Forever Deterred, the Highest Crimes? A Difficulty and Its Implications*

Weijia Li† Yang Xie‡

September 16, 2024

Abstract

We analyze a model of transgression and impeachment between a President and a Congress, where the circumstances determining whether the Congress will impeach a transgressing President become known only after the President has transgressed. We show a general difficulty in deterring such transgression: constant deterrence leads to a lack of new knowledge about the underlying circumstances, which evolve path-dependently and stochastically, making the threat of impeachment increasingly similar to a 50/50 coin toss over time. Forever deterrence is thus impossible whenever the President's gain from transgression is greater than the punishment from impeachment, however slightly. Stochastic cycles between deterrence and transgression may thus appear, with each transgression becoming a critical juncture predictive of future dynamics. For a Congress of many members, as time goes by with constant deterrence, any super-majority rule for conviction may almost surely acquit any transgressing President.

*We thank Scott Ashworth, Alberto Bisin, Kyle Chauvin, Gabriele Gratton, Guojun He, Alexander Jakobsen, Ye Jin, Matthew Kahn, Ethan Kaplan, Urmee Khan, Zhaotian Luo, Roger Myerson, Mariann Ollár, Marcin Pęski, Ariel Roginsky, Gérard Roland, Konstantin Sonin, Michael Ting, Scott Tyson, Fan Wang, Shaoda Wang, Melanie Wasserman, Yinxi Xie, Dingxin Zhao, Congyi Zhou, and participants in seminars at Monash and NYU Shanghai and 2023 SIOE and 2024 GAMES Meetings for their valuable comments.

†Department of Economics, Monash University; weijia.li@monash.edu.

‡Department of Economics, University of California, Riverside; yang.xie@ucr.edu.

1 Introduction

Given the often grave consequences incurred by the highest crimes, i.e., misconduct or transgression of the most senior government officials against law or political norms, it is natural for society to be concerned with deterrence of such crimes (e.g., Huq, 2018; Tribe and Matz, 2018). On deterrence, economists have long understood that it relies on credible threats to punish (e.g., Schelling, 1960; Becker, 1968; Tirole, 1988; Powell, 1990; Myerson, 2019). In the context of the highest crimes, such threats are often provided through a political or highly politicized process, such as electoral accountability, an impeachment proceeding, or potential rebellion of significant stakeholders (e.g., Madison, 1845, p. 528; Baumgartner and Kada, 2003; Myerson, 2008; Ohnesorge, 2020; Ginsburg et al., 2021; Monaghan et al., 2024). Regardless of the exact form of such deterrence, a difficulty may arise, exactly when the deterrence has been constantly successful: now that there have been no recent cases of such crimes, how could the threat to punish remain credible?

In this paper, in a simple model, we formalize this difficulty, analyze its theoretical foundation, and explore its implications for institutional dynamics and design. Not losing generality, the model is a stylized game of transgression of a President (P) and impeachment by the Congress (C). In each period, P will choose whether to transgress for a gain, followed by C choosing whether to punish P by impeaching him, whereas C's incentive to impeach, i.e., the state of the world, follows a potentially stochastic process. This reading of impeachment as punishment follows the literature on constitutional law, where to punish is viewed as a critical component in the purpose of impeachment (e.g., Tribe and Matz, 2018; Crespo, 2019).

The novel feature of our model is that the current state of the world is known to the players once and only once P transgresses in the current period. This feature follows the observation made in the literature of comparative and constitutional law, that whether a high crime will be punished is primarily a political, rather than a legal question, even when supposedly apolitical experts, such as prosecutors and judges, are involved, and the answer to the question often depends on the exact circumstances – political, social, economic, moral, etc. – at the time (e.g., Baumgartner and Kada, 2003; Tribe and Matz, 2018; Crespo, 2019; Ohnesorge, 2020; Ginsburg et al., 2021; Bowman, 2024; Monaghan et al., 2024). The general incompleteness of constitutional law on high crimes does not help much in clarifying the situation ex ante, either (e.g., Story, 1833; Huq, 2018; Tribe and Matz, 2018; Congressional Research Service, 2023;

Bowman, 2024). As a result, our knowledge about these exact circumstances at each time is generally limited ex ante, and many of them would become known only after the high crime has happened. We thus have to estimate ex ante whether a high crime would be punished based on what has happened before, i.e., the circumstances of the past cases of such crimes. So do the players in our model.

