The Emergence of Political Accountability[*]

Chris Bidner[†]   Patrick Francois[‡]

February 14, 2012

Abstract

Many factors are necessary for institutions to be effective in holding political leaders accountable to the interests of their citizens. A common delineation of such factors draws a distinction between formal constraints, like elections, that display clear rules and punishments for their violation, and less formal ones, like norms, that prescribe 'appropriate' acts and the customary consequences for not performing them. A promising stream of literature conceives of such informal norms as equilibria of a political agency game played between citizens and their leaders. Norms, and the punishments that support them, are then seen as mutually reinforcing outcomes of play in a setting where the formal structures do not fully capture all contingencies. However, previous formalizations along these lines yield little insight into the process by which norms (and the accountability institutions they imply) change. The present paper analyzes the process by which political norms can be made to change. Leaders play a key role. The acts of leaders convey information to imperfectly informed citizens about the underlying state of other citizens' willingness to tolerate political transgressions. Good political acts indicate this tolerance is likely to be low, both today and in future, and can trigger a transition in political institutions from prevailing norms of political permissiveness – where transgressions are widespread – to non-permissiveness – where they are rare. We show that the transitions modeled here can be used to understand widespread, but heretofore informal, concepts such as democratic capital, critical junctures, institutionalization and 'great' leaders.

Keywords: political norms, institutional change, accountability, voting.
JEL Codes: D72, P16, C73

[†]University of New South Wales, School of Economics. `c.bidner@unsw.edu.au`

[‡]University of British Columbia, Dept. of Economics, CEPR and CIFAR. `francois@interchange.ubc.ca`

1 Introduction

Political institutions are the means by which constraints of citizen accountability are imposed on political leaders. Such accountability constraints vary widely across countries. Well functioning democracies, at one extreme, feature institutions that tightly circumscribe their political leaders' discretion. Leadership constraints arise from recurring mechanisms, for example regular elections that are freely contested and fairly administered, that provide incentives for leaders to stay within proscribed bounds. However, many political systems still exhibit low accountability despite featuring such explicit democratic mechanisms. This is because, in addition to explicit rules and mechanisms, political 'norms' can also play a central role. Norms guide political actors on appropriate behaviour under contingencies that are necessarily little more than grey areas in the written rules. Strict political accountability generally corresponds with normative prohibitions on personal enrichment in office, promotion of relatives, criminality, scandal, respect for bureaucratic/legislative independence, etc. Functional democracies prescribe harsh punishments (electoral or otherwise) for leaders seen to have violated these norms.

One way of thinking about norms is as stationary equilibria of the political game being played between rulers and their citizens. This approach is central to formal models developed along such lines by Weingast (1997) and Myerson (2006). Both of these contributions demonstrate the possibility of widely varied, self-reinforcing political norms as outcomes. *Non-permissive* norms are one type of outcome: citizens stand ready to depose leaders not behaving accountably. Consequently leaders, fearing citizen reprisal, behave accountably. Citizens' threat of reprisal is credible because they rationally expect replacement leaders to behave accountably. Conversely, *permissive* norms describe a similarly self-reinforcing pattern whereby citizens are unwilling to stand against unaccountable leaders, who therefore act with impunity. The permissive behaviour is a best response since citizens are unwilling to incur the cost of leadership change (here modeled as an incumbency advantage) if the incumbent is likely to be replaced by another unaccountable leader.[1]

Approaches based on norms as stationary equilibria of a game have been useful in explaining the stability of both permissive and non-permissive political norms, but less so in providing theories of how non-permissive political norms may emerge. Such models do allow for norms to change, but changes typically stem from factors external to the model – a focal point shifts, exogenous changes make an equilibrium unstable, or a sun-spot type of transition occurs.

The present paper develops a theory of the emergence of political accountability where changes in political norms play a central role. Consistent with the theoretical tradition above, norms correspond to outcomes of a political game between leaders and cit-

---

[1] Below, we discuss the literature on political agency models in more detail and contrast it with the approach taken here. It is perhaps important to point out that in our model the nature of the underlying coordination problem is similar to that in Myerson (2006) and is rather more indirect than the more standard one whereby individual citizens are unwilling to engage in a costly struggle unless the struggle is likely to succeed (i.e. unless sufficiently many other individuals engage in the struggle); Weingast (1997), Fearon (2011), Persson and Tabellini (2009).

izens. An increase in political accountability occurs when non-permissive norms replace permissive ones, thereby forcing self-interested political actors to become accountable to their citizens. We demonstrate how such a transition endogenously arises in equilibrium, and in doing so provide a micro-foundation for many concepts that a largely informal literature on political transformations has argued to be important.

Central to our theory of accountability's emergence is the power of ideas. Ideas can grip a population and motivate mass action, and beliefs in the salience of a particular idea can change rapidly. Sometimes, as Kingdon (1995) argues, an "idea whose time has come" will be impossible to withstand. A population can become gripped by a particular idea. For example, a broad desire for democracy in an otherwise autocratic system, or an end to political corruption within a democracy featuring endemic corruption, or by a demand for responsive decision making when executives have previously been unconstrained. An obvious recent example of widely felt dramatic change in the salience of particular ideas is evidenced by the revolutionary wave of demonstrations and street protests that started in the Arab World (the "Arab Spring"). These widespread and dramatic shifts in the prominence of ideas – in this case, demands for political freedoms and democratic accountability – correspond precisely to our notion of a population becoming 'gripped'. Such ideas are not new to inhabitants of the region, but their salience to the mass of citizens increased dramatically in a short period of time, and clearly not due to any evident changes in structural conditions.

We explicitly allow such exogenous changes in the ideas held by citizens to occur. We show how this may lead to a type of 'bottom-up' transition to political accountability. When enough citizens become temporarily gripped by the importance of the need for accountability, and are willing to oppose non-accountable acting leaders, these leaders are not only removed, but their removal may usher in a permanent change in the political institution. Future leaders will be forced to act accountably – under threat of removal from office – even if the population reverts back to being no longer gripped by the importance of the accountability idea per se.

But the emergence of political accountability need not be driven by such dramatic changes. Indeed, many examples where polities have transitioned to tighter norms of political accountability occurred only gradually, after a long period of consolidation.[2] In these, leaders with vision, temperance and foresight seemed to have ushered in permanently improved accountability norms, through their own acts. These leaders were seen to have acted at "critical junctures", in ways that somehow seemed to embed their personal ideals on to their countries' institutions, with effects persisting long after their departure. We provide an explanation of these leader-driven 'top-down' transitions to political accountability as well. In these, we will see that idealistic leaders are able to leave a legacy of accountability that outlives them.

What we call 'top-down' transitions feature no changes in citizens' ideas about the importance of accountability per se, but instead changes in citizens' beliefs. Citizens

---

[2]The literature on 'democratization' has identified this latter phase (i.e., after deposing the autocrat, and implementing elections) as arguably the more difficult, and harder to understand, component in the process of democratic transitions. This is sometimes referred to as democracy's consolidation phase.

never know how widely shared a particular idea is by their compatriots, but they have beliefs about that. And the actions of political leaders provide them with clues that inform these beliefs. If prevailing political norms have been permissive, thus allowing leaders to act freely without censure, then counter-norm behaviour – a leader acting accountably – leads to some revision in citizens' beliefs. It could be that the leader is an inherently idealistic type and so acts accountably irrespective of prevailing political norms. Alternatively, even a non-idealistic but rational politician sensing a "bottom-up" change in people's ideas, i.e., facing a populous now widely valuing accountability per se, will self-interestedly act accountably. Since citizens can never know a politician's type for certain, they put some weight on both possibilities. After seeing counter-norm accountable behaviour, a citizen thus believes more strongly that his compatriots share widely held accountability ideals. Consider now a sequence of such accountable acting leaders. If long-enough, this sequence may eventually raise beliefs about the importance of accountability in one's compatriots' minds to such a point that one thinks any self-interested leader is very unlikely to act unaccountably. When this happens, one rationally chooses to oppose unaccountable acting leaders, thereby further weakening the incentive for rational leaders to act unaccountably. Thus, even non-idealistic leaders fearing reprisal from rational citizens will self-interestedly choose to respect accountability ideals. This can lead to a permanent change in the norms governing a political system.

Our model delivers a theory of political leadership. Transitions to political norms of accountability are engineered by 'great leaders' who change popular beliefs through their actions. The beliefs that they change are those that citizens hold about the importance of particular ideas in the minds of their compatriots. By acting accountably, leaders leverage the *possibility* of accountability ideals being widespread in the population, to convince rational citizens that they probably are. When so convinced, accountability may indeed become a political necessity. We establish conditions under which this type of change necessarily occurs even if citizens operate under the most pessimistic belief that it will never do so, and even if the chance of the population becoming 'gripped' is arbitrarily low. Importantly, leaders achieve this even if no citizen has changed her own personal idea about the value of accountability per se.

In our paper, rational politicians act according to electoral incentives, and citizens *may* have incentive to punish transgressing leaders because their transgressions *may* imply that they are a bad type. It fits in the standard rubric of political agency models where both moral hazard and selection operate.[3] Endemic to such models is the possibility of equilibria where moral hazard is unconstrained and leaders are effectively unaccountable.[4] We take this possibility seriously here, and will focus on transitions where

---

[3]See Ashworth *et al.* (2010). Early contributions in the moral hazard strand of this literature were provided by Key (1966), Barro (1973), Fiorina (1981), Ferejohn (1986), Austen-Smith and Banks (1989). The citizen selection branch is most often associated with Banks and Sundaram (1993) and Banks and Sundaram (1998), Alesina and Rosenthal (1995), Fearon (1999), Canes-Wrone *et al.* (2001), Ashworth (2005), Maskin and Tirole (2004) and Besley (2006).

[4]In this case, ignoring these to focus on the optimal (from the voter's perspective) one where the agent's moral hazard is punished is the approach taken in Barro (1973), Ferejohn (1986), and Persson and Tabellini (2000). As

equilibrium play involves a movement away from situations where leaders are unaccountable to ones where accountability arises from voters rationally punishing morally hazardous leaders.

This focus brings us closer to papers like Weingast (1997) and Myerson (2006). Essentially, both of these contributions take as their starting point the possibility of both non-permissive and permissive equilibria in political agency models. Both of these authors focus on factors external to the model that change the game's focal equilibrium. Weingast (1997) argues that the Glorious Revolution of 1688 in which William of Orange was brought onto the English throne, is one such set of external factors. The powers he assumed on ascending the throne after the deposed Stuart monarchy were relatively strongly circumscribed by the parliament. He argues this is because the Glorious Revolution represented a coordinated move from the previously permissive equilibrium to a non-permissive one. A different set of equilibrium changing factors are considered by Myerson (2006), who shows that electoral accountability depends on voter optimism which can be altered by a federalist structure. These models are useful in elucidating how a set of events or features – the glorious revolution, implementation of a federalist structure – could allow citizens to coordinate on a better equilibrium of the game (from citizen's perspective). But nothing within the models themselves illuminate the factors giving rise to such transitions. In these models transition from permissive to non-permissive outcomes have the flavour of sun-spots – the prime movers are external to the model.[5] In contrast, here transitions from permissive to non-permissive outcomes themselves are characterized along the equilibrium path, and we are able to say something about what causes them. We fully characterize a class of transition equilibria where the possibility of norm change is rationally anticipated by citizens. Transitions from non-accountable to accountable norms are affected by citizens' optimism about the possibility of change. But we show that even if citizens hold the most pessimistic beliefs about such positive changes occuring (i.e., that norms will never change), these cannot be sustained, implying that some sort of transition in norms must occur. In all such transitions to accountability, the actions of political leaders are key to making them happen, and can also provide long-lasting constraints on future leaders.

Our focus on political transitions shares in common with a large literature on 'democratization' an interest in factors causing improvement in political institutions. This literature typically focuses on the factors leading to wholesale changes in a polity, often from essentially autocratic and unrepresentative institutions to democratic ones.[6] In contrast, our focus is on changes in the norms of accountability that operate under given existing formal rules. The papers closest to ours in that literature are thus those interested in democracy's 'consolidation', as these investigators have emphasized the changes in soft features of the political system (norms, attitudes, beliefs) needing to follow upon

Besley (2006) persuasively argues, this type of argument for accountability deriving from moral hazard is fragile.

[5]Though we do not wish to argue that external factors are unimportant. In addition to these examples, Fearon (2011) makes a compelling argument that elections may well represent another such coordinating device helping to select non-permissive equilibria.

[6]For instance, see Acemoglu and Robinson (2000), (2001), and (2006), Persson and Tabellini (2009), Boix (2003), Fearon (2011), Przeworski *et al.* (1999), and Brender and Drazen (2009).

the formal changes (elections, a constitution) for democracy to consolidate. This literature asks why it is that polities adopting democracy's formal structures don't all end up functioning democratically. Factors that this literature emphasizes, such as time, experience and leadership, are all factors that will play an important role in our model. Therefore one way of viewing our contribution is in providing a theoretical underpinning to a more informal literature that has argued from examples for why these features matter.[7] We return to that literature after our main results.

The prominent role of ideas as catalysts of change in our model – both directly in bottom-up transitions to accountability, and indirectly through top-down transitions – is consistent with a stream of 'ideational' research in political science. Widespread ideas play a central role in Almond and Verba (1963), Moore (1967), Putnam et al. (1994), Diamond (1999), Linz and Stepan (1996).[8] Sometimes this role is quite direct: for example Barrington Moore's influential Social Origins of Democracy and Dictatorship (1967) that argues transitions occur because the politically powerful simply come to internally value democracy.[9] But the recent trends in this literature have emphasized the conceptual holes that pervade such direct explanations. Ideas are not salient to political change at all times – sometimes being widely held without precipitating change, and at others being absent from the change process altogether. Often ideas need to be instantiated by leaders, and great leadership only seems to have such a capacity at what has come to be called a "critical juncture". As Thelen (1999) argues, this literature has typically been weak in specifying the mechanisms that translate critical junctures into lasting political legacies. Our theory provides one such mechanism.

