z^*, \sigma^*, h_{t-1}) = \bar{A}(\pi, q, L, S, \delta).$$

But then (C.12) implies that

$$\mu_t(m_t, z_{h_{t-1}}^*)S \geq L + \delta \bar{A}(\pi, q, L, S, \delta) \quad \text{for all } m_t \in \text{Supp}(z_{h_{t-1}}^*). \quad (\text{C.13})$$

However, $\pi S < L + \delta \bar{A}(\pi, q, L, S, \delta)$. This contradicts (C.13) since Bayes' plausibility constraint implies that there exists at least one message $m' \in \text{Supp}(z_{h_{t-1}}^*)$ such that $\mu_t^*(m', z_{h_{t-1}}^*) \leq \pi$. ■

D A cheap talk model

We extend the results of our benchmark model to an alternative framework in which the illiberal government (agent) is unable to commit to a censorship policy, i.e., the illiberal

government chooses any (cheap talk) message upon observing the signal.

D.1 Setup and equilibrium concept

The voter's payoff and action set are unchanged. The liberal government's behavior and the illiberal government's payoffs are also unchanged. In contrast to the benchmark model, the illiberal government is unable to commit to a censorship policy: whenever elected ($g_{t-1} = i$), the illiberal government (privately) observes the underlying signal about the next period's state of the world $s(\theta_t) \in \{0, 1\}$ and then chooses a message $m_t \in \{0, 1\}$ to send to the voter.⁴⁴ Abusing notation, we denote the illiberal government's strategy by a mapping $c_t : \{0, 1\} \rightarrow [0, 1]$, where $c_t(s)$ is the probability that the illiberal government sends message $m_t = 1$ to the voter when $s(\theta_t) = s$,

We characterize perfect Bayesian equilibria in Markovian strategies. For consistency with our benchmark model, we focus on equilibria in which the voter chooses the illiberal government whenever she is indifferent. A Markovian strategy for the voter is a mapping g from the message observed by the voter, m_t , and the period- $(t-1)$ government, g_{t-1} , into the period- t government, g_t . A Markovian strategy for the period- t illiberal government is a mapping c from the period $t+1$ signal, $s(\theta_{t+1})$, into a message m_{t+1} . The voter's belief that $\theta_t = 1$, denoted by μ_t , is a mapping from the message observed by the voter, m_t , and the period- $(t-1)$ government, g_{t-1} , into a probability.

Definition D.1 (Equilibrium.) *An assessment $\sigma^* = (g^*, c^*, \{\mu_t^*\}_{t=1}^\infty)$ is an equilibrium if, for each period t :*

(i) $g^*(m_t, g_{t-1}) = i$ if and only if

$$L - \mu_t(m_t, g_{t-1})S + \mathbb{E} \left[\sum_{\tilde{t}=t+1}^{\infty} \delta^{\tilde{t}-t} v(g_{\tilde{t}}, \theta_{\tilde{t}}) \mid g_t = \ell, \sigma^* \right] \leq \mathbb{E} \left[\sum_{\tilde{t}=t+1}^{\infty} \delta^{\tilde{t}-t} v(g_{\tilde{t}}, \theta_{\tilde{t}}) \mid g_t = i, \sigma^* \right];$$

(ii) if $g_t = i$, then for all messages $m' \in \{0, 1\}$ and for all signal realizations $s \in \{0, 1\}$

$$\begin{aligned} & \mathbb{E} \left[\sum_{\tilde{t}=t+1}^{\infty} \delta^{\tilde{t}-t} R \mathcal{X}(\{g_{\tilde{t}}\}_{\tilde{t}=t+1}^{\tilde{t}}) \mid m_{t+1} \in \text{Supp}(c^*(s)), s(\theta_{t+1}) = s, \sigma^* \right] \\ & \geq \mathbb{E} \left[\sum_{\tilde{t}=t+1}^{\infty} \delta^{\tilde{t}-t} R \mathcal{X}(\{g_{\tilde{t}}\}_{\tilde{t}=t+1}^{\tilde{t}}) \mid m_{t+1} = m', s(\theta_{t+1}) = s, \sigma^* \right], \end{aligned}$$

⁴⁴The binary structure of the message space is without loss of generality.

where

$$\mathcal{X}(\{g_{t'}\}_{t+1}^{\tilde{t}}) = \begin{cases} 1 & \text{for all } t' \in \{t+1, \dots, \tilde{t}\}, \\ 0 & \text{otherwise;} \end{cases}$$

(iii) the voter's belief, $\mu_t(m_t, g_{t-1})$, is derived using Bayes' rule whenever possible, i.e.,

$$\mu_t(1, \ell) = \mu_t^*(1, \ell) := \Pr[\theta_t = 1 \mid s(\theta_t) = 1] = \frac{\pi q}{\pi q + (1 - \pi)(1 - q)},$$

and

$$\mu_t(0, \ell) = \mu_t^*(0, \ell) := \Pr[\theta_t = 1 \mid s(\theta_t) = 0] = \frac{\pi(1 - q)}{\pi(1 - q) + (1 - \pi)q},$$

and if $c^*(0) > 0$ or $c^*(1) > 0$,

$$\begin{aligned} \mu_t(1, i) &= \mu_t^*(1, i \mid c^*) := \Pr[\theta_t = 1 \mid m_t = 1, c^*] \\ &= \frac{\pi[c^*(1)q + c^*(0)(1 - q)]}{\pi[c^*(1)q + c^*(0)(1 - q)] + (1 - \pi)[c^*(1)(1 - q) + c^*(0)q]}, \end{aligned}$$

and if $c^*(0) < 1$ or $c^*(1) < 1$,

$$\begin{aligned} \mu_t(0, i) &= \mu_t^*(0, i \mid c^*) := \Pr[\theta_t = 1 \mid m_t = 0, c^*] \\ &= \frac{\pi[(1 - c^*(1))q + (1 - c^*(0))(1 - q)]}{\pi[(1 - c^*(1))q + (1 - c^*(0))(1 - q)] + (1 - \pi)[(1 - c^*(1))(1 - q) + (1 - c^*(0))q]}, \end{aligned}$$

D.2 Preliminaries

Let

$$V(g_t = \ell \mid \sigma) := \mathbb{E} \left[\sum_{\tilde{t}=t+1}^{\infty} \delta^{(\tilde{t}-t-1)} v(g_{\tilde{t}}, \theta_{\tilde{t}}) \mid g_t = \ell, \sigma \right]$$

and

$$V(g_t = i \mid \sigma) := \mathbb{E} \left[\sum_{\tilde{t}=t+1}^{\infty} \delta^{(\tilde{t}-t-1)} v(g_{\tilde{t}}, \theta_{\tilde{t}}) \mid g_t = i, \sigma \right]$$

denote the expected continuation payoffs from electing a liberal and illiberal government, respectively. We define the *accountability cost of illiberalism* as

$$A(\pi, q, L, S, \delta \mid \sigma^*) := V(g_t = \ell \mid \sigma^*) - V(g_t = i \mid \sigma^*).$$

Notice that $A(\pi, q, L, S, \delta \mid \sigma^*)$ is independent of t since we focus on Markovian strategies and, hence, $V(g_t = \ell \mid \sigma^*) = V(g_{t'} = \ell \mid \sigma^*)$ and $V(g_t = i \mid \sigma^*) = V(g_{t'} = i \mid \sigma^*)$ for all non-negative integers t, t' .