In this setting, we show that, whenever the gain from transgression for P is greater than the punishment from impeachment, it is always possible, depending on the nature of the stochastic process of the state of the world, for a particular phenomenon to emerge in equilibrium: constant deterrence of transgression over time will *cause* P to transgress eventually. The difficulty in deterring the highest crimes is thus formalized.

Further analysis reveals the theoretical foundation of this difficulty. We show that this difficulty arises only when the state of the world evolves path-dependently *and* stochastically, rather than path-independently or deterministically, *and* when the transgressions in focus are high crimes, instead of low, petty, ordinary crimes. This is because, it is only when the focal transgressions are high crimes that constant deterrence will generate a lack of new knowledge about the state of the world, since one has nowhere else to learn from when the transgressions are constantly deterred. It is also only when the world is path-dependent and uncertain that the effect of this lack of new knowledge will accumulate over time, making the past cases of transgressions increasingly irrelevant in estimating the current state of the world. This causes P to see whether he will be impeached for transgressing increasingly as a 50/50 coin toss, so he will eventually transgress at a time when the threat to punish becomes sufficiently uncredible, even if the gain from transgression is only slightly greater than the punishment. In this sense, the formalized difficulty in deterring the highest crimes is founded upon an intrinsic epistemic trap in constant deterrence of such crimes, which, only in a path-dependent yet uncertain world, will become increasingly consequential over time with constant deterrence.

Given that assuming a path-dependent yet uncertain world is not too wild (e.g., Callander and Martin, 2017, p. 13–15), focusing on high crimes in this setting delivers a few implications on institutional design and dynamics of institutional performance. First, because of the epistemic trap under constant deterrence of high crimes, between the threat to punish and the severity of punishment, the former is not always credible, and deterrence must thus rely on the latter. Indeed, in the limit over time with constant deterrence, any punishment greater than the potential gain would be able to deter P from transgressing. It is thus possible for transgression to be forever deterred, but only

when institutional arrangements are made to ensure the punishment for committing a high crime to be greater than the potential gain.

Second, since constant deterrence, at least for a limited period of time, can appear regardless of the institutional arrangements that may make the punishment greater or smaller than the gain from transgression, and since no new knowledge about the underlying state of the world is revealed under constant deterrence, observing only a still ongoing period of constant deterrence cannot reveal whether the current institution would allow the constant deterrence to last forever. In this sense, anyone outside the transgression-impeachment game, say a social scientist, political journalist, or concerned citizen, would not be able to learn much about the institution from an ongoing period of constant deterrence, and would thus have to wait for another time when the institution is under stress, i.e., when another transgression happens.

Third, given any punishment smaller than the potential gain from transgression, because the world is path-dependent, when transgression eventually happens, the revelation of the state of the world will be critical in determining the institutional dynamics that follows. A great current incentive for C to impeach P will deter transgression for a potentially lengthy period. A not-so-great current incentive for impeachment, on the contrary, can leave the threat to punish in the near future quite uncredible, leading to consecutive transgressions in potentially many periods to come. Therefore, stochastic cycles between constant deterrence and consecutive transgressions may appear, and each time when transgression happens – especially the time of the first transgression after a long period of constant deterrence – can be viewed as a critical juncture indicative of future institutional performance.

Finally, if C consists of multiple members, with their incentives to punish a transgressing P being path-dependent, uncertain, and mutually independent, and if the current state of these incentives will become known still once and only once P transgresses, then, as time goes by with constant success, the probability for each Congress member to vote to convict a transgressing P will converge to one-half. Therefore, in this limit, if C consists of many members, any super-majority rule will almost surely acquit any transgressing P, making the threat to punish hardly credible ex ante, regardless of the relative size of the transgression gain and punishment. A minority rule, on the other hand, will almost surely convict any transgressing P, so that forever deterrence would become possible, even when the transgression gain is greater than the punishment. Therefore, by adjusting the voting rule for conviction in impeachment, an institutional designer may make use of the epistemic trap in constant deterrence of high crimes to

achieve forever deterrence.

The economics literature on deterrence of crimes has primarily focused on ordinary crimes, and has shown theoretically and empirically that such crimes are more responsive to certainty of apprehension than to severity of punishment (e.g., survey by Chalfin and McCrary, 2017). We suggest that high crimes are fundamentally different, since there are few alternative settings where the underlying circumstances of such crimes can be learned. As a result, constant deterrence can eventually make the deterrence uncredible, and one may have to rely on severe punishment in deterring high crimes.