It has long been contended that political leadership matters for the development of functional political institutions. Recent empirical evidence (identified from random leadership changes) from Jones and Olken (2009) seems to confirm this. The importance of leaders is also reiterated in studies focused on the roles of political elites in the process of democratization – Di Palma (1990), O'Donnell et al. (1986), Przeworski (1991) and Rustow (1970). These studies emphasize leader agency: getting the right type of leader to usher in change; especially early in a new system. In providing a theory of leadership, our work is close in aim to Acemoglu and Jackson (2011). They are similarly interested in understanding how recurrent patterns of behavior in social and political contexts like social norms can be changed by the actions of leaders. Their context features overlapping generations of agents playing a symmetric coordination game in which a player's strategy is fixed by initial play. Players imperfectly observe past play, and use

---

[7]Our theory also provides an explanation for reversals – i.e., a failure of consolidation, and why democracies are more vulnerable to reversal when not long established (see Linz and Stepan (1978), Gasiorowski and Power (1998), Bernhard et al. (2003) and Brender and Drazen (2009)).

[8]Tangentially related is the literature relating accountability to information; Besley and Burgess (2002) and Snyder and Stromberg (2010) study this relationship empirically.

[9]A discussion of this literature, contrasting its approach with that of institutionalists, highlighting the complementarity between the approaches, and suggesting frictions between institutional structures and prevailing political ideas in explaining political change is developed in Lieberman (2002). Though there is not much empirical evidence on this, some evidence suggests the importance of democratic ideals for the consolidation of, rather than transition into, democracy (Persson and Tabellini (2009)).

these observations to infer the action taken by the previous player with whom they will initially interact. Leaders are agents endowed with permanent visibility of their actions to all future players, which allows them to enact a type of coordinated change in future play through their own visible acts. By playing 'good' and being seen to do so, a leader can induce good play from her immediate follower. Though subsequent successors will not perfectly observe the follower's actions, seeing the leader's allows them to infer that these were probably also good and can engineer coordination – in general only temporarily – on a 'good' social outcome. The context of our analysis is markedly different as it is directly rooted in a standard political agency framework: leaders and citizens are not symmetric players – leaders directly affect citizens payoffs through their actions, citizens periodically decide whether to support or replace leaders. The coordination aspect of our set-up derives from the interaction between citizens views regarding political transgressions today, and those views tomorrow. In our framework, institutional change derives from leaders' abilities to change beliefs about those views. Great leaders, by themselves acting well, raise expectations about the standards compatriots demand of leaders both today and in future. If raised enough, this can induce change (perhaps permanent) in civic institutions. The important function of leaders early in democratic transitions in changing voter expectations is the focus of Svolik (2010). According to his theory, a sequence of bad leaders early on can precipitate a slide back to autocratic rule by leading voters to believe that generally self-serving types will come to lead in democracies.[10]

The remainder of the paper is organized as follows. Section 2 lays out the model and results, beginning with a baseline version in which citizens are never gripped by ideals of accountability, then extending this to include such a feature. Various aspects of the model and results are discussed in Section 3, including a description of how our results illuminates informal notions of political change identified in the literature, and a discussion of how changes in modeling assumptions would (not) affect the main results. Conclusions are drawn in Section 4.

2 Model

2.1 Fundamentals

We consider an economy that unfolds in discrete time and is populated by two classes of agent: politicians and citizens. There is a large pool of politicians and a continuum of infinitely-lived citizens. Politicians are one of three privately known types: autocratic, democratic, or rational. It is common knowledge that the proportion of autocratic types is $\sigma_A > 0$, of democratic types is $\sigma_D > 0$, and of rational types is $1 - \sigma_A - \sigma_D > 0$.

One politician is in power at any given date $t \in \{0, 1, 2, ...\}$. Once a politician enters office, they decide whether to transgress (T) or to not transgress ($\tilde{T}$). A strategy for a politician is therefore $a \in \{T, \tilde{T}\}$.[11] Observing the action chosen by the politician, citizens

---

[10]The idea that current behaviour of members of a group (here politicians) is influenced by the behaviour of predecessors is present in Tirole (1996), and is also more closely related to Acemoglu and Jackson (2011).

[11]Although we do not allow politicians to change their action during their incumbency, it will become clear that this is for simplicity and is not restrictive in the sense that politicians would never want to switch their

decide whether to support (S) or not support ($\tilde{S}$) the politician. If the politician does not receive sufficient support from citizens, then they are removed from office (and never return). Otherwise, they return to office the following period with probability $\delta \in (0,1)$. Specifically, politicians need the support of at least proportion $z \in (0,1)$ of citizens.[12]

Autocratic types always transgress and democratic types never transgress. Rational types weigh up the costs and benefits. Specifically, politicians get a payoff normalized to zero when not in power and, while in power, action $a \in \{T, \tilde{T}\}$ produces a per-period payoff of $u(a)$ where $u(T) > u(\tilde{T}) > 0$. Citizens get a per-period payoff of $1 + D_I \cdot \alpha - D_T \cdot c$, where $D_I \in \{0,1\}$ is an indicator function that takes a value of 1 if and only if the politician is an incumbent (i.e. was in power in the previous period) and $D_T \in \{0,1\}$ is an indicator function that takes a value of 1 if and only if the politician transgresses. That is, $\alpha > 0$ is an incumbency advantage and $c > 0$ is the cost that a transgression imposes on citizens. Politicians and citizens maximize the present value of expected future payoffs using a discount factor of $\beta \in (0,1)$.

If politicians were supported regardless of whether they transgress, then they will always transgress. The interesting case is when citizen support is conditional on non-transgression. In this case, a politician's value to not-transgressing is $u(\tilde{T})/(1-\beta\delta)$ whereas the value to transgressing is $u(T)$. To rule out the uninteresting case where politicians always transgress regardless if they are supported, we make the following assumption.

Assumption 1. *The benefits from transgressing are not too great relative to the effective discount factor:*

$$\frac{u(T)}{u(\tilde{T})} < \frac{1}{1-\beta\delta}. \tag{1}$$

We know that politicians transgress if citizens always support transgressors and this assumption ensures that politicians will not transgress if citizens never support transgressors.

2.2 Political Norms and Stationary Equilibria

It is well-known that a wide range of behavior is typically supportable as equilibria in repeated games. We focus on behavior that seems reasonable insofar as it accords with our notion of a (political) 'norm': behavior that displays a degree of stability over time. In this section we begin with the most natural description of norm-like behavior - stationary strategies. That is, behavior in which the probability that rational politicians choose to transgress and the probability that citizens support politicians conditional on each of their two actions, is the same across time. We also focus on symmetric pure strategies. Specifically, let $p(a)$ be the probability with which citizens support a politician that takes action $a \in \{T, \tilde{T}\}$, and let q be the probability with which rational politicians transgress.

For citizens, the value to supporting a transgressor, $V(T)$, satisfies

$$V(T) = 1 + \alpha - c + \beta[\delta \cdot p(T) \cdot V(T) + (1 - \delta \cdot p(T)) \cdot \bar{V}], \tag{2}$$

chosen action.

[12]Since we focus on symmetric strategies, the value of z will not really matter: either all citizens support or none do. The value of z will matter when describing gripped citizens below.

where $\bar{V}$ is the expected value associated with an entrant politician. That is, the flow pay-off includes the incumbency advantage (since they are an incumbent in the future) net of the transgression burden, and the continuation payoff reflects that the politician is returned to power the following period with probability $\delta \cdot p(T)$, otherwise a new politician comes to power. Similarly, the value to supporting a non-transgressor, $V(\tilde{T})$, satisfies

$$V(\tilde{T}) = 1 + \alpha + \beta[\delta \cdot p(\tilde{T}) \cdot V(\tilde{T}) + (1 - \delta \cdot p(\tilde{T})) \cdot \bar{V}]. \tag{3}$$

Letting $\rho \equiv \sigma_D + (1 - \sigma_D - \sigma_H) \cdot q$ be the probability that a randomly drawn politician will transgress, we have that the value associated with an entrant politician satisfies

$$\bar{V} = \rho \cdot V(T) + (1 - \rho) \cdot V(\tilde{T}) - \alpha. \tag{4}$$

The "$-\alpha$" term reflects the fact that the entrant will not have an incumbency advantage. Solving these yields the three value functions $\{V(T), V(\tilde{T}), \bar{V}\}$; see section A.1 for details.

Citizens support a politician that takes action a if and only if $V(a) \geq \bar{V}$. Citizens always support a non-transgressor[13], and therefore all equilibria must have $p(\tilde{T}) = 1$. Given this, let $p \equiv p(T)$ from here on in order to simplify notation.

Given that politicians transgress if agents support transgressors and do not transgress if agents do not support transgressors, the two possible (pure-strategy) stationary equilibria are $(p = 1, q = 1)$ and $(p = 0, q = 0)$. The former is described as a case of 'permissive norms' and the latter is a case of 'non-permissive norms'. To derive the conditions under which these are equilibria, we derive the net benefit from supporting a transgressor. By subtracting $\bar{V}$ from both sides of (2) and re-arranging, we have that the net benefit to supporting a transgressing politician is

$$V(T) - \bar{V} = \frac{1 + \alpha - c - (1 - \beta) \cdot \bar{V}}{1 - \beta\delta \cdot p}. \tag{5}$$

The sign of this depends only on the numerator. The value of $\bar{V}$ depends on (p, q): let $\bar{V}^+$ be this value when $p = q = 0$ (non-permissive norms) and $\bar{V}^-$ be this value when $p = q = 1$ (permissive norms). That is,

$$\bar{V}^+ \equiv \frac{1 + \alpha\beta\delta - (c + \beta\delta \cdot (1 + \alpha - c)) \cdot \sigma_A}{(1 - \beta)(1 - \beta\delta\sigma_A)} \tag{6}$$

$$\bar{V}^- \equiv \frac{1 + \alpha\beta\delta - (1 - \sigma_D) \cdot c}{1 - \beta}. \tag{7}$$

Given these, we are ready to state the first proposition.

Proposition 1. *There exists a stationary equilibrium with non-permissive norms if and only if $\frac{\alpha}{c} \cdot (1 - \beta \cdot \delta) \leq 1 - \sigma_A$. There exists a stationary equilibrium with permissive norms if and only if $\frac{\alpha}{c} \cdot (1 - \beta \cdot \delta) \geq \sigma_D$.*

All proofs are in the appendix.

Existence of equilibrium is a direct consequence of this.[14] There is also an equilibrium in stationary mixed strategies, but we ignore this because it is not stable (e.g. a slight

---

[13]Subtracting $\bar{V}$ from both sides of (3) and re-arranging gives $V(\tilde{T}) - \bar{V} = (1 + \alpha - (1 - \beta) \cdot \bar{V})/(1 - \beta\delta p(\tilde{T}))$. The sign of this only depends on the numerator, and therefore on the sign of $(1 + \alpha)/(1 - \beta) - \bar{V}$. But this is positive since $(1 + \alpha)/(1 - \beta)$ is the value associated with getting a non-transgressing incumbent in every future period.

[14]By contradiction, non-existence would require $1 - \sigma_D < \frac{\alpha}{c} \cdot (1 - \beta \cdot \delta) < \sigma_H$, but this implies $\sigma_D + \sigma_H > 1$.

increase in the probability of supporting a transgressor raises the payoff to supporting a transgressor) and therefore uninteresting given our focus on transitions between norms. More interestingly, both (pure strategy) equilibria exist if the population of both behavioral types is sufficiently small that $\sigma_D \leq \frac{\alpha}{c} \cdot (1 - \beta \cdot \delta) \leq 1 - \sigma_A$.

2.2.1 Transitions

By focusing on stationary strategies, we have shown that there exist two equilibria with contrasting political norms. Our interest lies in understanding how political norms change, and therefore the conditions under which equilibrium behavior calls for transitions between these norms. While one could approach this by attempting to construct equilibria using more complicated strategies (based on sun-spots for instance), we instead enrich the model in order to retain our focus on norm-like strategies. In order to meaningfully focus on transitions between norms, we need to make some minimal assumption regarding the existence of multiple equilibria.

Assumption 2.

$$\sigma_D < \frac{\alpha}{c} \cdot (1 - \beta \cdot \delta) < 1 - \sigma_A. \tag{8}$$

2.3 Introducing Gripped Agents

2.3.1 Fundamentals

Now suppose that a proportion $z' > z$ of citizens occasionally become 'gripped' by the idea that transgressors should never be supported (not necessarily the same citizens over time). We model this by supposing that at any given date the economy exists in either a 'gripped' state (G) or a 'non-gripped' state ($\tilde{G}$). Since $z' > z$, a transgressing politician is not supported in the gripped state. We assume that the state is constant throughout a politician's incumbency, and that the politician in power observes the state. Citizens do not observe the state unless they are gripped and a transgression occurs, but of course, citizens may infer the state from observed outcomes. It is important for our results only that citizens are at an informational disadvantage with regards to the state vis a vis politicians, not that they are completely ignorant, nor that politicians are perfectly informed.[15]

Citizens' beliefs are updated at various points. Consider a new politician that comes to power. Citizens' prior belief regarding the probability of being in state G is denoted π_0. Once the action taken by the politician is observed, beliefs are updated to π_1. Then the support decision is made by citizens, and the result is observed. Beliefs are then once again updated to π_2. Beliefs will not change again until the politician leaves power. The state stays the same for the next politician with probability s. This parameter acts as our measure of *persistence* and will play an important role in what follows. With the remaining probability $1 - s$ a new state is drawn, in which case the drawn state is G with probability λ.

---

[15]This seems the most reasonable information assumption given the immense resource advantage and interest that politicians have in knowing the state – from focus groups, polling and party machinery.

Thus, if the probability of being in the G state is π_2 when a politician leaves office, the probability that their replacement is in G - i.e. the new prior - is $\pi_0' \equiv s \cdot \pi_2 + (1-s) \cdot \lambda$. The highest that this belief can be is $\overline{\pi}_0 \equiv s + (1-s) \cdot \lambda$ (i.e. when $\pi_2 = 1$) and the lowest that it can be is $\underline{\pi}_0 \equiv (1-s) \cdot \lambda$ (i.e. when $\pi_2 = 0$).