Lemma D.1 says that in equilibrium the voter elects an illiberal government if and only if the expected cost of security, $\mu_t(m_t, g_{t-1})S$, is greater than the sum of the value of liberty, L , and the discounted accountability cost of illiberalism, $\delta A(\pi, q, L, S, \delta \mid \sigma^*)$. The proof of Lemma D.1 follows the proof of Lemma 1.

Lemma D.1 (The voter's optimal strategy.) *In every equilibrium, the voter elects the illiberal government in period t if and only if*

$$\mu_t(m_t, g_{t-1})S \geq L + \delta A(\pi, q, L, S, \delta \mid \sigma^*).$$

We now turn to the illiberal government's problem. Lemma D.2 establishes that the illiberal government's problem reduces to the problem of choosing a message that maximizes their reelection probability.

Lemma D.2 (The illiberal government's problem.) *In every equilibrium, for any t such that $g_t = i$ and for any signal realization $s = s(\theta_{t+1})$, every message $m_{t+1} \in \text{Supp}(c^*(s))$ maximizes the government's reelection probability in period $t + 1$.*

Proof. By Definition D.1 (ii) and using the law of iterated expectations (and noting that period- t deviations do not affect continuation payoffs after period $t + 1$), for all m' ,

$$\begin{aligned} \delta R \Pr[g_{t+1} = i \mid m_{t+1} \in \text{Supp}(c^*(s)), s(\theta_{t+1}) = s] &+ \mathbb{E} \left[\sum_{\tilde{t}=t+2}^{\infty} \delta^{\tilde{t}-t} R \mathcal{X}(\{g_{t'}\}_{t+1}^{\tilde{t}}) \mid \sigma^* \right] \\ &\geq \delta R \Pr[g_{t+1} = i \mid m_{t+1} = m', s(\theta_{t+1}) = s] + \mathbb{E} \left[\sum_{\tilde{t}=t+2}^{\infty} \delta^{\tilde{t}-t} R \mathcal{X}(\{g_{t'}\}_{t+1}^{\tilde{t}}) \mid \sigma^* \right], \end{aligned}$$

and therefore

$$\Pr[g_{t+1} = i \mid m_{t+1} \in \text{Supp}(c^*(s)), s(\theta_{t+1}) = s] \geq \Pr[g_{t+1} = i \mid m_{t+1} = m', s(\theta_{t+1}) = s].$$

■

When $g_{t-1} = i$, the voter does not observe $s(\theta_t)$. Therefore, combined with Lemma D.2, the illiberal government's reelection probability must be independent of $s(\theta_t)$ and must be constant across all messages that the illiberal government sends with positive probability.

Furthermore, the voter follows a pure strategy in equilibrium and so, whenever $g_{t-1} = i$, the illiberal government's reelection probability is either 0 or 1. This result is stated in Corollary D.1.

Corollary D.1 *In every equilibrium, the illiberal government's reelection probability is either 0 or 1, is independent of the underlying signal, and is constant across all messages $m \in \text{Supp}(c^*)$.*

D.3 Results

We now present our equilibrium characterization. The characterization follows closely Proposition 1. However, there is one key distinction. When a cycling liberal-illiberal democracy is induced, the illiberal government is never reelected. This is inefficient for the voter. In fact, if the voter could observe the underlying information, she would reelect the illiberal government with strictly positive probability.

Proposition D.1 *In the essentially unique equilibrium,*

- (i) $L \leq \mu_t^*(0, \ell)S$ induces an efficient stable illiberal democracy;
- (ii) $\mu_t^*(0, \ell)S < L \leq \pi S - \delta \bar{A}(\pi, q, L, S, \delta)$ induces an inefficient stable illiberal democracy;
- (iii) $\pi S - \delta \bar{A}(\pi, q, L, S, \delta) < L \leq \mu_t^*(1, \ell)S$ induces a cycling liberal-illiberal democracy in which the illiberal government is elected if and only if $m_t = 1$ when $g_{t-1} = \ell$ and, when $g_{t-1} = i$, the illiberal government is reelected with probability 0;
- (iv) $\mu_t^*(1, \ell)S < L$ induces an efficient stable liberal democracy.

Proof. Section D.4 below proves with a series of lemmas that, for each of the 4 cases in Proposition D.1, if an equilibrium exists, then it is characterized by Proposition D.1.

We now turn to give a brief argument for equilibrium existence. We provide the essential elements to construct an equilibrium in each of the 4 cases of Proposition D.1. For all cases, let the illiberal government adopt a (completely uninformative) messaging strategy $c^* : c^*(0) = c^*(1) = 1/2$. Then the voter's beliefs are pinned down by Definition D.1: when $g_{t-1} = i$, the voter's belief is $\mu_t^*(m_t, i) = \pi$ for all $m_t \in \{0, 1\}$. The voter's unique sequentially optimal strategy, $g^*(m_t, g_{t-1})$, is then derived using Lemma D.1 and the continuation payoffs given by: for Part (i),

$$\begin{aligned} V(g_t = \ell \mid \sigma^*) &= \delta V(g_t = i \mid \sigma^*) = 0 \\ \implies A(\pi, q, L, S, \delta \mid \sigma^*) &= 0; \end{aligned}$$

for Part (ii),

$$\begin{aligned}
V(g_t = \ell \mid \sigma^*) &= \Pr[s(\theta_t) = 0](L - \mu_t^*(0, \ell)S + \delta V(g_t = \ell \mid \sigma^*)) \\
&\quad + \Pr[s(\theta_t) = 1]\delta V(g_t = i \mid \sigma^*) \\
V(g_t = i \mid \sigma^*) &= 0 \\
\implies A(\pi, q, L, S, \delta \mid \sigma^*) &= \frac{\Pr[s(\theta_t) = 0](L - \mu_t^*(0, \ell)S)}{1 - \delta \Pr[s(\theta_t) = 0]} \\
&= \bar{A}(\pi, q, L, S, \delta);
\end{aligned}$$

for Part (iii),

$$\begin{aligned}
V(g_t = \ell \mid \sigma^*) &= \Pr[s(\theta_t) = 0](L - \mu_t^*(0, \ell)S + \delta V(g_t = \ell \mid \sigma^*)) \\
&\quad + \Pr[s(\theta_t) = 1]\delta V(g_t = i \mid \sigma^*) \\
V(g_t = i \mid \sigma^*) &= L - \pi S + \delta V(g_t = \ell \mid \sigma^*) \\
\implies A(\pi, q, L, S, \delta \mid \sigma^*) &= \frac{\Pr[s(\theta_t) = 0](L - \mu_t^*(0, \ell)S) - (L - \pi S)}{1 + \delta \Pr[s(\theta_t) = 1]};
\end{aligned}$$

for Part (iv),

$$\begin{aligned}
V(g_t = \ell \mid \sigma^*) &= V(g_t = i \mid \sigma^*) = \frac{L - \pi S}{1 - \delta} \\
\implies A(\pi, q, L, S, \delta \mid \sigma^*) &= 0.
\end{aligned}$$

To conclude the proof, it is sufficient to notice that, given the voter's equilibrium strategy, all messages provide equal payoff to the illiberal government. ■

D.4 Essential uniqueness proofs

We provide a sequence of lemmas that collectively prove the essential uniqueness result stated in Proposition D.1. In particular, for each part of Proposition D.1:

Part (i). By Lemma D.6, in every equilibrium: for any $g_{t-1} \in \{i, \ell\}$, the illiberal government is elected with probability 1. Therefore, all equilibria are essentially unique.