2.4 Political Norms

We begin by searching for stationary equilibria. As above, let p be the probability that rational voters support a transgressor, and q be the probability that a rational politician will transgress (in the non-gripped state).[16] In general, the relative payoff to supporting a transgressor depends on beliefs π_1 (recall that this is the belief that is updated according to the action taken by the politician). Let $V(\pi_1)$ be the expected value to supporting a transgressor and $\tilde{V}(\pi_1)$ be the expected value to not supporting a transgressor at beliefs π_1. These are given by

$$V(\pi_1) \equiv \pi_1 \cdot G(T) + (1-\pi_1) \cdot \tilde{G}(T) \tag{9}$$

$$\tilde{V}(\pi_1) \equiv \pi_1 \cdot E + (1-\pi_1) \cdot \tilde{E}, \tag{10}$$

where $G(T)$ is the value to citizens of having an transgressing incumbent in the gripped state, $\tilde{G}(T)$ is the value of having a transgressing incumbent in the non-gripped state, E is the value associated with drawing a new politician given that the current state is gripped, and $\tilde{E}$ is the value associated with drawing a new politician given that the current state is non-gripped. The net benefit to supporting a transgressor is therefore

$$V(\pi_1) - \tilde{V}(\pi_1) = \pi_1 \cdot [G(T) - E] + (1-\pi_1) \cdot [\tilde{G}(T) - \tilde{E}]. \tag{11}$$

The values of $(G(T), \tilde{G}(T), E, \tilde{E})$ are related according to the following. First, $G(T)$ comprises a flow payoff incorporating the incumbency advantage net of the transgression cost, but the continuation payoff is E since this transgression is not supported (due to the state being gripped). Therefore $G(T)$ satisfies

$$G(T) = 1 + \alpha - c + \beta \cdot E. \tag{12}$$

In contrast, $\tilde{G}(T)$ has a continuation value that reflects the fact that this politician is supported with probability p, and therefore returns to power with probability $p \cdot \delta$. Therefore $\tilde{G}(T)$ satisfies

$$\tilde{G}(T) = 1 + \alpha - c + \beta \cdot [p \cdot \delta \cdot \tilde{G}(T) + (1 - p \cdot \delta) \cdot \tilde{E}]. \tag{13}$$

The value to having a new politician in the gripped state, E, satisfies

$$E = \overline{\pi}_0 \cdot [\sigma_A \cdot G(T) + (1-\sigma_A) \cdot G(\tilde{T})] + (1-\overline{\pi}_0) \cdot [\rho \cdot \tilde{G}(T) + (1-\rho) \cdot \tilde{G}(\tilde{T})] - \alpha, \tag{14}$$

where $G(\tilde{T})$ is the value associated with having a non-transgressing incumbent in the gripped state and $\tilde{G}(\tilde{T})$ is the value associated with having a non-transgressing incumbent in the non-gripped state. That is, the entrant will be in the gripped state with

_____
[16]Assumption 1 ensures that rational politicians will never transgress in the gripped state.

probability $\overline{\pi}_0$ and conditional on this state, will transgress if and only if they are an autocratic type. They will be in the not gripped state with the complementary probability, and conditional on this state, will transgress with probability $\rho = \sigma_A + (1 - \sigma_A - \sigma_D) \cdot q$.

The value to having a new politician in the non-gripped state, $\tilde{E}$, is derived similarly except that the entrant will be in the gripped state with probability $\underline{\pi}_0$. That is, $\tilde{E}$ satisfies:

$$\tilde{E} = \underline{\pi}_0 \cdot [\sigma_A \cdot G(T) + (1 - \sigma_A) \cdot G(\tilde{T})] + (1 - \underline{\pi}_0) \cdot [\rho \cdot \tilde{G}(T) + (1 - \rho) \cdot \tilde{G}(\tilde{T})] - \alpha. \tag{15}$$

Finally, the values of $G(\tilde{T})$ and $\tilde{G}(\tilde{T})$, for analogous reasons, satisfy:

$$G(\tilde{T}) = 1 + \alpha + \beta \cdot [\delta \cdot G(\tilde{T}) + (1 - \delta) \cdot E] \tag{16}$$

$$\tilde{G}(\tilde{T}) = 1 + \alpha + \beta \cdot [\delta \cdot \tilde{G}(\tilde{T}) + (1 - \delta) \cdot \tilde{E}]. \tag{17}$$

Equations (12) to (17) form a linear system that can be solved for the six unknowns, which will of course depend on (p, q). See section A.2 for details. Subtracting E and $\tilde{E}$ from (12) and (13) respectively and re-arranging allows us to express (11) as a function of E and $\tilde{E}$ only:

$$\pi_1 \cdot [1 + \alpha - c - (1 - \beta) \cdot E] + (1 - \pi_1) \cdot \left[\frac{1 + \alpha - c - (1 - \beta) \cdot \tilde{E}}{1 - \beta \cdot p \cdot \delta} \right]. \tag{18}$$

2.4.1 Non-Permissive Norms

We begin by exploring whether there continues to exist a stationary equilibrium with non-permissive norms after the introduction of the gripped state. Under these norms, the only time that a politician transgresses is when an autocratic type comes to power. But since this event is independent of whether the current state is gripped or not, the states that we have introduced here have no bearing under non-permissive norms. Therefore we have the following.

Proposition 2. *For any $\lambda > 0$, a stationary equilibrium with non-permissive norms exists.*

The existence of a permanently non-permissive norm equilibrium is unaffected by the addition of gripped actors to the basic model. This norm entails the belief that political transgressors will be punished in equilibrium, and this belief ensures individual citizens find it optimal to withdraw support from transgressors as part of their equilibrium strategies. The addition of gripped actors who would act the same (though in their case for purely behavioral reasons) does nothing to affect the optimality of this equilibrium behavior for the rational remainder. Rational citizens may come to be intolerant of political transgressors if they believe the populace is gripped, but it is not necessary that it continues to be gripped for them to continue to demand accountability from their leaders. This is a point to which we return in our discussion of accountability transitions later.

2.4.2 Permissive Norms

When norms are permissive we have the net benefit to supporting a transgressor is

$$\gamma(\pi_1) \equiv \pi_1 \cdot [1 + \alpha - c - (1 - \beta) \cdot E] + (1 - \pi_1) \cdot \left[\frac{1 + \alpha - c - (1 - \beta) \cdot \tilde{E}}{1 - \beta \delta} \right], \tag{19}$$

where E and $\tilde{E}$ are calculated using $p = q = 1$. Recall that the existence of a stationary equilibrium with permissive norms requires that $\gamma(\pi_1) \leq 0$ for all π_1 that arise with positive probability (for some initial belief) in equilibrium.

We begin exploring the properties of γ by exploring the consequences of relatively high persistence.

Lemma 1. *If $p = q = 1$, then E is increasing in s with $\lim_{s \to 1} E = \bar{V}^+$ and $\tilde{E}$ is decreasing in s with $\lim_{s \to 1} \tilde{E} = \bar{V}^-$.*

The simple intuition is as follows. As s goes to one, citizens perceive that the current state will persist well into the future. For instance, if citizens knew that they were in the gripped state, then they would perceive that much of the future will play out in the gripped state. Since this implies that politicians only transgress if they are autocratic types, this means that behavior is similar to that arising under permanent non-permissive norms. As a result, the value of E approaches $\bar{V}^+$ as s gets large. For analogous reasons, the the value of $\tilde{E}$ approaches $\bar{V}^-$ as s gets large.

This result then implies that

$$\lim_{s \to 1} \gamma(\pi_1) = \pi_1 \cdot \left[1 + \alpha - c - (1 - \beta) \cdot \bar{V}^+ \right] + (1 - \pi_1) \cdot \left[\frac{1 + \alpha - c - (1 - \beta) \cdot \bar{V}^-}{1 - \beta\delta} \right]. \tag{20}$$

The first bracketed term is negative and the second bracketed term is positive (by assumption 2), implying that γ is decreasing in π_1 and becomes negative for sufficiently high π_1. Thus, *if s is sufficiently large, then the existence of a stationary equilibrium with permissive norms requires that beliefs do not become too high on the equilibrium path*. To see whether this is possible we must examine how beliefs evolve as play occurs.

2.4.3 Belief Updating

Fix some prior, $\pi_0 \in (0, 1)$, and suppose that the politician chooses to transgress. This could either be due to the state being gripped and the politician being an autocratic type, or to the state being non-gripped and the politician not being a democratic type. As such, we have via Bayes' rule:

$$\pi_1(T, \pi_0) \equiv \frac{\pi_0 \cdot \sigma_A}{\pi_0 \cdot \sigma_A + (1 - \pi_0) \cdot (1 - \sigma_D)}. \tag{21}$$

A transgression involves a lowering of beliefs: $\pi_1(T, \pi_0) < \pi_0$. Following the transgression, if the politician does receive sufficient support, then the state must be non-gripped. In this case we have $\pi_2 = 0$ and $\pi_0' = \underline{\pi}_0$. On the other hand, if the politician does not receive sufficient support then the state must be gripped. In this case we have $\pi_2 = 1$ and $\pi_0' = \overline{\pi}_0$. If the next politician also transgresses, then beliefs get updated to $\pi_1 = \overline{\pi}_1 \equiv \pi_1(T, \overline{\pi}_0)$. This is the highest possible value of π_1 since it updates the highest possible prior.

Now consider how beliefs are updated following a non-transgression. In this case it could either be that the state is gripped and the politician is not an autocratic type, or that the state is non-gripped and the politician is a democratic type. As such, Bayes' rule yields:

$$\pi_1(\tilde{T}, \pi_0) \equiv \frac{\pi_0 \cdot (1 - \sigma_A)}{\pi_0 \cdot (1 - \sigma_A) + (1 - \pi_0) \cdot \sigma_D}. \tag{22}$$

13

A non-transgression involves a raising of beliefs: $\pi_1(\tilde{T}, \pi_0) > \pi_0$. As no transgression occurs, no further information is revealed by the observation that the politician is supported. Thus, in this case we have $\pi_2 = \pi_1(\tilde{T}, \pi_0)$ and

$$\pi_0' = f(\pi_0) \equiv s \cdot \pi_1(\tilde{T}, \pi_0) + (1 - s) \cdot \lambda. \tag{23}$$

For an arbitrary initial belief, π_0, suppose that the following n politicians in office do not transgress. Then, the prior for the $(n+1)^{\text{th}}$ politician is $\hat{\pi}_0(n, \pi_0) \equiv f^n(\pi_0)$, where $f^0(\pi_0) = \pi_0$ and $f^n(\pi_0) = f\left(f^{n-1}(\pi_0)\right)$. If this politician transgresses, then the posterior becomes $\hat{\pi}_1(n, \pi_0) \equiv \pi_1(T, \hat{\pi}_0(n, \pi_0))$.

Since f is a strictly increasing and concave function with $f(0) = \underline{\pi}_0 \geq 0$ and $f(1) = \overline{\pi}_0 \in (0, 1]$, we have that f has a unique positive fixed point, $\pi_0^* \in (0, 1]$, and furthermore we have that $\lim_{n \to \infty} \hat{\pi}_0(n, \pi_0) = \pi_0^*$ and that $\lim_{n \to \infty} \hat{\pi}_1(n, \pi_0) = \pi_1(T, \pi_0^*) \equiv \pi_1^*$. The functions $\pi_1(T, \pi_0)$, $\pi_1(\tilde{T}, \pi_0)$, and $f(\pi_0)$ are illustrated in Figure 1a, and the points π_0^*, π_1^*, and $\overline{\pi}_1$ are illustrated in Figure 1b.
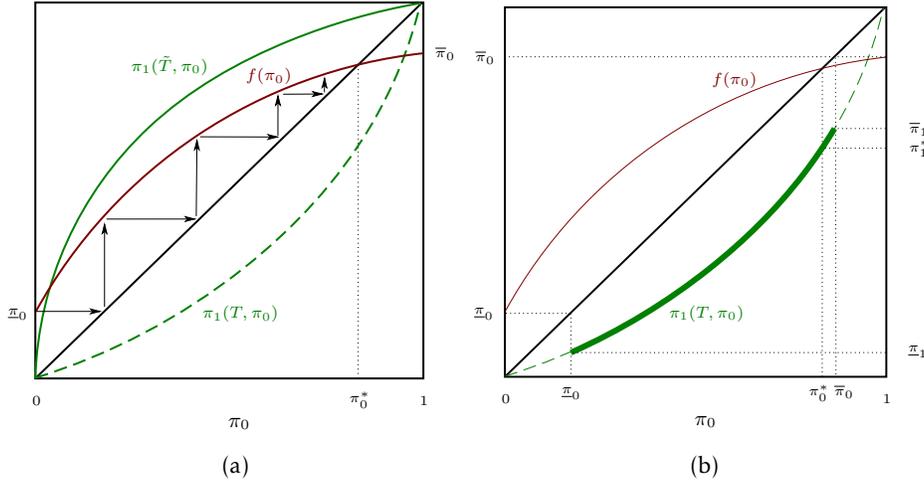
Figure 1: Belief Updating

To summarize, equilibrium beliefs rise when successive politicians do not transgress, and when there is a transgression in the gripped state. Beliefs in the former case can rise up to the point π_1^* in the limit, whereas in the latter case beliefs jump to $\overline{\pi}_1$. The following result indicates that these 'upper limits' on equilibrium beliefs become large when the degree of persistence becomes large.

Lemma 2. $\lim_{s \to 1} \pi_0^* = \lim_{s \to 1} \pi_1^* = \lim_{s \to 1} \overline{\pi}_1 = 1$.

This result, when combined with the previous, indicates that

$$\lim_{s \to 1} \gamma(\overline{\pi}_1) = \lim_{s \to 1} \gamma(\pi_1^*) = 1 + \alpha - c - (1 - \beta) \cdot \bar{V}^+ < 0, \tag{24}$$

where the inequality is implied by assumption 2. This fact delivers the following proposition.

Proposition 3. *For any $\lambda > 0$, a stationary equilibrium with permissive norms does not exist if s is sufficiently large.*

14

Note that this proposition holds even if the 'gripped' state is extremely unlikely, i.e., for $\lambda \to 0$. As λ falls, the degree of persistence must rise (but does not go to one: see the numerical results in section 2.6). The dependence on persistence stems from the benefit to removing a transgressing incumbent extending into the future, and hence rising, the greater is persistence – as discussed after Lemma 1.

A further implication of this feature of the model is that it is possible to solve for the length of the sequence of non-transgressing leaders beyond which citizens holding the most pessimistic beliefs about the possibility of permissive norms changing will necessarily no longer act permissively.

This upper bound, denoted $\bar{N}$, is determined as follows. Define $\tilde{\gamma}(n, \pi_0)$ as the net payoff to supporting a transgressor when beliefs, π_1, are those arising following n consecutive non-transgressors given an initial prior of π_0: i.e. $\tilde{\gamma}(n, \pi_0) \equiv \gamma(\hat{\pi}_1(n, \pi_0))$. Then $\bar{N}$ is defined as

$$\bar{N} \equiv \min\{n \in \mathbb{N} \mid \tilde{\gamma}(n, \underline{\pi}_0) < 0\}. \tag{25}$$

This is well-defined when s is sufficiently large.[17]

To summarize, the possibility of gripped citizens still allows an equilibrium with non-permissive norms to exist but, sufficient persistence in states does not permit an equilibrium with *permanent* permissive norms. The reason is that citizens prefer to not support transgressors when beliefs about being in the gripped state become sufficiently high, and beliefs become sufficiently high when (i) a transgression occurs in the gripped state, or (ii) a sufficiently long sequence of consecutive non-transgressors, length $\bar{N}$, is observed.

2.5 The Emergence of Accountability: Transition Equilibria

So far we have seen that permanent non-permissive norms *may* still arise with the introduction of gripped agents, and that permanent permissive norms, (the most pessimistic beliefs about the possibility of transitions) *necessarily can not* be consistent with sufficient persistence. The non-existence of permanent permissive norms naturally leads to the question of whether some form of impermanent permissive norms can exist, and whether the model can tell us anything general about the factors that precipitate improvements from permissive to non-permissive norms where rational politicians self-interestedly act accountably to their citizens.

Because we are analyzing a dynamic infinite horizon game with an arbitrarily long history, any partition of which can be conditioned on by players' strategies, it is well known that little can be said *in general* about equilibrium outcomes. To make some progress, a typical strategy in such situations is to fully characterize a class of equilibria satisfying some reasonable restrictions. One such restriction that we have explored limits the relationship between the history of past citizen permissiveness and that expected by citizens in future. Using a monotonicity restriction between past and future

---

[17]Specifically, when s is large enough that (i) γ is decreasing in π_1, and (ii) $\gamma(\pi_1^*) < 0$. In this case, $\tilde{\gamma}(n, \pi_0)$ is decreasing in n (since $\hat{\pi}_1(n, \pi_0)$ is increasing in n), and $\tilde{\gamma}(n, \pi_0) \le \tilde{\gamma}(\infty, \pi_0) = \gamma(\pi_1^*)$. We numerically calculate the critical persistence levels in section 2.6.

permissiveness, which we call *permissive monotonicity* we have been able to provide a full characterization of equilibrium outcomes.[18] We relegate this analysis to appendix B.2 while noting here that all of the results we develop in the main text apply also in that class of equilibria. The body of the paper instead proceeds in two complementary directions.

We first characterize a set of transition equilibria in which agents understand the possibility of norms changing from permissive to non-permissive along the equilibrium path. Here, we particularly focus on the process of belief updating that precipitates these transitions. We will see that beliefs about the population being 'gripped' play a key role in shifting norms, and that these beliefs are directly affected by the actions of political leaders. As will be seen, these correspond to both 'bottom-up' and 'top-down' transitions, as we have referred to them in the introduction. The second direction we pursue attempts to quantify the difference in outcomes between these transition equilibria in which citizens carry the belief that norm changes are possible, but not guaranteed, and the outcomes arising from beliefs that are maximally pessimistic about the possibility of a transition – i.e., that it will never happen. As we have already seen from Proposition 3, under these most pessimistic of beliefs about transitions, after a long enough sequence of non-transgressing leaders, all citizens will rationally not support a further transgressor. We numerically compare the sequence length obtained from this most pessimistic scenario to that obtained from the transition equilibria that we fully characterize. It will be seen that, with the exception of a very small range of the parameter space, these sequence lengths are very close.

2.5.1   N-Transition Equilibria

Motivated by the above analysis, we search for an equilibrium in which norms are initially permissive but transition to non-permissive when beliefs about being in the gripped state are sufficiently high. Such transitions can potentially occur in a 'bottom-up' way, i.e., when a transgression occurs in the gripped state, or a 'top-down' way, i.e., when citizens observe N consecutive non-transgressors irrespective of state. Once norms are non-permissive, they remain that way permanently. We call such equilibria *N-transition equilibria*.

In the initial permissive phase, suppose that a politician transgresses. If the state is not gripped, the prior belief that the next politician will operate in the gripped state is $\underline{\pi}_0$. If this politician does not transgress, then the prior belief that the next politician will operate in the gripped state is $f(\underline{\pi}_0)$. If this politician too does not transgress, then the prior that the next politician will operate in the gripped state is $f(f(\underline{\pi}_0)) = f^2(\underline{\pi}_0)$, and so on. In general, if a politician enters after n consecutive non-transgressors then citizens hold prior beliefs of $\pi_0(n) \equiv f^n(\underline{\pi}_0)$. If this politician then transgresses, the updated belief about being in the gripped state is $\pi_1(n) \equiv \pi_1(T, \pi_0(n))$.

As previously, the value associated with supporting this transgressing politician - this

---

[18] Specifically the restriction is that a history which has been strictly more permissive than another one cannot lead to expectations of a strictly less permissive future than that stemming from the strictly less permissive history.

time expressed as a function of the number of previous consecutive non-transgressors, n - is

$$V(n) \equiv \pi_1(n) \cdot G(T) + (1 - \pi_1(n)) \cdot \tilde{G}(T), \tag{26}$$

and the value of not supporting this transgressor is

$$\tilde{V}(n) \equiv \pi_1(n) \cdot E_0 + (1 - \pi_1(n)) \cdot \tilde{E}_0, \tag{27}$$

where $G(T)$ and $\tilde{G}(T)$ are the expected value associated with a transgressing incumbent politician that operates in states G and $\tilde{G}$ respectively, and E_n and $\tilde{E}_n$ are the expected value associated with an entrant politician that enters office after n consecutive non-transgressors, when the current politician operates in states G and $\tilde{G}$ respectively. Since a transgression in the gripped state triggers a transition to permanent non-permissive norms, we have that $E_0 = \overline{V}^+$. As for the others, for $n \in \{0, 1, 2, ..., N - 1\}$, these are related according to:

$$G_n(\tilde{T}) = 1 + \alpha + \beta \cdot [\delta \cdot G_n(\tilde{T}) + (1 - \delta) \cdot E_{n+1}] \tag{28}$$

$$G(T) = 1 + \alpha - c + \beta \cdot E_0 \tag{29}$$

$$\tilde{G}_n(\tilde{T}) = 1 + \alpha + \beta \cdot [\delta \cdot \tilde{G}_n(\tilde{T}) + (1 - \delta) \cdot \tilde{E}_{n+1}] \tag{30}$$

$$\tilde{G}(T) = 1 + \alpha - c + \beta \cdot [\delta \cdot \tilde{G}(T) + (1 - \delta) \cdot \tilde{E}_0], \tag{31}$$

where $G_n(\tilde{T})$ and $\tilde{G}_n(\tilde{T})$ are the expected value associated with a non-transgressing incumbent politician that operates in the gripped and non-gripped states respectively, having entered office after n consecutive non-transgressors, and

$$
\begin{aligned}
E_n = \overline{\pi}_0 \cdot \sigma_D \cdot G(T) + (1 - \overline{\pi}_0) \cdot (1 - \sigma_H) \cdot \tilde{G}(T) \\
+ \overline{\pi}_0 \cdot (1 - \sigma_D) \cdot G_n(\tilde{T}) + (1 - \overline{\pi}_0) \cdot \sigma_H \cdot \tilde{G}_n(\tilde{T}) - \alpha
\end{aligned}
\tag{32}
$$

$$
\begin{aligned}
\tilde{E}_n = \underline{\pi}_0 \cdot \sigma_D \cdot G(T) + (1 - \underline{\pi}_0) \cdot (1 - \sigma_H) \cdot \tilde{G}(T) \\
+ \underline{\pi}_0 \cdot (1 - \sigma_D) \cdot G_n(\tilde{T}) + (1 - \underline{\pi}_0) \cdot \sigma_H \cdot \tilde{G}_n(\tilde{T}), - \alpha
\end{aligned}
\tag{33}
$$

except that $E_0 = \bar{V}^+$ as previously mentioned. Also, since the transition occurs once N non-transgressors are observed, we have $E_N = \tilde{E}_N = \bar{V}^+$. Once the transition occurs, all value functions correspond to those under permanent non-permissive norms as described above.

Let $g(n)$ be the net value associated with supporting a transgressor; $g(n) \equiv V(n) - \tilde{V}(n)$. That is,

$$g(n) = \pi(n) \cdot [G(T) - E_0] + (1 - \pi(n)) \cdot [\tilde{G}(T) - \tilde{E}_0] \tag{34}$$

$$= \pi(n) \cdot [(1 + \alpha - c) - (1 - \beta) \cdot E_0] + (1 - \pi(n)) \cdot \left[\frac{(1 + \alpha - c) - (1 - \beta) \cdot \tilde{E}_0}{1 - \beta \delta} \right], \tag{35}$$

where the last equality is derived by using (29) and (31) to solve for $G(T)$ and $\tilde{G}(T)$, then subtracting E_0 and $\tilde{E}_0$ respectively. We see that $g(n)$ can be expressed as a function of E_0 and $\tilde{E}_0$. We already know that $E_0 = \bar{V}^+$, and the procedure for calculating $\tilde{E}_0$ for an arbitrary N is provided in section A.3 in the appendix.

The proposed behavior is an equilibrium if and only if $g(n) \geq 0$ for all $n \in \{0, 1, ..., N - 1\}$. Checking this condition is made considerably simpler by the following Lemma.

Lemma 3. $g(n) \geq 0$ for all $n \in \{0, 1, ..., N-1\}$ if and only if $g(N-1) \geq 0$.

In other words, we need only check the sign of $g(N-1)$: an N-transition equilibrium exists if and only if $g(N-1) \geq 0$. Notice that $g(N-1)$ need not be monotonic as there are counteracting effects. Higher N means that beliefs regarding the possibility of being in the gripped state reach a higher level, thereby acting to reduce the net benefit to supporting a transgressor. However, higher N also means that it is more difficult for a transition to occur, thus making it more beneficial to support a transgressor.

For what values of N does an N-transition equilibrium exist? First, one can show that N is bounded by $\bar{N}$ (where $\bar{N}$ is defined in (25)).[19] That is, if we let N^* denote the highest N for which a N-transition equilibrium exists, i.e.

$$N^* \equiv \max\{N \in \mathbb{N} \mid g(N-1) \geq 0\}, \tag{36}$$

then we can be sure that $N^* \leq \bar{N}$. For N^* to be well-defined we need to be sure that an N-transition equilibrium exists for *some* N. The following provides a sufficient condition for this.

Proposition 4. *A* 1*-transition equilibrium exists for sufficiently large s if:*

$$\sigma_D < \frac{\alpha(1-\beta)(1-\beta\delta\sigma_A)}{c(1-\beta\sigma_A) - \alpha\beta(1-\delta)} \tag{37}$$

Intuitively, a 1-transition equilibrium will fail to exist if the proportion of democratic types is too great because the probability of drawing a type that will induce the transitions provides insufficient incentive to support the first transgressor.

Together, we have the following implication.

Corollary 1. *If s is sufficiently large that a stationary equilibrium with permissive norms does not exist and (37) holds, then there is an N^* where $1 \leq N^* \leq \bar{N}$ (given by (36)), such that an N-transition equilibrium exists for $N = N^*$ but not for any $N > N^*$.*

In keeping with our focus on norms, it seems reasonable to focus on the N^*-transition equilibrium as this equilibrium displays the most persistent permissive behaviour. That is, we know that if s is large enough then permissive norms must give way to non-permissive norms after observing a sufficiently long sequence of non-transgressors (Proposition 3). In general, how long norms subsequently remain non-permissive cannot be specified without additional structure, but permanent non-permissiveness can never be ruled out (Proposition 2). If beliefs about the gripped state fall low enough, there always exists the possibility of transitioning from non-permissive norms back to permissive ones, in the spirit of a sun-spot transition. The impetus for that sort of change does not derive from factors internal to the model, so the model tells us little about it.[20]

In contrast, the N^* transition equilibrium is a leadership driven top-down transition that occurs along the equilibrium path of play. Along this sequence, beliefs about being in the gripped state are continually rising due to the continued non-transgression of

---

[19]This follows from the observation that N-transition equilibria are permissive monotone, and proposition 5 in the appendix.

[20]This does not make such changes any less compelling as explanations for observed outcomes – as for example Weingast (1997) illustrates with his dissection of the English Glorious Revolution. It instead only means our model has little to say about them.

leaders, until a transition occurs. This is the main norm changing effect of good leaders. In section 3.1 we will show that other features of this sequence correspond well, and can be used to understand, observed accountability transitions described by political scientists. In the next sub-section we turn to exploring the quantitative difference between these rationally anticipated N^* transitions, and the upper bound on N inducing a transition when agents are maximally pessimistic (i.e., they believe a transition cannot occur): $\bar{N}$.

2.6   Numerical Solutions

There are three facets of the model investigated in the numerical simulations that we conduct in this section. We first investigate the critical persistence level above which perpetual permissiveness cannot persist, and hence under which $\bar{N}$ is defined. Under these persistence levels, we then quantify a range of bounds for the sequence lengths N^* and $\bar{N}$ for a baseline range of parameter values. We are particularly interested in whether the critical Ns that are generated by the model are small enough under reasonable parameters for these sorts of top-down transitions to ever be feasible in reality. Clearly Ns in the hundreds or even dozens are going to render the model mostly irrelevant in real world applications.

The final aspect we consider here is the distance between N^* and $\bar{N}$. Recall that these are two bounds on N computed under the two cases we characterized: (i) assuming that agents rationally anticipate the possibility of top-down institutional change along the equilibrium path, but that permissive norms are persistent, so that N^* is the longest length consistent with a transition, and (ii) assuming that agents are as pessimistic about such change as possible ($\bar{N}$). The point of this exercise is to provide some estimate of the imprecision arising from the fact that different levels of citizen pessimism can generate different bounds on the sequence length.

2.6.1   Parameter values

Citizens receive direct benefits from political leaders each period normalized to 1. Two parameter values then determine per period citizen utility costs from transgression, c, and utility benefits from incumbency, α. In selecting α, it makes sense to consult the political economy literature that has attempted to quantify the political incumbency advantage. Lee *et al.* (2004) and Lee (2001) found, by exploiting regression discontinuity methods for closely contested seats in the US congress, that though incumbency generated small values in terms of seat shares, these had large effects on probabilities of reelection. It is not straightforward to map these numbers back to citizen utility estimates – as is required for calibrating our model – fortunately, the absolute value of α does not matter greatly, instead as will be seen it is the ratio α/c that is critical. So we proceed by only showing results for a baseline value of $\alpha = 0.1$ (implying citizens receive a 10 per cent premium from the re-election of an incumbent, ceteris parabus) and vary the range of c widely to cover what we think are the extreme cases. At the lower end, these costs are posited to be no greater than the benefits of incumbency ($\alpha = c$), and we consider a range extending up to $c = 5\alpha$. We vary the probability of the gripped state, λ,

from a low of 0.01 to a high of 0.2 again attempting to be comprehensive in covering a reasonable range. The persistence parameter can then be run for the full range (from 0 to 1). Since it reaches a critical value above which $\bar{N}$ is defined and below which permissive norms can be permanent, it plays a key role in determining the relative magnitudes of N^* and $\bar{N}$.