Part (ii). By Lemma D.9, in every equilibrium: if $g_{t-1} = \ell$, the illiberal government is elected if and only if $m_t = 1$; if $g_{t-1} = i$, the illiberal government is reelected with probability one. Therefore, all equilibria are essentially unique.

Part (iii). By Lemma D.10, in every equilibrium: if $g_{t-1} = \ell$, the illiberal government is elected if and only if $m_t = 1$; if $g_{t-1} = i$, the illiberal government is reelected with probability zero. Therefore, all equilibria are essentially unique.

Part (iv). By Lemma D.8, in every equilibrium the liberal government is reelected with probability 1. Therefore, all equilibria are essentially unique.

Before proceeding with the proofs of the above lemmas, we introduce an auxiliary lemma that will be repeatedly applied in the proofs.

Lemma D.3 *In every equilibrium,*

$$(i) \quad V(g_t = i \mid \sigma^*) \geq 0;$$

$$(ii) \quad V(g_t = \ell \mid \sigma^*) \geq \delta V(g_t = i \mid \sigma^*) \geq 0; \text{ and}$$

$$(iii) \quad V(g_t = \ell \mid \sigma^*) \geq \frac{L-\pi S}{1-\delta} \text{ and } V(g_t = i \mid \sigma^*) \geq \frac{L-\pi S}{1-\delta}.$$

Proof. Starting with Part (i),

$$\begin{aligned} V(g_t = i \mid \sigma^*) &= \sum_{m' \in \text{Supp}(c^*)} \Pr[m_{t+1} = m' \mid c^*] \max\{L - \mu_{t+1}^*(m', i)S + \delta V(g_{t+1} = \ell \mid \sigma^*), \\ &\quad \delta V(g_{t+1} = i \mid \sigma^*)\} \\ &\geq \delta V(g_{t+1} = i \mid \sigma^*) \\ &= \delta V(g_t = i \mid \sigma^*), \end{aligned}$$

where the final equality follows from the observation that $V(g_{t+1} = i \mid \sigma^*) = V(g_t = i \mid \sigma^*)$. Therefore, $V(g_t = i \mid \sigma^*) \geq 0$.

Turning to Part (ii),

$$\begin{aligned} V(g_t = \ell \mid \sigma^*) &= \sum_{m' \in \{0,1\}} \Pr[m_{t+1} = m' \mid g_t = \ell] \max\{L - \mu_{t+1}^*(m', \ell)S + \delta V(g_{t+1} = \ell \mid \sigma^*), \\ &\quad \delta V(g_{t+1} = i \mid \sigma^*)\} \\ &\geq \delta V(g_{t+1} = i \mid \sigma^*) \\ &= \delta V(g_t = i \mid \sigma^*) \\ &\geq 0, \end{aligned}$$

where the final inequality follows from Part (i).

For Part (iii),

$$\begin{aligned}
V(g_t = \ell \mid \sigma^*) &= \sum_{m' \in \{0,1\}} \Pr[m_{t+1} = m' \mid g_t = \ell] \max\{L - \mu_{t+1}^*(m', \ell)S + \delta V(g_{t+1} = \ell \mid \sigma^*), \\
&\quad \delta V(g_{t+1} = i \mid \sigma^*)\} \\
&\geq L - \pi S + \delta V(g_{t+1} = \ell \mid \sigma^*) \\
&= L - \pi S + \delta V(g_t = \ell \mid \sigma^*),
\end{aligned}$$

where the final equality follows from the observation that $V(g_{t+1} = \ell \mid \sigma^*) = V(g_t = \ell \mid \sigma^*)$. Rearranging the final inequality gives

$$V(g_t = \ell \mid \sigma^*) \geq \frac{L - \pi S}{1 - \delta}. \quad (\text{D.1})$$

Similarly,

$$\begin{aligned}
V(g_t = i \mid \sigma^*) &= \sum_{m' \in \text{Supp}(c^*)} \Pr[m_{t+1} = m' \mid c^*] \max\{L - \mu_{t+1}^*(m', i)S + \delta V(g_{t+1} = \ell \mid \sigma^*), \\
&\quad \delta V(g_{t+1} = i \mid \sigma^*)\} \\
&\geq L - \pi S + \delta V(g_{t+1} = \ell \mid \sigma^*) \\
&\geq L - \pi S + \delta \frac{L - \pi S}{1 - \delta} \\
&= \frac{L - \pi S}{1 - \delta},
\end{aligned}$$

where the second last inequality follows from (D.1). ■

Lemma D.4 *Suppose $L \leq \mu_t^*(1, \ell)S$. In every equilibrium, when $g_{t-1} = \ell$, there is a strictly positive probability of electing an illiberal government ($g_t = i$).*

Proof. Suppose $L \leq \mu_t^*(1, \ell)S$ and, for sake of a contradiction, suppose there exists an equilibrium σ^* such that the voter always reelects a liberal government. This implies that

$$L - \mu_t^*(1, \ell)S + \delta V(g_t = \ell \mid \sigma^*) > \delta V(g_t = i \mid \sigma^*).$$

Because $L - \mu_t^*(1, \ell)S \leq 0$, the above inequality implies that

$$V(g_t = \ell \mid \sigma^*) > V(g_t = i \mid \sigma^*).$$

Furthermore, because the voter always reelects the liberal government:

$$V(g_t = \ell \mid \sigma^*) = \frac{L - \pi S}{1 - \delta}. \quad (\text{D.2})$$

But, by Lemma D.3 (iii), we have $V(g_t = i \mid \sigma^*) \geq \frac{L - \pi S}{1 - \delta}$ —a contradiction. ■

Lemma D.5 *Suppose $L \leq \mu_t^*(0, \ell)S$. In every equilibrium, when $g_{t-1} = \ell$, the illiberal government is elected ($g_t = i$) with probability one.*

Proof. Suppose $L \leq \mu_t^*(0, \ell)S$. Let σ^* be an equilibrium. By Lemma D.4, under σ^* , the voter elects the illiberal government with positive probability when $g_{t-1} = \ell$. For sake of a contradiction, suppose that the voter elects the illiberal government with positive but non-unit probability when $g_{t-1} = \ell$. This implies that

$$L - \mu_t^*(1, \ell)S + \delta V(g_t = \ell \mid \sigma^*) \leq \delta V(g_t = i \mid \sigma^*)$$

and

$$L - \mu_t^*(0, \ell)S + \delta V(g_t = \ell \mid \sigma^*) > \delta V(g_t = i \mid \sigma^*).$$

Because $L - \mu_t^*(0, \ell)S \leq 0$, the second inequality above implies that

$$V(g_t = \ell \mid \sigma^*) > V(g_t = i \mid \sigma^*).$$

Furthermore, by Lemma D.3 (i), we have that $V(g_t = i \mid \sigma^*) \geq 0$ and, hence,

$$V(g_t = \ell \mid \sigma^*) > 0. \quad (\text{D.3})$$

But the continuation payoff from $g_t = \ell$ is bounded:

$$V(g_t = \ell \mid \sigma^*) \leq \sum_{\tilde{t}=t+1}^{\infty} \delta^{(\tilde{t}-t-1)} \mathbb{E}[\max\{0, L - \mu_{\tilde{t}}^*(m_{\tilde{t}}, \ell)S\}],$$

where the right hand side is the highest possible payoff for the voter if the voter were to always observe the underlying signal and then choose the government. However, the right hand side equals zero since $L - \mu_{\tilde{t}}^*(m_{\tilde{t}}, \ell)S \leq 0$ for all $m_{\tilde{t}}$, which contradicts (D.3). ■

Lemma D.6 *Suppose $L \leq \mu_t^*(0, \ell)S$. In every equilibrium, for any $g_{t-1} \in \{i, \ell\}$, the illiberal government is elected with probability one*

Proof. Suppose $L \leq \mu_t^*(0, \ell)S$. Let σ^* be an equilibrium. Lemma D.5 and Corollary D.1 imply that, when $g_{t-1} = i$, the illiberal government is reelected with probability 0 or 1. For sake of a contradiction, suppose that the illiberal government is reelected with probability 0; this implies

$$V(g_t = i \mid \sigma^*) = L - \pi S + \delta V(g_{t+1} = \ell \mid \sigma^*).$$

But, by Lemma D.3 (i), $V(g_t = i \mid \sigma^*) \geq 0$. Therefore,

$$L - \pi S + \delta V(g_{t+1} = \ell \mid \sigma^*) \geq 0.$$

Because $L \leq \mu_t^*(0, \ell)S \implies L - \pi S < 0$, the above inequality implies that

$$V(g_t = \ell \mid \sigma^*) > 0.$$

But the continuation payoff from $g_t = \ell$ is bounded:

$$V(g_t = \ell \mid \sigma^*) \leq \sum_{\tilde{t}=t+1}^{\infty} \delta^{(\tilde{t}-t-1)} \mathbb{E}[\max\{0, L - \mu_{\tilde{t}}^*(m_{\tilde{t}}, \ell)S\}],$$

where the right hand side is the highest possible payoff for the voter if the voter were to always observe the underlying signal and then choose the government. However, the right hand side equals zero since $L - \mu_{\tilde{t}}^*(m_{\tilde{t}}, \ell)S \leq 0$ for all $m_{\tilde{t}}$, which is a contradiction. ■

Lemma D.7 Suppose $\mu_t^*(0, \ell)S < L \leq \mu_t^*(1, \ell)S$. In every equilibrium, when $g_{t-1} = \ell$, the illiberal government is elected ($g_t = i$) if and only if $m_t = 1$.

Proof. Suppose $\mu_t^*(0, \ell)S < L \leq \mu_t^*(1, \ell)S$. Let σ^* be an equilibrium. By Lemma D.4, when $g_{t-1} = \ell$, the illiberal government is elected with positive probability. Because the voter elects an illiberal government whenever they are indifferent, the voter must either elect the illiberal government for all messages or the lemma is proven. For sake of a contradiction, suppose the voter elects the illiberal government for all messages:

$$L - \mu_t^*(1, \ell)S + \delta V(g_t = \ell \mid \sigma^*) \leq \delta V(g_t = i \mid \sigma^*), \quad (\text{D.4})$$

and

$$L - \mu_t^*(0, \ell)S + \delta V(g_t = \ell \mid \sigma^*) \leq \delta V(g_t = i \mid \sigma^*). \quad (\text{D.5})$$

The continuation payoff from electing a liberal government is

$$V(g_t = \ell \mid \sigma^*) = \delta V(g_t = i \mid \sigma^*),$$

because the voter always elects an illiberal government when $g_t = \ell$. Using $\mu_t^*(0, \ell) \leq \pi \leq \mu_t^*(1, \ell)$, (D.4), and (D.5), we have

$$L - \pi S + \delta V(g_t = \ell \mid \sigma^*) \leq \delta V(g_t = i \mid \sigma^*). \quad (\text{D.6})$$

Corollary D.1 implies that, when $g_{t-1} = i$, the illiberal government is reelected with probability 1 or 0. Assume the former case, then the voter always elects the illiberal government (regardless of g_t) and, hence,

$$V(g_t = \ell \mid \sigma^*) = \delta V(g_t = i \mid \sigma^*) = 0.$$

But then (D.5) implies that

$$L - \mu_t^*(0, \ell)S \leq 0,$$

which is a contradiction. Now assume the latter case, then

$$V(g_t = i \mid \sigma^*) = L - \pi S + \delta V(g_t = \ell \mid \sigma^*)$$

and for all $m \in \text{Supp}(c^*)$

$$L - \mu_t^*(m, i)S + \delta V(g_t = \ell \mid \sigma^*) > \delta V(g_t = i \mid \sigma^*).$$

Taking the expectation over the distribution c^* gives

$$L - \pi S + \delta V(g_t = \ell \mid \sigma^*) > \delta V(g_t = i \mid \sigma^*),$$

which contradicts (D.6). ■

Lemma D.8 *Suppose $L > \mu_t^*(1, \ell)S$. In every equilibrium, when $g_{t-1} = \ell$, the liberal government is reelected ($g_t = \ell$) with probability one.*

Proof. Suppose $L > \mu_t^*(1, \ell)S$ and, for sake of a contradiction, suppose there exists an equilibrium σ^* such that the voter reelects a liberal government with probability less than one. This implies that

$$L - \mu_t^*(1, \ell)S + \delta V(g_t = \ell \mid \sigma^*) \leq \delta V(g_t = i \mid \sigma^*).$$