2.6.2 Critical Persistence

Table 1 computes the critical persistence level - i.e. the level of s such that a stationary permissive equilibrium fails to exist for all higher s. When the stationary permissive equilibrium fails to exist, it is possible to compute $\bar{N}$, the sequence length of non-transgressing leaders such that even the most pessimistic citizens will not support a subsequent transgressing politician. With sufficient persistence, high enough beliefs about the gripped state today make citizens willing to withdraw support from transgressing leaders even if they believe all other rational citizens would never do so.

The table reports values for the full range of α/c pairs.[21] For some parameter configurations assumption 2 is violated: such cells are marked "N/A". For some parameter configurations, a stationary permissive equilibrium fails to exist for *any* value of s: these cells are marked with "–".

Table 1 illustrates that the critical value s^* is high when c is low relative to α. This happens because citizens have a low relative benefit to deposing incumbents in these cases, and high persistence is required so that a gripped state today will likely lead to one in future, and hence a non transgressing leader. Though the critical persistence values are all quite high in the table, they become low when λ is close to the boundary at which a stationary permissive equilibrium exists. For example, using the parameters in the last set of results ($\beta = \delta = 0.8$, $\sigma_A = \sigma_D = 0.1$, and $c = 3\alpha$) yields $s^* = 0.2943$ when $\lambda = 0.0276$. But the parameter ranges under which s^* is low are small, suggesting that, typically a high degree of persistence is required to rule out perpetual non-permissiveness. [22]

2.6.3 Values of $\bar{N}$ and N^*

For s above the critical s^* values we can compute $\bar{N}$, allowing comparison with N^*. Table 2 reports $\bar{N}$ and N^* for values of s above s^*, under five different sets of parameters (specifications I, II, III, IV, V). Each of the parameter sets are chosen from among those reported in Table 1. The parameters used are as follows:[23]

- I: $\sigma_A = \sigma_D = 0.1$, $\beta = \delta = 0.7$, $c = 3\alpha$, $\lambda = 0.05$, $\alpha = 0.1$

- II: $\sigma_A = \sigma_D = 0.1$, $\beta = \delta = 0.7$, $c = 3\alpha$, $\lambda = 0.1$, $\alpha = 0.1$

- III: $\sigma_A = \sigma_D = 0.1$, $\beta = \delta = 0.8$, $c = 3\alpha$, $\lambda = 0.01$, $\alpha = 0.1$

- IV: $\sigma_A = \sigma_D = 0.05$, $\beta = \delta = 0.7$, $c = 5\alpha$, $\lambda = 0.05$, $\alpha = 0.1$

- V: $\sigma_A = \sigma_D = 0.1$, $\beta = \delta = 0.8$, $c = \alpha$, $\lambda = 0.1$, $\alpha = 0.1$

---

[21] The value of α is immaterial in these calculations since the critical value depends only on the ratio α/c.
[22] But note that the critical persistence level still remains less than one as λ goes to zero.
[23] We do not report these for varying α as variation in α only matters relative to c.

	λ		$c = \alpha$	$c = 3\alpha$	$c = 5\alpha$
$\beta = \delta = 0.7$	0.01		0.9773	0.8810	0.7745
&	0.05		0.9763	0.8698	0.7110
$\sigma_A = \sigma_D = 0.05$	0.10		0.9749	0.8513	–
	0.20		0.9714	0.7547	–
$\beta = \delta = 0.8$	0.01		0.9684	0.8553	0.6780
&	0.05		0.9669	0.8320	–
$\sigma_A = \sigma_D = 0.05$	0.10		0.9647	0.7696	–
	0.20		0.9594	–	–
$\beta = \delta = 0.7$	0.01		0.9583	0.7495	–
&	0.05		0.9564	0.7056	–
$\sigma_A = \sigma_D = 0.1$	0.10		0.9536	0.5590	–
	0.20		0.9469	–	–
$\beta = \delta = 0.8$	0.01		0.9377	0.5778	N/A
&	0.05		0.9343	–	N/A
$\sigma_A = \sigma_D = 0.1$	0.10		0.9294	–	N/A
	0.20		0.9169	–	N/A

Table 1: The Critical Persistence Level, s^*

The most important thing to note in this table is that the absolute values of $\bar{N}$ and N^* suggest top-down transitions occur for between 2 and up to 5 non-transgressing leaders across the range of baseline cases. This suggests that citizen beliefs regarding the gripped state rise quickly enough with non-transgressors to make top-down transitions feasible for small sequences of such leaders.

Though relatively small values for the Ns arise in our baseline, it will still be discomforting if small departures from the baseline lead these Ns to blow up. To check for this, we proceed in the other direction. We attempt to find ranges of parameters required to generate high values of the critical Ns. Intuitively, we need beliefs of being in the 'gripped' state to start very low and take a long time to update. The former is ensured by setting λ low, and the latter by setting σ_D, the probability of leaders being inherently good types, high. In this case, citizens generally interpret non-transgressing politicians as simply good types, and thus require a long sequence to believe the gripped state has occurred.

For reasonable values of β and δ, such high values of N only arise for very low values of c/α. From assumption 2, we need $c/\alpha \leq \frac{1-\beta\delta}{\sigma_D}$. For instance, the value of c/α must be less than one whenever $\beta\delta > 1 - \sigma_D$, and we've already argued that σ_D must be high to make the updating slow. In other words, the costs of politicians being corrupt need to be

	Persistence, s									
		0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95	$s^* + 0.01$
s^*	0.7056									
$\bar{N}$		–	–	–	4	3	3	3	3	5
N^*		–	–	–	3	3	3	3	3	3
s^*	0.5590									
$\bar{N}$		3	3	3	3	3	3	3	3	4
N^*		2	2	2	2	3	3	3	3	2
s^*	0.5778									
$\bar{N}$		5	4	4	4	4	4	4	4	5
N^*		3	3	3	3	3	3	3	4	3
s^*	0.7110									
$\bar{N}$		–	–	–	3	3	3	3	3	4
N^*		–	–	–	3	3	3	3	3	3
s^*	0.9294									
$\bar{N}$		–	–	–	–	–	–	–	5	5
N^*		–	–	–	–	–	–	–	4	5

Table 2: Values of $\bar{N}$ and N^* as function of persistence

outweighed by the incumbency benefits for long Ns to arise. This means that, in evaluating single period benefits of a leader, citizens get more utility from the continuation of a corrupt incumbent over a known non-corrupt challenger, which seems unreasonable. However, under these conditions, it is indeed possible to generate large values for $\bar{N}$ and N^*. For example, if we take $\sigma_D = 0.9$ and $\sigma_A = 0.09$, as well as $\beta = \delta = 0.8$, then we obtain a range that the critical ratio has to fall within to satisfy assumption 2. By taking c/α to be the mid-point in this range (0.3978) and taking $\lambda = 0.001$, then s^* is 0.99719. Taking persistence to be just above the critical value, at $s = 0.9972$, leads to extremely high sequence lengths: $\bar{N} = 2037$ and $N^* = 1889$.[24]

But this exercise points out a second difficulty that generally arises when generating large values for $\bar{N}$ and N^*. In addition to the ratio c/α becoming small, the range of admissible values in which it must lie also becomes extremely small. In this example it is only possible to generate such large values for the critical Ns when $c/\alpha \in (0.3956, 0.4)$.

---

[24]Varying λ is not as important for obtaining high N values as is raising σ_D: if we take $c/\alpha = 0.399$ as before but this time set $\lambda = 0.05$, then $s^* = 0.9934$. If we set $s = 0.994$ then we get $\bar{N} = 751$ and $N^* = 738$. Both are still high and close to one another.

This is due to the second restriction in Assumption 2, namely that $c/\alpha \geq \frac{1-\beta\delta}{1-\sigma_A}$. This implies that for any values of β and δ the ratio of the upper bound on c/α (i.e. $\frac{1-\beta\delta}{\sigma_D}$) to the lower bound (i.e. $\frac{1-\beta\delta}{1-\sigma_A}$) is simply $\frac{\sigma_D}{1-\sigma_A}$. This becomes very close to unity as σ_D becomes large (noting that σ_A is necessarily small when σ_D is large since they must sum to no more than one). Specifically, the ratio is not less than σ_D and goes to one as σ_A goes to $1 - \sigma_D$.

We take these numerical simulations to imply a fairly strong conclusion: Reasonable parameters suggest relatively small sequences of good leaders (low critical Ns) are sufficient to generate a breakdown in permissive norms. Though it is possible to find parameters that make these critical Ns large, this leads to problems both in realism (incumbency needs to generate large benefits relative to the costs of transgression) and also implies extremely slim ranges of parameters under which transition equilibria can exist.

A final point to note is that the distance between $\bar{N}$ and N^* is small. This is evident in the baseline examples reported, and is also 'generally' true except for tiny ranges of the parameter space. To see this, the last column of table 2 reports a value of s slightly above s^* (specifically, $s^* + 0.01$) for each of the parameter configurations. Since $\bar{N}$ goes to infinity as s falls to s^*, values of s near s^* will produce the greatest difference in $\bar{N}$ and N^*. However, as the final column indicates, the difference is small at even 0.01 above s^*. Moreover it shrinks further as s increases. This suggests that the bounds on the sequence length of non-transgressing leaders required to generate some non-permissiveness do not differ greatly with underlying citizen pessimism.

3   Discussion

3.1   Connection to Informal Notions of Political Change

3.1.1   Democratic Capital

There is a widely held view that, like physical or human capital, a country has a stock of Democratic Capital, that can take time to build. A recent elaboration is by Persson and Tabellini:

> Consolidation of democracy requires that citizens learn to cherish and respect democracy as a method of government. A common perception of democracy as a valuable form of government will not pop up overnight, or in a vacuum. Rather, a gradual appreciation of democracy can be envisaged as an accumulation of a stock of civic and social assets that takes place through a country's learning from its own historical experience or from its neighboring countries. We refer to this consolidation process as the accumulation of "democratic capital". Persson and Tabellini (2006)

The building of such democratic capital comes without change in the formal structures of the democracy – instead it is attitudinal – "citizens learn to cherish and respect democracy". One interpretation of this is that citizens preferences change. Another interpretation is that beliefs are altered in such a way that make citizens move from not valuing their country's nascent democratic institutions to doing so. The question is

why should this take time, and what is the process by which experience with democracy changes such beliefs? The answer we provide is illustrated by the transition from permissive to non-permissive norms in an N^* transition equilibrium. Rational citizens will oppose a transgressing leader after enough inherently democratic leaders have preceded him. Such opposition does not derive from the state being gripped, something citizens can never observe, but because a long enough sequence of good (non-transgressing) leaders, leads them to think there is a high enough chance of the state being gripped that no rational leader would ever transgress. In our analysis, there is no substitute for this experience by citizens and it leads to their respect and cherishing of the institution. By opposing a transgressor after such a sequence, they expect to be opposing an inherently bad leader, because the institution now works in ensuring rational leaders will not transgress. Enough democratic capital has been built.

3.1.2   Huntington's Two-Turnover Test

Our model also explains why Samuel Huntington's (1991) empirically motivated "two turn-over test" may hold. According to him, democracy becomes consolidated when . . .

> "the party or group that takes power in the initial election at the time of transition loses a subsequent election and turns over power to those election winners, and if those election winners then peacefully turn over power to the winners of a later election. "

In terms of the model, his test can be interpreted as a statement that $N^* = 2$, under the assumption that running a fair election and vacating after a loss corresponds with a non-transgressing leader. In the model, following an $N^* = 2$ sequence, subsequent self-interested leaders must respect electoral outcomes. Not doing so will lead to citizens withdrawing support. Democracy then becomes consolidated. Moreover a failed democratic consolidation arises for any sequence less than N^*. Citizens will act permissively to a subsequent transgression, for $N < N^*$, so self-interested leaders will continue to transgress.[25]

3.1.3   Inherent V. Materialist Democratic Motivations

Do citizens value democracy inherently or because it furnishes material well-being? As Fearon (2011) notes, the idea that two different sorts of motivations may underpin opposition to autocrats is not new:

> . . . the usual recourse is to suggest that for democracy to be self-enforcing, the public, or some significant part of it, must be motivated to protest or even rebel if democracy is threatened. . . . People might have this motivation either because they have internalized democratic norms or culture - a commitment

---

[25]Indirect evidence in support of Huntington's test has been obtained from African opinion poll data in the 1990's, by Moehler and Lindberg (2009). The supporters of winning and losing parties in elections initially have highly polarized views regarding the legitimacy of political institutions – with losers not seeing them as legitimate. This polarization declines after peaceful electoral turnovers.

to the "rules of the game" or because they expect that they will be materially worse off if dictatorship prevails.'

Our analysis includes these motivations embeded in two different types of citizens. The gripped are the internally motivated actors who oppose leaders seen to undermine democratic ideals (transgressors) no matter what. That is, irrespective of the consequence of their opposition, which they do not consider. They oppose transgressors on principle. The rational are those who choose their opposition based on calculation of expected values. What the model adds to the mere observation of two differing sources of motivation is the demonstration that the two are linked, and that the link between them can be activated by leaders. The N^* transitions show how *great* leaders are able to make rational citizens believe that a significant part of the public are gripped, and hence motivated to bring down leaders who transgress, no matter what. As the model illustrates, leaders do this by themselves not transgressing and thereby raising public beliefs about the ubiquity of intolerance for such leadership behavior. Once these beliefs are high, citizens indeed do believe that they will be materially worse off by keeping the current transgressor and are rationally motivated to oppose them.

This explanation makes sense of appeals to public disattisfaction uttered by revolutionizing leaders. A recent example of such a leader is Nitish Kumar, the reformist Chief Minister of what has previously been seen as the most corrupt and dysfunctional state in India, Bihar. In reference to the Hazare corruption protests in India he said:

> "Going by the response of the people in support of Sri Hazare's crusade against corruption, *it is clear that people are not going to take it anymore.*" Nitish Kumar, *Patna Daily*, April 11 2011, our emphasis.