Because $L - \mu_t^*(1, \ell)S > 0$, we have $V(g_t = \ell \mid \sigma^*) < V(g_t = i \mid \sigma^*)$. But, by Lemma D.3 (iii),

$$V(g_t = \ell \mid \sigma^*) \geq \frac{L - \pi S}{1 - \delta}$$

and therefore

$$V(g_t = i \mid \sigma^*) > \frac{L - \pi S}{1 - \delta}.$$

But the continuation payoff from electing the illiberal government is bounded:

$$\begin{aligned} V(g_t = i \mid \sigma^*) &\leq \sum_{\bar{t}=t+1}^{\infty} \delta^{(\bar{t}-t-1)} \mathbb{E}[\max\{0, L - \mu_{\bar{t}}^*(m_{\bar{t}}, \ell)S\}] \\ &= \frac{L - \pi S}{1 - \delta}, \end{aligned}$$

since $L - \mu_{\bar{t}}^*(m_{\bar{t}}, \ell)S > 0$ for all $m_{\bar{t}}$ —a contradiction. ■

Lemma D.9 *Suppose $\mu_t^*(0, \ell)S < L$ and $L \leq \pi S - \delta \bar{A}(\pi, q, L, S, \delta)$. In every equilibrium, when $g_{t-1} = \ell$, the illiberal government is elected if and only if $m_t = 1$ and, when $g_{t-1} = i$, the illiberal government is reelected with probability 1.*

Proof. Suppose $\mu_t^*(0, \ell)S < L$ and $L \leq \pi S - \delta \bar{A}(\pi, q, L, S, \delta)$. Because $\bar{A}(\pi, q, L, S, \delta) \geq 0$, $\mu_t^*(0, \ell)S < L < \mu_t^*(1, \ell)S$ and, hence, Lemma D.7 implies that, when $g_{t-1} = \ell$, the voter elects the illiberal government if and only if $m_t = 1$. Hence,

$$V(g_t = \ell \mid \sigma^*) = \Pr[s_t = 1] \delta V(g_t = i \mid \sigma^*) + \Pr[s_t = 0] (L - \mu_t^*(0, \ell)S + \delta V(g_t = \ell \mid \sigma^*)).$$

Corollary D.1 then implies that the illiberal government is reelected (when $g_{t-1} = i$) with probability zero or probability one. For sake of a contradiction, suppose the illiberal government is reelected with probability zero, i.e.,

$$V(g_t = i \mid \sigma^*) = L - \pi S + \delta V(g_t = \ell \mid \sigma^*)$$

and, for all $m \in \text{Supp}(c^*)$,

$$\delta V(g_t = i \mid \sigma^*) < L - \mu_t^*(m, i)S + \delta V(g_t = \ell \mid \sigma^*).$$

Taking the expectation across all $m \in \text{Supp}(c^*)$ gives

$$\delta V(g_t = i \mid \sigma^*) < L - \pi S + \delta V(g_t = \ell \mid \sigma^*). \tag{D.7}$$

Thus,

$$\begin{aligned}
V(g_t = i \mid \sigma^*) &= L - \pi S + \delta V(g_t = \ell \mid \sigma^*) \\
&> \delta V(g_t = i \mid \sigma^*) \\
\implies V(g_t = i \mid \sigma^*) &> 0.
\end{aligned} \tag{D.8}$$

It follows that

$$\begin{aligned}
V(g_t = \ell \mid \sigma^*) &= \Pr[s_t = 1] \delta V(g_t = i \mid \sigma^*) + \Pr[s_t = 0] (L - \mu_t^*(0, \ell) S + \delta V(g_t = \ell \mid \sigma^*)) \\
&> \Pr[s_t = 0] (L - \mu_t^*(0, \ell) S + \delta V(g_t = \ell \mid \sigma^*)) \\
\implies V(g_t = \ell \mid \sigma^*) &> \frac{\Pr[s_t = 0] (L - \mu_t^*(0, \ell) S)}{1 - \Pr[s_t = 0] \delta}.
\end{aligned}$$

Recall that

$$\begin{aligned}
\bar{A}(\pi, q, L, S, \delta) &= \max \left\{ 0, \frac{\Pr[s(\theta_t) = 0]}{1 - \delta \Pr[s(\theta_t) = 0]} (L - \Pr[\theta_t = 1 \mid s(\theta_t) = 0] S) \right\} \\
&= \frac{\Pr[s_t = 0] (L - \mu_t^*(0, \ell) S)}{1 - \Pr[s_t = 0] \delta},
\end{aligned}$$

where the final equality follows because $\mu_t^*(0, \ell) = \Pr[\theta_t = 1 \mid s(\theta_t) = 0]$ and $\mu_t^*(0, \ell) S < L$. Thus,

$$V(g_t = \ell \mid \sigma^*) > \bar{A}(\pi, q, L, S, \delta). \tag{D.9}$$

Returning to (D.7) and by (D.8), we have

$$\begin{aligned}
0 < \delta V(g_t = i \mid \sigma^*) &< L - \pi S + \delta V(g_t = \ell \mid \sigma^*) \\
\implies L > \pi S + \delta V(g_t = \ell \mid \sigma^*).
\end{aligned}$$

Combining this with (D.9) gives

$$L > \pi S + \delta \bar{A}(\pi, q, L, S, \delta),$$

which contradicts the assumption that $L \leq \pi S - \delta \bar{A}(\pi, q, L, S, \delta)$. ■

Lemma D.10 *Suppose $\pi S - \delta \bar{A}(\pi, q, L, S, \delta) < L$ and $L \leq \mu_t^*(1, \ell) S$. In every equilibrium, when $g_{t-1} = \ell$, the illiberal government is elected if and only if $m_t = 1$ and, when $g_{t-1} = i$, the illiberal government is reelected with probability 0.*

Proof. Suppose $\pi S - \delta \bar{A}(\pi, q, L, S, \delta) < L$ and $L \leq \mu_t^*(1, \ell)S$. By the argument within the proof of Proposition A.1, $\mu_t^*(0, \ell)S < L \leq \mu_t^*(1, \ell)S$ and, hence, Lemma D.7 implies that, when $g_{t-1} = \ell$, the voter elects the illiberal government if and only if $m_t = 1$. Hence,

$$V(g_t = \ell \mid \sigma^*) = \Pr[s_t = 1] \delta V(g_t = i \mid \sigma^*) + \Pr[s_t = 0](L - \mu_t^*(0, \ell)S + \delta V(g_t = \ell \mid \sigma^*)).$$

Corollary D.1 then implies that the illiberal government is reelected (when $g_{t-1} = i$) with probability zero or probability one. For sake of a contradiction, suppose the illiberal government is reelected with probability one, i.e.,

$$V(g_t = i \mid \sigma^*) = 0$$

and, for all $m \in \text{Supp}(c^*)$,

$$\delta V(g_t = i \mid \sigma^*) \geq L - \mu_t^*(m, i)S + \delta V(g_t = \ell \mid \sigma^*).$$

Taking the expectation across all $m \in \text{Supp}(c^*)$ implies that

$$\delta V(g_t = i \mid \sigma^*) \geq L - \pi S + \delta V(g_t = \ell \mid \sigma^*). \quad (\text{D.10})$$

Now notice that

$$\begin{aligned} V(g_t = \ell \mid \sigma^*) &= \Pr[s_t = 0](L - \mu_t^*(0, \ell)S + \delta V(g_t = \ell \mid \sigma^*)) \\ \iff V(g_t = \ell \mid \sigma^*) &= \frac{\Pr[s_t = 0](L - \mu_t^*(0, \ell)S)}{1 - \delta \Pr[s_t = 0]} \\ &= \bar{A}(\pi, q, L, S, \delta). \end{aligned}$$

Returning to (D.10) gives

$$0 \geq L - \pi S + \delta V(g_t = \ell \mid \sigma^*) = L - \pi S + \delta \bar{A}(\pi, q, L, S, \delta),$$

which implies

$$\pi S - \delta \bar{A}(\pi, q, L, S, \delta) \geq L;$$

a contradiction. ■

E A model with symmetric censorship

We extend the model by adding a strategic liberal government who chooses whether—and how much—to censor information. Such a model may be useful in thinking of a principal who can only choose between types of agents who offer different services but can all manipulate information to induce the principal to believe that their specific services are needed. Thus, in contrast to our benchmark application to the rise and fall of illiberal democracies, the amount of information visible to the principal (voter) is independent from the constraints on the action of the agent (executive).

E.1 Setup and equilibrium concept

The voter's and the illiberal government's payoffs and action sets are unchanged. Similarly, the information structure on θ_t is unchanged. In contrast to the benchmark model, the liberal government is now strategic and chooses c_t to maximize their payoff, which is defined as follows: a liberal government elected for the first time in period t receives a rent $R > 0$ for period t and each subsequent period until the first period $t' > t : g_{t'} = i$, at which point an illiberal government is elected and the previous liberal government is replaced by a new liberal government. This payoff structure is identical to the illiberal government's payoff structure.

Similar to the benchmark model, we characterize the perfect Bayesian equilibria of our model in which the voter and governments play pure Markovian strategies. Following the Bayesian persuasion literature ([Kamenica and Gentzkow, 2011](#)), we focus on equilibria in which the voter chooses the incumbent government whenever she is indifferent. Abusing notation slightly, we denote the liberal government's censorship policy by c_ℓ and the illiberal government's and the liberal government's equilibrium choice of censorship by c_i^* and c_ℓ^* , respectively.

Definition E.1 (Equilibrium.) *An assessment $\sigma^* = (g^*, c_i^*, c_\ell^*, \{\mu_t^*\}_{t=1}^\infty)$ is an equilibrium if, for each period t :*

(i) *If $g_{t-1} = i$, then $g^*(m_t, c_t, g_{t-1}) = i$ if and only if*

$$L - \mu_t(m_t, c_t)S + \mathbb{E} \left[\sum_{\tilde{t}=t+1}^{\infty} \delta^{(\tilde{t}-t)} v(g_{\tilde{t}}, \theta_{\tilde{t}}) \mid g_t = \ell, \sigma^* \right] \leq \mathbb{E} \left[\sum_{\tilde{t}=t+1}^{\infty} \delta^{(\tilde{t}-t)} v(g_{\tilde{t}}, \theta_{\tilde{t}}) \mid g_t = i, \sigma^* \right];$$

If $g_{t-1} = \ell$, then $g^*(m_t, c_t, g_{t-1}) = \ell$ if and only if

$$L - \mu_t(m_t, c_t)S + \mathbb{E}\left[\sum_{\tilde{t}=t+1}^{\infty} \delta^{(\tilde{t}-t)} v(g_{\tilde{t}}, \theta_{\tilde{t}}) \mid g_t = \ell, \sigma^*\right] \geq \mathbb{E}\left[\sum_{\tilde{t}=t+1}^{\infty} \delta^{(\tilde{t}-t)} v(g_{\tilde{t}}, \theta_{\tilde{t}}) \mid g_t = i, \sigma^*\right];$$

(ii) for $g_t \in \{i, \ell\}$, then, for all censorship policies c' ,

$$\mathbb{E}\left[\sum_{\tilde{t}=t+1}^{\infty} \delta^{\tilde{t}-t} R \mathcal{X}_{g_t}(\{g_{t'}\}_{t+1}^{\tilde{t}}) \mid c_{t+1} = c_{g_t}^*, \sigma^*\right] \geq \mathbb{E}\left[\sum_{\tilde{t}=t+1}^{\infty} \delta^{\tilde{t}-t} R \mathcal{X}_{g_t}(\{g_{t'}\}_{t+1}^{\tilde{t}}) \mid c_{t+1} = c', \sigma^*\right],$$

where

$$\mathcal{X}_{g_t}(\{g_{t'}\}_{t+1}^{\tilde{t}}) = \begin{cases} 1 & \text{if } g_{t'} = g_t \text{ for all } t' \in \{t+1, \dots, \tilde{t}\}, \\ 0 & \text{otherwise;} \end{cases}$$

(iii) the voter's belief, $\mu_t(m_t, c_t)$, is derived using Bayes' rule, i.e.,

$$\begin{aligned} \mu_t(1, c_t) &= \mu_t^*(1, c_t) := \Pr[\theta_t = 1 \mid m_t = 1, c_t] \\ &= \frac{\pi[c_t(1)q + c_t(0)(1-q)]}{\pi[c_t(1)q + c_t(0)(1-q)] + (1-\pi)[c_t(1)(1-q) + c_t(0)q]}, \end{aligned}$$

and

$$\begin{aligned} \mu_t(0, c_t) &= \mu_t^*(0, c_t) := \Pr[\theta_t = 1 \mid m_t = 0, c_t] \\ &= \frac{\pi[(1-c_t(1))q + (1-c_t(0))(1-q)]}{\pi[(1-c_t(1))q + (1-c_t(0))(1-q)] + (1-\pi)[(1-c_t(1))(1-q) + (1-c_t(0))q]}. \end{aligned}$$

E.2 Preliminaries

Let

$$V(g_t = \ell \mid \sigma) := \mathbb{E}\left[\sum_{\tilde{t}=t+1}^{\infty} \delta^{(\tilde{t}-t-1)} v(g_{\tilde{t}}, \theta_{\tilde{t}}) \mid g_t = \ell, \sigma\right]$$

and

$$V(g_t = i \mid \sigma) := \mathbb{E}\left[\sum_{\tilde{t}=t+1}^{\infty} \delta^{(\tilde{t}-t-1)} v(g_{\tilde{t}}, \theta_{\tilde{t}}) \mid g_t = i, \sigma\right]$$

denote the expected continuation payoffs from electing a liberal and illiberal government, respectively. We define the *accountability cost of illiberalism* as

$$A(\pi, q, L, S, \delta \mid \sigma^*) := V(g_t = \ell \mid \sigma^*) - V(g_t = i \mid \sigma^*).$$

Notice that $A(\pi, q, L, S, \delta \mid \sigma^*)$ is independent of t since $V(g_t = \ell \mid \sigma^*) = V(g_{t'} = \ell \mid \sigma^*)$ and $V(g_t = i \mid \sigma^*) = V(g_{t'} = i \mid \sigma^*)$ for all non-negative integers t, t' .