According to our model, his actions in reducing his own regime's corruption serve to raise citizen's beliefs of how intolerant their compatriots will now be of corruption. If these beliefs reach the point where citizens are widely viewed to be strongly intolerant of corruption – as his quote indicates they are – then a transition to non-permissive norms necessarily occurs, and leaders (even the merely self-interested that may follow) will choose not to be corrupt.

3.1.4 Institutionalization

A related question raised in this literature is why the values forwarded by some, but not all, leaders end up getting embedded in the institutions that follow them. Leaders are seen to play a key role in making sure these values outlive their own tenure, but how are they able to do this?

> "New ideas do not achieve political prominence on their own; they must be championed by "carriers", individuals or groups capable of persuading others to reconsider the ways they think and act ...What is really being investigated here is the process of institutionalization – that is, how ideas become embedded in organizations, patterns of discourse, collective identities and so forth and manage to outlast the original conditions that gave rise to them." Berman (2001)

25

One answer has been to posit that these leaders act at a 'critical juncture'. A special confluence of events at which they are both free to act (i.e., that they are somehow able to choose their own path) and can constrain the leaders that will follow.

> "...critical junctures are characterized by a situation in which the structural (that is, economic, cultural, ideological, organizational) influences on political action are significantly relaxed for a relatively short period, with two main consequences: the range of plausible choices open to powerful political actors expands substantially and the consequences of their decisions for the outcome of interest are potentially much more momentous ...those moments when the freedom of political actors and impact of their decisions is heightened ..." Capoccia and Kelemen (2007), p 343.

But a problem with critical junctures explanations is there 'just so' nature. It seems to simply beg the question about when these instances occur.

> "There are always innumerable concepts, beliefs, policy models, etc., vying for attention; why do some and not others achieve prominence in the political realm at particular moments?" Berman (2001)

Alternatively put, the goal is:

> "...to develop, in other words, a framework for predicting when the "formative moments" in history will occur–for understanding when political spaces will open up and the possibilities for ideational or cultural shifts will arise." Rothstein (1992).

This is also a principle aim of a literature under the heading of "Historical Institutionalism". Which is an attempt to provide:

> "...a theoretical bridge between men who make history and the circumstances under which they are able to do so." Thelen and Steinmo (1992), p. 12

Play along the N^* transition path provides answers to all of these questions: why great leaders can embed their preferences into an institution; how and when this can consolidate the institution, and when it is not enough; when a critical juncture occurs; and why leaders acting in the right way at such times has long-lasting institutional effects. We can also point precisely to the sequence of actions leading up to a formative moment and conceptually at least say what such a moment is.

Under permissive norms, leaders materially gain from transgressing, as citizens do not oppose transgressors so that political institutions do not render leaders accountable. Following a sequence of non-transgressing leaders, of length $N^* - 1$, we have a critical juncture. At such a point, a leader is also free to transgress without restriction. A great leader is an inherently democratic type entering at this point. He has the opportunity to embed his ideals into the institution. If he does not transgress, norms transition to non-permissive. Necessarily, his successor will face self-interested incentives to also not transgress as norms now dictate non-transgressors are deposed. However, if he transgresses, his successor will be similarly free from institutional constraint, and permissive political norms persist. This leader's decision at the critical juncture is momentous.

3.1.5 Structure V. Political Agency

Our model also provides a means to reconcile the tensions between a literature (familiar to economists) emphasizing structural determinants of political choice by leaders, i.e., incentives, with a view espoused more commonly by some political scientists emphasizing the agency and values of leaders.

Fearon (1999) exemplifies the structural view:

> "Democracy might be rendered self-enforcing in a country if its potential rulers internalize a strong normative attachment to democratic 'rules of the game'. At least since Locke's Second Treatise on Government, however, liberal political theorists have not imagined that the problem could be solved by assuming (or creating) virtuous leaders. The temptations of power are too strong, and in any event, intense political competition could make almost any faction believe that it was doing the right thing by preventing the other side from taking over. Instead, the usual recourse is to suggest that for democracy to be self-enforcing, the public, or some significant part of it, must be motivated to protest or even rebel if democracy is threatened."

And the alternative is emphasized by Diamond *et al.* (1999)

> " The cases in this volume strongly suggest a reciprocal relationship between the policies culture and the policies system. Democratic culture helps to maintain and also pressure for the return of democracy, but historically, the choice of democracy by political elites clearly preceded in many of our cases the presence of democratic values among the general public. This elite choice of democracy was no doubt influenced by values, including those induced by international diffusion and demonstration effects. p.39 "

Both incentives and preferences of leaders play an important role in our analysis. In our 'top-down' N^* transitions, under the initially permissive norms, the advent of an elite (leader) whose values are inherently democratic does, in fact, precede the emergence of what may be called democratic values in the public as a whole. Under permissive norms, the public will not oppose a transgressing leader, so only leaders acting under internal values will do so, consistent with Diamond *et al.* (1999). But, under permissive norms, democracy is not self-enforcing, and any such virtuous leaders are aberrant and cannot lead to consolidation on their own. However, their actions can be part of a sequence that makes democracy self-enforcing, in the structural sense advocated by Fearon (2011) above. An N^* sequence of such virtuous leaders forces future non-virtuous (but self-interested) ones to act accountably to their citizens. This is akin to the public now exhibiting democratic values as a whole. That is, now being willing to stand against leaders seen to have acted unaccountably.

3.1.6 Great Leaders

There exist numerous examples of leaders whose values seemed to precede and, in some sense, mold those of their political constituents in their image.

"The period in which a new democratic regime is founded and begins to function provides a particularly wide scope for political leadership to shape the actual character of politics and political institutions. In several of our cases, political leaders stand out, individually or collectively for choices, initiatives, and strategies that crucially contributed to democratic development at a formative moment. ....The early development of democracy in Chile cannot be understood without appreciating the role of General Manuel Bulnes ...like George Washington he chose to leave office at the end of his term ...The role of Jose Figueres in institutionalizing democracy in Costa Rica a century later presents some interesting parallels ...he was in a position to 'do anything he wanted, including setting up a personal dictatorship.' Instead he administered honest parliamentary elections, which his party lost badly, accepting the defeat of his proposed constitution ...and then transferring power to the conservative victor ...whose party had blocked his legislative agenda." Diamond *et al.* (1999)

In addition to these three historic cases, a recent example of such a leader driven transition exhibiting features consistent with our model is provided by Matthieu Kerekou's post 1990 record in Benin. Kerekou was the Marxist/Leninist Military ruler of Benin who came to power in 1972. In 1990, reacting to a macroeconomic crisis, and after the collapse of the Soviet Union, he convened a 'National Conference' for dialogue on the country's political and economic future. The conference called for, and he implemented, a presidential election which was held in 1991. It was perceived as fair and legitimate, and moreover, Kerekou lost the runoff ceding power to a political outsider – Nicephore Soglo. At this point, Kerekou is notable for being the first African president to lose office after an election defeat. He re-contests the next election in 1996 and wins the presidential runoff against Soglo, after which Soglo cedes power back to him. After 5 more years of incumbency, he stands for re-election in 2001 and wins another term against Soglo – though this time many irregularities in the electoral process were reported. In 2006 Kerekou hits a constitutionally imposed age (and term) limit prohibiting him from standing for another term. Many reports from the time indicate his reticence to leave office, for which he will need to either suspend or change the constitution. He informally canvasses the degree of political support he would receive for such a move. At the time, it is widely seen that his decision on how to proceed will be pivotal for Benin's democracy. For example, *The Economist* magazine on May 16th 2006 editorializes:

"So far, however, his promise to give up power has to be taken at face value. Benin's democracy-loving people should make him stick to his word: it could be the former general's greatest legacy."

But why should he step down from power? According to our model we interpret his initial ceding of power after the 1991 election as the first step in an $N^* = 2$ transition from permissive to non-permissive norms. The next step was Soglo's peaceful electoral transfer of power back to him in 1996. At this point, political norms became non-permissive. These norms were not tested in the subsequent election as Kerekou won. Though there were numerous questions about the legitimacy of this election, leading Soglo to abandon

the run-off, there was no conclusive proof of a transgression. However, Kerekou's hitting of the constitutional age (and a 2-term) limit in 2006 was unequivocal. As was his desire to continue beyond it. As reported in the same *Economist* article, the reason he had to respect this limit is intimately linked to the change in political norms that *he* took the first steps in engineering. Beninese citizens' views of political legitimacy had been altered by this sequence of peaceful transitions, and the implication is that such a further transgression would not be tolerated.

> "Since then, democracy has implanted itself strongly in the minds of Benin's citizens. 'Our history is so terrible, with coups and years of problems, that now we all care about democracy very deeply,' says one of them."

Kerekou eventually chose to stand down, and played no part in the accession of his successor – Yayi Boni – a strong critic of his who won the Presidential election while Kerekou was still in office.

3.2   Addressing Specific Modeling Choices

Throughout we have restricted politicians to make a once-and-for-all decision about whether to transgress. This certainly has made the analysis cleaner, but at what cost? To address this, suppose that politicians could change their actions mid-way through their incumbency. If the state does not change within an incumbency, it seems reasonable to restrict attention to strategies of citizens that do not change within an incumbency. If norms are permissive, then there is no incentive to change behavior by ceasing transgression. If norms are non-permissive, then changing behavior by starting to transgress will reveal to citizens that the politician is a rational type and that the state is non-gripped. Still, since future politicians are using a strategy of not transgressing such a deviation will be punished by citizens. In short, nothing changes.

This argument relies on no state changes within an incumbency. How would relaxing this change matter? The behavior of autocratic and democratic types is obviously unchanged. There is clearly no impact under non-permissive norms as we have shown that states are irrelevant there. There are competing forces when norms are permissive. A sudden commencement of transgressions reveals that the state was gripped but now is non-gripped. This lowers beliefs to the lowest possible level, thereby making the 'top-down' transitions occur less often. A sudden cessation of transgressions (still under permissive norms) implies that the state has changed to gripped. This raises beliefs to the highest possible level and thereby makes the 'top-down' transitions occur more often.

Here we have bad-to-good transitions arising from the actions of politicians that were not motivated to leave an institutional legacy *per se*. If we introduced motivated politicians, then the transitions we describe would presumably only be strengthened; the rational politician following $N-1$ consecutive non-transgressors now has the potential to leave a legacy by choosing to refrain from transgressing. However, now upon observing a non-transgression, citizens put some weight on the politician being a rational-motivated type and therefore less weight on the possibility of being in the gripped state, thereby

undermining the 'top-down' mechanism stressed here. There is also the strange possibility that a motivated type would transgress in the gripped state in order to (selflessly) incite a 'bottom-up' transition.

We have assumed a binary transgression decision. If there were instead degrees of transgression, e.g. transgression intensity is a continuous variable, then one can imagine the same sort of multiplicity arising - if politicians are expected to transgress extensively, then citizens are willing to put up with a great deal of transgression before finding it optimal to lose the incumbency advantage. If democratic types transgress with a lower intensity (say zero) then their actions will always reveal them. Same for autocratic types - their high intensity of transgression will reveal them. If being gripped now means a tolerance of exactly what the democratic types do, then the top-down transition will remain. This is a strong assumption, but it could easily be relaxed by having politicians that differ with respect to a continuous 'moral cost of transgressing'. All this would greatly obscure the main insights here without adding a great deal.

4   Conclusions

We have developed a dynamic political agency model to investigate the dynamics of political norms. Specifically, we study the endogenous transition to non-permissive norms, whereby citizens punish transgressors and leaders act accountably, from permissive norms, whereby citizens optimally tolerate transgressing leaders and leaders are unaccountable. The model predicts both 'bottom-up' change, where transitions are sudden and are triggered by politician transgressions in a gripped state, and 'top-down' change, where transitions are gradual and are triggered by a consecutive sequence of non-transgressing politicians. We have argued that our model formalizes various informal notions related to political change, including democratic capital, critical junctures, and institutionalization.

Perhaps the main insight obtained from the model is an explanation for the vital role that leaders can play in engineering long-lived institutional change. Virtuous, good acting, leaders achieve change by modifying citizens' beliefs regarding the willingness of their compatriots to tolerate political transgressors. A sequence of such leaders can serve to modify citizen beliefs so much that it becomes rational for non-virtuous, but self-interested, leaders to also act well in office. When this happens, institutional incentives, manifested through norms of political accountability, ensure that only autocratic types will transgress on citizens rights, and citizens will rationally punish them.

The model is simple, but could be extended in various interesting directions. Chief among these is introducing features that give rise to deteriorating political norms - i.e. from non-permissive to permissive behaviour.[26] One possible way this could be done is to introduce windfall gains to office, and allow the existence of such gains to persist across successive politicians. Observing a transgression under non-permissive norms

---

[26]It is tempting to speculate that such 'reverse' transitions to permissive norms may emerge when sufficiently many transgressors are observed during a period of non-permissiveness. In appendix section B.3 we show that this is not the case.

would raise beliefs about being in the 'windfall' state and may eventually make citizens sufficiently certain that challenging politicians will transgress in order to secure such gains. This chain of events would make citizens permissive of transgressions, thereby leading to a deterioration of political norms. This and related extensions are left for future research.

Appendix

A Supporting Details

A.1 Stationary Equilibria without Gripped Citizens

Equations (2) to (4) can be written in matrix form, $\mathbf{Y_1 v} = \mathbf{Y_2}$, where

$$\mathbf{Y_1} = \begin{pmatrix} 1 + \alpha - c \\ 1 + \alpha \\ -\alpha \end{pmatrix}, \tag{38}$$

$$\mathbf{v} = \begin{pmatrix} V(T) \\ V(\tilde{T}) \\ \bar{V} \end{pmatrix}, \tag{39}$$

and

$$\mathbf{Y_2} = \begin{pmatrix} 1 - \beta\delta \cdot p & 0 & -\beta(1 - \delta \cdot p) \\ 0 & 1 - \beta\delta & -\beta * (1 - \delta) \\ -\rho & -(1 - \rho) & 1 \end{pmatrix}. \tag{40}$$

The value functions are then simply $\mathbf{v} = \mathbf{Y_1}^{-1}\mathbf{Y_2}$.

A.2 Stationary Equilibria with Gripped Citizens

Equations (12) to (17) can be written in matrix form, $\mathbf{Y_3 y} = \mathbf{Y_4}$, where

$$\mathbf{Y_3} = \begin{pmatrix} 1 - \beta \cdot \delta & 0 & 0 & 0 & -\beta(1 - \delta) & 0 \\ 0 & 1 & 0 & 0 & -\beta & 0 \\ 0 & 0 & 1 - \beta\delta & 0 & 0 & -\beta(1 - \delta) \\ 0 & 0 & 0 & 1 - \beta\delta \cdot p & 0 & -\beta(1 - p \cdot \delta) \\ -\overline{\pi}_0(1 - \sigma_A) & -\overline{\pi}_0\sigma_A & -(1 - \overline{\pi}_0)(1 - \rho) & -(1 - \overline{\pi}_0)\rho & 1 & 0 \\ -\underline{\pi}_0(1 - \sigma_A) & -\underline{\pi}_0\sigma_A & -(1 - \underline{\pi}_0)(1 - \rho) & -(1 - \underline{\pi}_0)\rho & 0 & 1 \end{pmatrix} \tag{41}$$

$$\mathbf{y} = \begin{pmatrix} G(\tilde{T}) \\ G(T) \\ \tilde{G}(\tilde{T}) \\ \tilde{G}(T) \\ E \\ \tilde{E} \end{pmatrix} \tag{42}$$

31

$$\mathbf{Y_4} = \begin{pmatrix} 1+\alpha \\ 1+\alpha-c \\ 1+\alpha \\ 1+\alpha-c \\ -\alpha \\ -\alpha \end{pmatrix}. \tag{43}$$

The value functions are then simply $\mathbf{y} = \mathbf{Y_3}^{-1}\mathbf{Y_4}$.

A.3 N-transition Equilibria

From (28)-(31), we have

$$G_n(\tilde{T}) = \frac{1+\alpha}{1-\beta\delta} + \frac{\beta\cdot(1-\delta)}{1-\beta\delta}\cdot E_{n+1} \tag{44}$$

$$G(T) = 1+\alpha-c+\beta\cdot E_0 \tag{45}$$

$$\tilde{G}_n(\tilde{T}) = \frac{1+\alpha}{1-\beta\delta} + \frac{\beta\cdot(1-\delta)}{1-\beta\delta}\cdot \tilde{E}_{n+1} \tag{46}$$

$$\tilde{G}(T) = \frac{1+\alpha-c}{1-\beta\delta} + \frac{\beta\cdot(1-\delta)}{1-\beta\delta}\cdot \tilde{E}_0. \tag{47}$$

In matrix form, this is

$$\begin{pmatrix} G_n(\tilde{T}) \\ G(T) \\ \tilde{G}_n(\tilde{T}) \\ \tilde{G}(T) \end{pmatrix} = \begin{pmatrix} \frac{1+\alpha}{1-\beta\delta} \\ 1+\alpha-c \\ \frac{1+\alpha}{1-\beta\delta} \\ \frac{1+\alpha-c}{1-\beta\delta} \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ \beta & 0 \\ 0 & 0 \\ 0 & \frac{\beta(1-\delta)}{1-\beta\delta} \end{pmatrix} \begin{pmatrix} E_0 \\ \tilde{E}_0 \end{pmatrix} + \begin{pmatrix} \frac{\beta(1-\delta)}{1-\beta\delta} & 0 \\ 0 & 0 \\ 0 & \frac{\beta(1-\delta)}{1-\beta\delta} \\ 0 & 0 \end{pmatrix} \begin{pmatrix} E_{n+1} \\ \tilde{E}_{n+1} \end{pmatrix}. \tag{48}$$

By making the obvious definitions, write this as

$$\mathbf{X_n} = \mathbf{L_0} + \mathbf{L_1 Q_0} + \mathbf{L_2 Q_{n+1}}. \tag{49}$$

We can express (32) and (33) in matrix form as

$$\begin{pmatrix} E_n \\ \tilde{E}_n \end{pmatrix} = \begin{pmatrix} -\alpha \\ -\alpha \end{pmatrix} + \begin{pmatrix} \overline{\pi}_0(1-\sigma_A) & \overline{\pi}_0\sigma_A & (1-\overline{\pi}_0)\sigma_D & (1-\overline{\pi}_0)(1-\sigma_D) \\ \underline{\pi}_0(1-\sigma_A) & \underline{\pi}_0\sigma_A & (1-\underline{\pi}_0)\sigma_D & (1-\underline{\pi}_0)(1-\sigma_D) \end{pmatrix} \begin{pmatrix} G_n(\tilde{T}) \\ G(T) \\ \tilde{G}_n(\tilde{T}) \\ \tilde{G}(T) \end{pmatrix}. \tag{50}$$

Again using the obvious definitions, write this as:

$$\mathbf{Q_n} = \mathbf{R_0} + \mathbf{R_1 X_n}. \tag{51}$$

Using (49) in (51) therefore gives

$$\mathbf{Q_n} = [\mathbf{R_0} + \mathbf{R_1 L_0}] + [\mathbf{R_1 L_1}]\mathbf{Q_0} + [\mathbf{R_1 L_2}]\mathbf{Q_{n+1}} \tag{52}$$

$$= \mathbf{A_0} + \mathbf{A_1 Q_0} + \mathbf{B Q_{n+1}}, \tag{53}$$

where the final equality makes use of the obvious definitions. By successive substitution, we get

$$\mathbf{Q_0} = \mathbf{Z}\cdot[\mathbf{A_0} + \mathbf{A_1 Q_0}] + \mathbf{B}^N \cdot \mathbf{Q_N}, \tag{54}$$

32

where $\mathbf{Q}_N = [\bar{V}^+ \ \bar{V}^+]^T$, and assuming that the absolute value of the eigenvalues of $\mathbf{B}$ are less than unity,

$$\mathbf{Z} \equiv \mathbf{I} + \mathbf{B} + \mathbf{B}^2 + \cdots + \mathbf{B}^{N-1} = [\mathbf{I} - \mathbf{B}]^{-1} \left[\mathbf{I} - \mathbf{B}^N\right]. \tag{55}$$

Re-arranging (54) gives

$$\mathbf{Q}_0 = [\mathbf{I} - \mathbf{Z}\mathbf{A_1}]^{-1} [\mathbf{Z} \cdot \mathbf{A_0} + \mathbf{B}^N \cdot \mathbf{Q}_N]. \tag{56}$$

The second element of $\mathbf{Q}_0$ gives us $\tilde{E}_0$ (recalling that we already know E_0 equals $\bar{V}^+$, and not the first element of $\mathbf{Q}_0$).

B Proofs and Supporting Results

B.1 Proofs

Proof of Proposition 1

Proof. Using the arguments made in the text, there is an equilibrium with non-permissive norms if and only if $(1-\beta) \cdot \bar{V}^+ \geq 1 + \alpha - c$. Using $\bar{V}^+$ from (6) and simple algebra delivers the stated condition. Similarly, permissive norms are supported iff $(1-\beta) \cdot \bar{V}^- \leq 1 + \alpha - c$. Using $\bar{V}^-$ from (7) and simple algebra delivers the stated condition. $\square$

Proof of Proposition 2

Proof. When $p = q = 0$, we have $E = \tilde{E}$ where the common value is $\bar{V}^+$. See section A.2 for how E and $\tilde{E}$ are calculated. As such, the net benefit to supporting a transgressor is $1 + \alpha - c - (1-\beta) \cdot \bar{V}^+$, which must be non-positive by Assumption 2. $\square$

Proof of Lemma 1

Proof. Section A.2 describes how to calculate E and $\tilde{E}$ as the solution to a linear system. The claimed effect of s on the solutions is easily verified, as is the fact that the solutions are continuous in s. Continuity implies that the limits of E and $\tilde{E}$ as $s \to 1$ equal the values of E and $\tilde{E}$ at $s = 1$, which are easily confirmed to be the stated values. $\square$

Proof of Lemma 2

Proof. Being a fixed point, the value of π_0^* satisfies

$$\pi_0^* = s \cdot \frac{\pi_0^* \cdot (1 - \sigma_A)}{\pi_0^* \cdot (1 - \sigma_A) + (1 - \pi_0^*) \cdot \sigma_D} + (1 - s) \cdot \lambda. \tag{57}$$

This gives a quadratic function of π_0^* with one root that lies in $(0, 1)$ and one that is negative for all $s < 1$. The fixed point lies in $[0, 1]$ and therefore is the former root. This root is increasing in s (both roots are, in fact), approaching 1 as $s \to 1$. The value of π_1^*, by definition, equals $\frac{\pi_0^* \cdot \sigma_A}{\pi_0^* \cdot \sigma_A + (1 - \pi_0^*) \cdot (1 - \sigma_D)}$ which goes to 1 since $\pi_0^* \to 1$ as $s \to 1$. Since $\overline{\pi}_1 \in [\pi_1^*, 1]$, the fact that $\pi_1^* \to 1$ implies $\overline{\pi}_1 \to 1$ also. $\square$

Proof of Proposition 3

Proof. By (24), there exists values of $s < 1$ such that $\gamma(\overline{\pi}_1) < 0$. If s is sufficiently large that $\gamma(\overline{\pi}_1) < 0$, then citizens will find it optimal to not support a transgressor that follows a period in which there is a transgression in the gripped state (necessarily by an autocratic type).

Alternatively, (24) implies that there exists values of $s < 1$ such that $\gamma(\pi_1^*) < 0$. This implies that there exists a finite $\bar{N}$ such that $\gamma(\hat{\pi}_1(n, \underline{\pi}_0)) < 0$ for all $n \geq \bar{N}$. But if s is sufficiently large that $\gamma(\hat{\pi}_1(\bar{N}, \underline{\pi}_0)) < 0$, then citizens will find it optimal to not support a transgressor that follows $\bar{N}$ consecutive non-transgressors. □

Proof of Lemma 3.

Proof. The 'only if' part is obvious. For the 'if' part, suppose that $g(N - 1) \geq 0$. Since the first bracketed term in g is negative, it must be that the second bracketed term is positive. But then g is decreasing in n since $\pi_1(n)$ is increasing in n. But then it follows that $g(n) \geq 0$ for all $n \in \{0, 1, ..., N - 1\}$. □

Proof of Proposition 4.

Proof. As $s \to 1$ we have that $\pi_1(0) \to 0$. Therefore $g(0) \geq 0$ requires that $\tilde{E}_0$ (when $N = 1$) is sufficiently small. Specifically, we must have $(1 + \alpha - c) - (1 - \beta) \cdot \tilde{E}_0 > 0$. The stated condition is that required to ensure this when $s = 1$. Details on how $\tilde{E}_0$ is calculated are provided in section A.3. □

B.2 Permissive Monotonicity

Let $a_t \in \{T, \tilde{T}\}$ be the action taken by the period t politician, and let $s_t \in [0, 1]$ be the proportion of citizens that choose to support the politician in period t. Let these period t outcomes be summarized by $h_t = \{a_t, s_t\}$ and let the history up to the beginning of period t be $h^t = \{h_{t-1}, h_{t-2}, ...\}$.

Say that history h^{t+1} is the *permissive extension* of history h^t if $h^{t+1} = \{\{T, 1\}, h^t\}$. Say that history h^{t+1} *is more permissive than* history h^t if either (i) h^{t+1} is the permissive extension of h^t, or (ii) h^t is the permissive extension of h^{t-1} and $h^{t+1} = \{\{T, s'\}, h^t\}$ and $h^t = \{\{T, s''\}, h^{t-1}\}$, where $s' \geq s''$.

Intuitively, the permissive extension of h is more permissive than h since it simply appends one extra period in which the most permissive actions are played. Similarly, suppose that we take two histories, one more permissive than the other. To the more permissive history we add a period in which a transgressor is supported with a given probability. To the less permissive history we add a period in which a transgressor is supported with a lower probability. Then, the former history remains more permissive than the latter.

Fixing some equilibrium strategies, let $\bar{V}(h^t)$ be the value of drawing a new politician at the start of date t given history h^t. Say that an equilibrium is *permissive monotone* if $\bar{V}(h^{t+1}) \leq \bar{V}(h^t)$ whenever h^{t+1} is more permissive than h^t. Simply put, this property captures the notion that a more permissive past should not be indicative of a strictly less permissive future (and therefore one with a higher value). It is straightforward to verify

that the stationary permissive, stationary non-permissive, and N-transition equilibria are all permissive monotone.

Proposition 5. *If s is sufficiently large, then in* any *permissive monotone equilibrium, permissive play must cease following $\bar{N}$ consecutive non-transgressors, where $\bar{N}$ is given by (25).*

Proof. Let s be sufficiently large that $\gamma(\pi_1^*) < 0$. Let $\tilde{\gamma}(n, \pi_0) \equiv \gamma(\hat{\pi}_1(n, \pi_0))$. Note that $\lim_{n \to \infty} \tilde{\gamma}(n, \pi_0) = \gamma(\pi_1^*) < 0$. Therefore, for any π_0 there is a finite smallest n such that $\tilde{\gamma}(n, \pi_0) < 0$. Since $\tilde{\gamma}(n, \pi_0)$ is decreasing in π_0 (since $\gamma(\pi_1)$ is decreasing in π_1 when s is sufficiently large), we have $\tilde{\gamma}(n, \pi_0) \le \tilde{\gamma}(n, \underline{\pi}_0)$. Let $\bar{N}$ be the smallest value of n such that $\tilde{\gamma}(n, \underline{\pi}_0) < 0$.

Consider a politician, i, that enters after $\bar{N}$ consecutive non-transgressors and suppose to the contrary, that there is a permissive monotone (PM) equilibrium in which citizens are permissive towards i. If i transgresses, the value to supporting him at date t if the current state is gripped is given by

$$G_{t+1}^*(T) = 1 + \alpha - c + \beta \cdot E_{t+2}^*, \tag{58}$$

where E_{t+2}^* is the expected value of drawing a new politician at the start of period $t + 2$ given that the state is gripped. If supported, we arrive at history $h^{t+1} = \{\{T, 1\}, h^t\}$. Since the state is gripped and play is permissive by supposition, in the following period a proportion $1 - \tilde{z}$ of citizens support the transgression. As a result we arrive at history $h^{t+2} = \{\{T, 1 - \tilde{z}\}, h^{t+1}\}$. Since h^{t+1} is the permissive extension of h^t and $1 - \tilde{z} < 1$, we have that h^{t+2} is more permissive than h^{t+1}. Since we are in a PM equilibrium, we must have $E_{t+2}^* \le E_{t+1}^*$. Therefore

$$G_{t+1}^*(T) - E_{t+1}^* \le 1 + \alpha - c - (1 - \beta) \cdot E_{t+1}^* \le 1 + \alpha - c - (1 - \beta) \cdot E, \tag{59}$$

where E is the value of drawing a new politician under permanent permissive norms. The final inequality follows from $E_{t+1}^* \ge E$ since the probability of a transgression occurring in any given period is weakly greater with permanent permissive norms.