It will be useful to introduce terminology and notation for two censorship policies. The *full censorship* policy refers to any of the completely uninformative mappings and, abusing notation slightly, is denoted by c_F . The *no-censorship* policy is the most-informative mapping $c(s(\theta_t)) = s(\theta_t)$ and is denoted by c_N .

Lemma E.1 says that in equilibrium the voter elects an illiberal government if and only if the expected cost of security, $\mu_t(m_t, c_t)S$, is greater than the sum of the value of liberty, L , and the discounted accountability cost of illiberalism, $\delta A(\pi, q, L, S, \delta \mid \sigma^*)$. The lemma is split in two parts because, whenever indifferent between either government, the voter reelects the incumbent; otherwise, the proof of Lemma E.1 is similar to the proof of Lemma 1.

Lemma E.1 *In every equilibrium,*

- *if $g_{t-1} = i$, the voter elects the illiberal government in period t if and only if*

$$\mu_t(m_t, c_t)S \geq L + \delta A(\pi, q, L, S, \delta \mid \sigma^*).$$

- *if $g_{t-1} = \ell$, the voter elects the illiberal government in period t if and only if*

$$\mu_t(m_t, c_t)S > L + \delta A(\pi, q, L, S, \delta \mid \sigma^*).$$

We now turn to the incumbent government's problem. Lemma E.2 establishes that the incumbent government's problem can be reduced to the problem of choosing the censorship policy that maximizes their probability of being reelected in the following period. Standard Bayesian persuasion results then allow for the optimal censorship policy to be characterized. After substituting "illiberal government" with "incumbent government," the proof of Lemma E.2 follows from the proofs of Lemmas 2 and 3.

Lemma E.2 *In every equilibrium, in period t , the period- t government, $g_t \in \{i, \ell\}$, chooses a censorship policy $c_{t+1} = c^*$ that maximizes its reelection probability in period $t + 1$. Furthermore, in equilibrium, if an incumbent government $g_t \in \{i, \ell\}$ is not indifferent between all censorship policies and can't guarantee their reelection, then*

1. *they choose $c_{t+1} = c_i^*$ such that $c_i^*(1) = 1$ and $c_i^*(0) \in [0, 1]$:*

$$\mu_i^*(1, c_i^*)S = L + \delta A(\pi, q, L, S, \delta \mid \sigma^*)$$

if $g_t = i$, and

2. they choose $c_{t+1} = c_\ell^*$ such that $c_\ell^*(0) = 0$ and $c_\ell^*(1) \in [0, 1]$:

$$\mu_t^*(0, c_\ell^*)S = L + \delta A(\pi, q, L, S, \delta \mid \sigma^*)$$

if $g_t = \ell$.

Lemma E.3 says that, in equilibrium, the accountability cost of illiberalism is zero.

Lemma E.3 *In every equilibrium, the accountability cost of illiberalism $A(\pi, q, L, S, \delta \mid \sigma^*)$ is zero.*

Proof. Suppose we have an equilibrium σ^* . Notice that exactly one of the following is true:

- (i) $\pi S > L + \delta A(\pi, q, L, S, \delta \mid \sigma^*)$;
- (ii) $\pi S < L + \delta A(\pi, q, L, S, \delta \mid \sigma^*)$;
- (iii) $\pi S = L + \delta A(\pi, q, L, S, \delta \mid \sigma^*)$.

Furthermore, notice that if $c_t = c_F$, then $\mu_t(m_t, c_t) = \pi$ for all messages $m_t \in \text{Supp}(c_F)$.

In case (i), it is immediate that full censorship ($c_t = c_F$) guarantees the illiberal government reelection. Therefore, by Lemma E.2, the illiberal government's equilibrium strategy must guarantee its reelection forever and $V(g_t = i \mid \sigma^*) = 0$. Turning to the liberal government, there are two cases to consider. If $\mu_t(0, c_N)S > L + \delta A(\pi, q, L, S, \delta \mid \sigma^*)$, then for all censorship policies c and messages $m_t \in \text{Supp}(c)$

$$\mu_t(m_t, c)S > L + \delta A(\pi, q, L, S, \delta \mid \sigma^*).$$

It follows that, when $g_{t-1} = \ell$, the liberal government is reelected with probability zero and, hence,

$$V(g_t = \ell \mid \sigma^*) = \delta V(g_{t+1} = i \mid \sigma^*) = 0.$$

Otherwise, $\mu_t(0, c_N)S \leq L + \delta A(\pi, q, L, S, \delta \mid \sigma^*)$. Notice that the liberal government can't guarantee their own reelection because $\pi S > L + \delta A(\pi, q, L, S, \delta \mid \sigma^*)$. Given Lemma E.2, it follows that, when $g_{t-1} = \ell$, the liberal government chooses an optimal censorship policy

c_ℓ^* such that, when $m_t = 0$, the voter chooses $g_t = \ell$ and is indifferent between electing either government and, when $m_t = 1$, the voter chooses $g_t = i$. Thus,

$$\begin{aligned}
V(g_t = \ell \mid \sigma^*) &= \Pr[m_t = 0 \mid c_t = c_\ell^*](L - \mu_t^*(0, c_\ell^*)S + \delta V(g_{t+1} = \ell \mid \sigma^*)) + \Pr[m_t = 1 \mid c_t = c_\ell^*]\delta V(g_{t+1} = i \mid \sigma^*) \\
&= \Pr[m_t = 0 \mid c_t = c_\ell^*]\delta V(g_{t+1} = i \mid \sigma^*) + \Pr[m_t = 1 \mid c_t = c_\ell^*]\delta V(g_{t+1} = i \mid \sigma^*) \\
&= \delta V(g_{t+1} = i \mid \sigma^*) \\
&= 0.
\end{aligned}$$

In both cases, we conclude that $A(\pi, q, L, S, \delta \mid \sigma^*) = 0$.

In case (ii), a similar but mirrored argument as case (i) can be used to show that $A(\pi, q, L, S, \delta \mid \sigma^*) = 0$.

In case (iii), it is immediate that full censorship ($c_t = c_F$) guarantees the incumbent government reelection (regardless of whether it is the liberal or illiberal government). Thus, by Lemma E.2, in equilibrium, the incumbent government must maintain power forever. Therefore, $V(g_t = i \mid \sigma^*) = 0$ and

$$V(g_t = \ell \mid \sigma^*) = \frac{L - \pi S}{1 - \delta}.$$

Substituting these two equations into the case (iii) condition gives

$$\pi S = L + \delta \frac{L - \pi S}{1 - \delta} \iff L = \pi S;$$

hence, $A(\pi, q, L, S, \delta \mid \sigma^*) = 0$. ■

Lemmas E.2 and E.3 immediately imply:

Corollary E.1 *In every equilibrium,*

1. *if $\pi S > L$ and $g_{t-1} = i$, the illiberal government is reelected with probability 1;*
2. *if $\pi S < L$ and $g_{t-1} = \ell$, the liberal government is reelected with probability 1;*
3. *if $\pi S = L$ and $g_{t-1} \in \{i, \ell\}$, the incumbent government is reelected with probability 1.*

E.3 Equilibrium characterization

Lemma E.4 characterizes the probability of an illiberal government coming to power.