Similarly, if i transgresses, the value to supporting him at date t if the current state is not gripped is given by

$$\tilde{G}_{t+1}^*(T) = 1 + \alpha - c + \beta \cdot [\delta \cdot \tilde{G}_{t+2}^*(T) + (1 - \delta) \cdot \tilde{E}_{t+2}^*], \tag{60}$$

where $\tilde{E}_{t+2}^*$ is the expected value of drawing a new politician at the beginning of period $t + 2$ given that the state is not gripped.

If supported, we arrive at history $h^{t+1} = \{\{T, 1\}, h^t\}$. Since the state is not gripped but play is permissive by supposition, in the following period all citizens support the transgression. As a result we arrive at history $h^{t+2} = \{\{T, 1\}, h^{t+1}\}$. Since h^{t+1} is the permissive extension of h^t and $1 \le 1$, we have that h^{t+2} is more permissive than h^{t+1}. Since we are in a PM equilibrium, we must have $\tilde{E}_{t+2}^* \le \tilde{E}_{t+1}^*$. Thus

$$\tilde{G}_{t+1}^*(T) - \tilde{E}_{t+1}^* = 1 + \alpha - c - (\tilde{E}_{t+1}^* - \beta \cdot \tilde{E}_{t+2}^*) + \beta \cdot \delta \cdot (\tilde{G}_{t+2}^*(T) - \tilde{E}_{t+2}^*) \tag{61}$$

$$\le 1 + \alpha - c - (1 - \beta) \cdot \tilde{E}_{t+1}^* + \beta \cdot \delta \cdot (\tilde{G}_{t+2}^*(T) - \tilde{E}_{t+2}^*). \tag{62}$$

Let $X_t \equiv \tilde{G}_t^*(T) - \tilde{E}_t^*$ and $z_t \equiv 1 + \alpha - c - (1 - \beta) \cdot \tilde{E}_t^*$, so that we have

$$X_{t+1} = z_{t+1} + \beta \cdot \delta \cdot X_{t+2} = z_{t+1} + \beta \cdot \delta \cdot z_{t+2} + (\beta \cdot \delta)^2 \cdot X_{t+3} = \cdots = \sum_{\tau=0}^{\infty} (\beta \cdot \delta)^\tau \cdot z_{t+1+\tau}. \quad (63)$$

But

$$\sum_{\tau=0}^{\infty} (\beta \cdot \delta)^\tau \cdot z_{t+1+\tau} \leq \sum_{\tau=0}^{\infty} (\beta \cdot \delta)^\tau \cdot [1 + \alpha - c - (1 - \beta) \cdot \tilde{E}] = \frac{1 + \alpha - c - (1 - \beta) \cdot \tilde{E}}{1 - \beta \cdot \delta}, \quad (64)$$

where $\tilde{E}$ is the value of drawing a new politician under permanent permissive norms conditional on being in the non-gripped state. Again the inequality follows from $\tilde{E}_{t+\tau}^* \geq \tilde{E}$ since the probability of a transgression occurring in any given period is weakly greater with permanent permissive norms. Therefore, we have

$$\tilde{G}_{t+1}^*(T) - \tilde{E}_{t+1}^* \leq \frac{1 + \alpha - c - (1 - \beta) \cdot \tilde{E}}{1 - \beta \cdot \delta}. \quad (65)$$

The net value to supporting politician i in period t is

$$\tilde{\gamma}^*(\bar{N}, \pi_0) = \pi_1(\bar{N}, \pi_0) \cdot [G_t^*(T) - E_t^*] + (1 - \pi_1(\bar{N}, \pi_0)) \cdot \left[\tilde{G}_{t+1}^*(T) - \tilde{E}_{t+1}^* \right]. \quad (66)$$

Using (59) and (65), we have

$$\tilde{\gamma}^*(\bar{N}, \pi_0) \leq \tilde{\gamma}(\bar{N}, \pi_0) \leq \tilde{\gamma}(\bar{N}, \underline{\pi}_0) < 0. \quad (67)$$

But this is a contradiction since it is evidently not optimal to support i if they transgress.

□

B.3 Reverse Transitions

We have shown how counter-norm behaviour, if sufficiently pervasive, can change political norms. Specifically, superior norms (from the citizens' perspective) arise when politicians unexpectedly refrain from transgressions. One may speculate that the reverse may also be true: inferior norms arise when politicians unexpectedly engage in transgressions. We show that this is not the case in the simple baseline setting without gripped citizens.

Proposition 6. *Let $\lambda = 0$ so that we are in the base case without gripped citizens. Apart from the special case in which $\sigma_D = (\alpha/c) \cdot (1 - \beta \cdot \delta)$, there is never an equilibrium in which play starts out non-permissive and transitions to permissive once $N \geq 1$ transgressors are observed.*

Proof. Suppose that politician i enters after $N - 1$ transgressors have already been observed. If i transgresses, then since norms are non-permissive toward i and since the transition has been triggered, the payoff to supporting i when they transgress is

$$V_i^* = 1 + \alpha - c + \beta \cdot \bar{V}^-. \quad (68)$$

The value to not supporting i when they transgress is $\bar{V}^-$, which is the same for all future politicians. The net benefit to supporting i when they transgress is therefore

$$1 + \alpha - c - (1 - \beta) \cdot \bar{V}^- \leq 0, \quad (69)$$

where the inequality must hold since equilibrium play calls for citizens to be non-permissive i. Following the transition (i.e. in the permissive phase) we have that the net value to supporting a transgressor is

$$V - \bar{V}^- = \frac{1 + \alpha - c - (1 - \beta) \cdot \bar{V}^-}{1 - \beta\delta} \geq 0, \tag{70}$$

where the inequality must hold since equilibrium play calls for citizens to be permissive in this phase. Therefore both (69) and (70) can only hold in the special case in which $\bar{V}^- = (1 - \alpha - c)/(1 - \beta)$, which is equivalent to $\sigma_D = (\alpha/c) \cdot (1 - \beta \cdot \delta)$. In all other cases one of the two conditions will be violated. $\qquad\square$

References

ACEMOGLU, D. and JACKSON, M. O. (2011). *History, Expectations, and Leadership in the Evolution of Social Norms*. NBER Working Papers 17066, National Bureau of Economic Research, Inc.

— and ROBINSON, J. (2006). *Economic origins of dictatorship and democracy*. Cambridge University Press.

— and ROBINSON, J. A. (2000). Why did the west extend the franchise? democracy, inequality, and growth in historical perspective. *The Quarterly Journal of Economics*, **115** (4), 1167–1199.

— and — (2001). A theory of political transitions. *American Economic Review*, **91** (4), 938–963.

ALESINA, A. and ROSENTHAL, H. (1995). *Partisan Politics, Divided Government, and the Economy*. Cambridge, UK: Cambridge University Press.

ALMOND, G. and VERBA, S. (1963). *The civic culture: political attitudes and democracy in five nations*. Princeton University Press.

ASHWORTH, S. (2005). Reputational dynamics and political careers. *Journal of Law, Economics and Organization*, **21** (2), 441–466.

—, BUENO DE MESQUITA, E. and FRIEDENBERG, A. (2010). *Creating Incentives and Selecting Good Types Revisited*. Tech. rep., Harris School of Public Policy Studies.

BANKS, J. S. and SUNDARAM, R. K. (1998). Optimal retention in agency problems. *Journal of Economic Theory*, **82** (2), 293–323.

BARRO, R. J. (1973). The control of politicians: An economic model. *Public Choice*, **14** (1), 19–42.

BERMAN, S. (2001). Review: Ideas, norms, and culture in political analysis. *Comparative Politics*, **33** (2), 231–250.

BERNHARD, M., REENOCK, C. and NORDSTROM, T. (2003). Economic performance and survival in new democracies. *Comparative Political Studies*, **36** (4), 404–431.

BESLEY, T. (2006). *Principled Agents? The Political Economy of Good Government*. Oxford University Press.

— and BURGESS, R. (2002). The political economy of government responsiveness: Theory and evidence from india. *The Quarterly Journal of Economics*, **117** (4), 1415–1451.

BOIX, C. (2003). *Democracy and redistribution*. Cambridge studies in comparative politics, Cambridge University Press.

BRENDER, A. and DRAZEN, A. (2009). Consolidation of new democracy, mass attitudes, and clientelism. *American Economic Review*, **99** (2), 304–09.

CANES-WRONE, B., HERRON, M. C. and SHOTTS, K. W. (2001). Leadership and pandering: A theory of executive policymaking. *American Journal of Political Science*, **45** (3), 532–550.

CAPOCCIA, G. and KELEMEN, R. D. (2007). The study of critical junctures: Theory, narrative, and counterfactuals in historical institutionalism. *World Politics*, **59** (3), 341–369.

CHAUVET, L., COLLIER, P. and HOEFFLER, A. (2007). *The Cost of Failing States and the Limits to Sovereignty*. Working papers, World Institute for Development Economic Research (UNU-WIDER).

DI PALMA, G. (1990). *To craft democracies: an essay on democratic transitions*. University of California Press.

DIAMOND, L. (1999). *Developing democracy: toward consolidation*. Developing Democracy, Johns Hopkins University Press.

—, HARTLYN, J., LINZ, J. J. and LIPSET, S. M. (1999). *Larry Diamond*, *Jonathan Hartlyn*, *Juan J. Linz*, *and Seymour MartinLipset*. Lynne Rienner, 2nd edn.

FEARON, J. (2011). Self-enforcing democracy. *Quarterly Journal of Economics*, **126** (4), forthcoming.

FEARON, J. D. (1999). Electoral Accountability and the Control of Politicians: Selecting Good Types versus Sanctioning Poor Performance. In A. Przeworski and S. C. Stokes (eds.), *Democracy, Accountability, and Representation*, *99*, Cambridge.

FEREJOHN, J. (1986). Incumbent performance and electoral control. *Public Choice*, **50** (1), 5–25.

GASIOROWSKI, M. J. and POWER, T. J. (1998). The Structural Determinants of Democratic Consolidation: Evidence from the Third World. *Comparative Political Studies*, **31** (6), 740–771.

HUNTINGTON, S. (1991). *The third wave: democratization in the late twentieth century*. The Julian J. Rothbaum distinguished lecture series, University of Oklahoma Press.

Jones, B. F. and Olken, B. A. (2009). Hit or miss? the effect of assassinations on institutions and war. *American Economic Journal: Macroeconomics*, **1** (2), 55–87.

Kingdon, J. W. (1995). *Agendas, alternatives, and public policies / John W. Kingdon*. Longman, New York :, 2nd edn.

Lee, D. (2001). *The electoral advantage to incumbency and voters' valuation of politicians' experience: a regression discontinuity analysis of close elections*. Tech. rep., Center for Labor Economics, University of California, Berkeley.

Lee, D. S., Moretti, E. and Butler, M. J. (2004). Do voters affect or elect policies? evidence from the u. s. house. *The Quarterly Journal of Economics*, **119** (3), 807–859.

Lieberman, R. C. (2002). Ideas, institutions, and political order: Explaining political change. *American Political Science Review*, **96** (04), 697–712.

Linz, J. and Stepan, A. (1978). *The Breakdown of democratic regimes*. Johns Hopkins University Press.

— and — (1996). *Problems of democratic transition and consolidation: southern Europe, South America, and post-communist Europe*. Problems of Democratic Transition and Consolidation, Johns Hopkins University Press.

Maskin, E. and Tirole, J. (2004). The politician and the judge: Accountability in government. *American Economic Review*, **94** (4), 1034–1054.

Moehler, D. C. and Lindberg, S. I. (2009). Narrowing the legitimacy gap: Turnovers as a cause of democratic consolidation. *The Journal of Politics*, **71**, 1448 – 1466.

Moore, B. (1967). *Social origins of dictatorship and democracy: lord and peasant in the making of the modern world*. Beacon paperback, Beacon Press.

Myerson, R. B. (2006). Federalism and incentives for success of democracy. *Quarterly Journal of Political Science*, **1** (1), 3–23.

O'Donnell, G., Schmitter, P. and Whitehead, L. (1986). *Transitions from authoritarian rule: comparative perspectives*. No. v. 3 in Transitions from Authoritarian Rule, Johns Hopkins University Press.

Persson, T. and Tabellini, G. (2000). *Political Economics: Explaining Economic Policy*. The MIT Press.

— and — (2006). *Democratic Capital: The Nexus of Political and Economic Change*. NBER Working Papers 12175, National Bureau of Economic Research, Inc.

— and — (2009). Democratic capital: The nexus of political and economic change. *American Economic Journal: Macroeconomics*, **1** (2), 88–126.

Przeworski, A. (1991). *Democracy and the market: political and economic reforms in Eastern Europe and Latin America*. Studies in rationality and social change, Cambridge University Press.

—, Stokes, S. and Manin, B. (1999). *Democracy, accountability, and representation*. Cambridge studies in the theory of democracy, Cambridge University Press.

Putnam, R., Leonardi, R. and Nanetti, R. (1994). *Making democracy work: civic traditions in modern Italy*. Princeton paperbacks, Princeton University Press.

Rothstein, B. (1992). Explaining swedish corporatism: The formative moment. *Scandinavian Political Studies*, **15** (3), 173–191.

Rustow, D. A. (1970). Transitions to Democracy: Toward a Dynamic Model. *Comparative Politics*, **2** (3), 337–363.

Snyder, J. M. and Stromberg, D. (2010). Press coverage and political accountability. *Journal of Political Economy*, **118** (2), 355–408.

Svolik, M. W. (2010). *Learning to Love Democracy: Electoral Accountability, Government Performance, and the Consolidation of Democracy*. Tech. rep., University of Illinois at Urbana-Champaign.

Thelen, K. (1999). Historical institutionalism in comparative politics. *Annual Review of Political Science*, **2** (1), 369–404.

— and Steinmo, S. (1992). Historical institutionalism. In K. T. Sven Steinmo and F. Longstreth (eds.), *Structuring Politics*, New York: Cambridge University Press.

Tirole, J. (1996). A theory of collective reputations (with applications to the persistence of corruption and to firm quality). *Review of Economic Studies*, **63** (1), 1–22.

Weingast, B. (1997). The political foundation of democracy and the rule of law. *American Political Science Review*, **91** (2), 245–263.