Lemma E.4 *Suppose $g_{t-1} = \ell$. In every equilibrium,*

- (i) If $\mu_t^*(1, c_N)S < L$, the illiberal government is elected in period t with probability zero;
- (ii) If $\pi S \leq L \leq \mu_t^*(1, c_N)S$, the illiberal government is elected in period t with probability zero;
- (iii) If $\mu_t^*(0, c_N)S \leq L < \pi S$, the illiberal government is elected in period t with probability

$$c_\ell^*(1)(\pi q + (1 - \pi)(1 - q)) \in (0, 1)$$

where $c_\ell^*(1) \in (0, 1]$ satisfies

$$\frac{\pi[q(1 - c_\ell^*(1)) + (1 - q)]}{\pi[q(1 - c_\ell^*(1)) + (1 - q)] + (1 - \pi)[(1 - q)(1 - c_\ell^*(1)) + q]} = L;$$

- (iv) If $L < \mu_t^*(0, c_N)S$, the illiberal government is elected in period t with probability one.

Proof. For (i) and (ii), the parameter restrictions imply that $\pi S \leq L$ and, by Corollary E.1, the liberal government is reelected with probability one. Therefore, whenever $g_{t-1} = \ell$, the illiberal government is elected with probability zero.

For (iii), notice that $L < \pi S$ and $\mu_t^*(0, c_N)S \leq L$ imply

$$\mu_t^*(0, c_N)S \leq L < \mu_t^*(1, c_N)S.$$

Furthermore, by Corollary E.1, the illiberal government is reelected with probability one and, by Lemma E.3, $A(\pi, q, L, S, \delta \mid \sigma^*) = 0$. Therefore, when $g_{t-1} = \ell$, the liberal government is reelected if and only if

$$\mu_t^*(m_t, c_t)S \leq L.$$

Given Lemma E.2, the liberal government's optimal censorship policy is such that $c_\ell^*(0) = 0$ and $c_\ell^*(1) \in (0, 1]$ such that $\mu_t^*(0, c_\ell^*)S = L$, i.e.,

$$\frac{\pi[q(1 - c_\ell^*(1)) + (1 - q)]}{\pi[q(1 - c_\ell^*(1)) + (1 - q)] + (1 - \pi)[(1 - q)(1 - c_\ell^*(1)) + q]} = L.$$

Therefore, the illiberal government is elected with probability

$$c_\ell^*(1) \Pr[s(\theta_t) = 1] = c_\ell^*(1)(\pi q + (1 - \pi)(1 - q)) \in (0, 1).$$

For (iv), the parameter restriction implies that $L \leq \pi S$ and, by Corollary E.1, the illiberal government is reelected with probability one. Furthermore, by Lemma E.3, $A(\pi, q, L, S, \delta \mid$

$\sigma^*) = 0$. Therefore, when $g_{t-1} = \ell$, the liberal government is reelected if and only if

$$\mu_t^*(m_t, c_t)S \leq L.$$

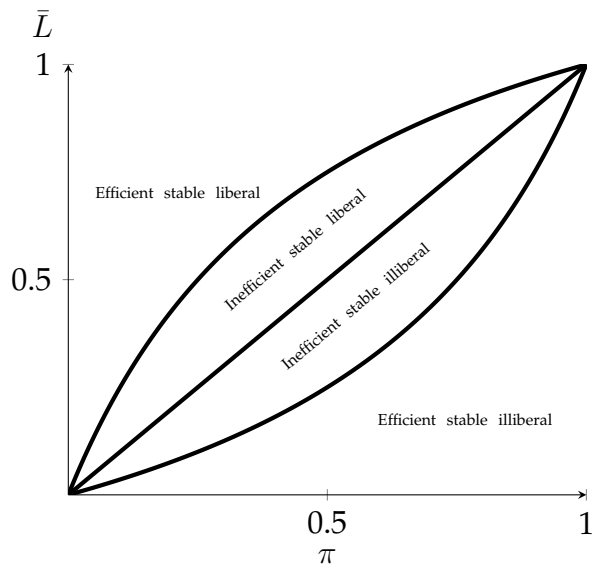
Because $L < \mu_t^*(0, c_N)S$, for any censorship policy c_t and any message $m_t \in \text{Supp}(c_t)$, the above inequality is not satisfied and the illiberal government is elected with probability one. ■

Combining Corollary E.1 and Lemma E.4 provides the following equilibrium characterization,⁴⁵ which is illustrated in Figure 5.

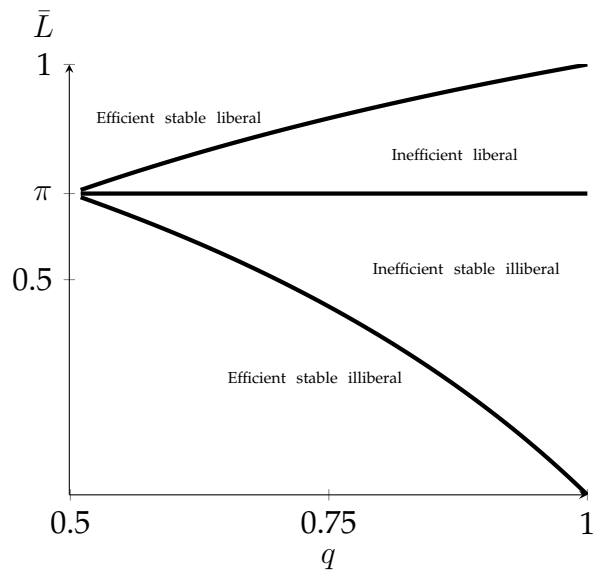
Corollary E.2 *In the essentially unique equilibrium,*

1. $L < \mu_t^*(0, c_N)S$ induces an efficient stable illiberal democracy;
2. $\mu_t^*(0, c_N)S \leq L < \pi S$ induces an inefficient stable illiberal democracy in which an illiberal government is elected with strictly positive probability and, once elected, maintains power forever;
3. $\pi S \leq L \leq \mu_t^*(1, c_N)S$ induces an inefficient stable liberal democracy in which the liberal government maintains power forever;
4. $\mu_t^*(1, c_N)S < L$ induces an efficient stable liberal democracy.

⁴⁵For brevity's sake, we omit the proof of equilibrium existence. Given the proofs of Lemmas E.3 and E.4, it is straightforward to construct an equilibrium for any set of parameters.



(a) $\delta = 0.9$ and $q = 0.75$.



(b) $\delta = 0.9$ and $\pi = 0.7$.

Figure 5: Typology of regimes.