On high crimes, a thread of literature in political economy has focused on institutions that deter abuse of power, mainly through facilitating a coordinated threat to punish such abuse (e.g., Przeworski, 2006; Weingast, 1997; Myerson, 2008; Fearon, 2011; Svolik, 2012). We show that this approach, and any approach that focus on making a credible threat, may be weakened by their own success, because of the epistemic trap associated with constant deterrence of high crimes. Institutions that make use of this epistemic trap, such as severe punishment and minority rule for conviction in impeachment, may thus deserve more attention in institutional design.

Recent history has shown that long-established, advanced democracies are not exempt from the highest crimes and, in particular, transgressions of democratic institutions and norms by the head of the government (e.g., Bowman, 2024). This observation is at odds with the conventional wisdom that long, successful experience with democratic governance should foster and even immunize democracy against such transgressions (e.g., Przeworski, 2006; Fearon, 2011; Bidner and Francois, 2013; Besley and Persson, 2019; Acemoğlu and Robinson, 2022; Acemoğlu et al., Forthcoming). Reconciling the literature with the observation, we show that constant deterrence of such transgressions can become its own enemy. Unlike most of the literature on democratic backsliding and executive transgressions (e.g., survey by Grillo et al., 2024), our result does not depend on any bounded rationality or informational disadvantages on the side of the custodian of democracy, i.e., the Congress in our model, but originates only from the high-crime nature of such transgressions and the political nature of its punishment.[1]

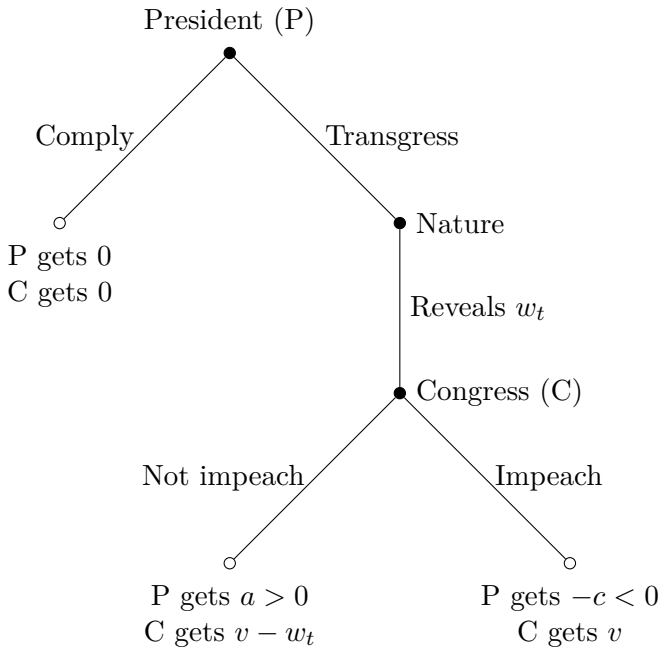
In the most general sense, our paper also contributes to the economics literature on the consequences of knowledge deficiency, which has focused on implications of *exogenous* lacks of knowledge on managing central planning, catastrophic risks, and

[1]For examples of the literature on democratic backsliding, see Acemoğlu et al. (2013), Howell and Wolton (2018), Svolik (2020), Chiopris et al. (2021), Matějka and Tabellini (2021), Miller (2021), Grillo and Prato (2023), Luo and Przeworski (2023), and Gratton and Lee (2024).

power relations (von Hayek, 1945; Weitzman, 1974, 2009, 2011; Stern, 2013; Huang et al., 2017; Conte and Kelly, 2021; Li et al., 2022). In our setting, new knowledge about the world can be *endogenously* hidden by a desirable outcome, i.e., successful deterrence. As a result, information about the key institution, i.e., whether the punishment for transgression is greater than the potential gain, cannot be revealed by an ongoing period of desirable outcomes, i.e., constant deterrence. This suggests a limit upon institutional learning where new knowledge can be generated only in a narrow set of contexts, for example, the deterrence of the highest crimes.

This paper is unfolded as follows. Section 2 presents the baseline setup of the model. Section 3 analyzes the model, formalizes the difficulty in deterring high crimes, and delivers the implications, except for the one about voting rules for conviction in impeachment, which is delivered in Section 4 by incorporating multiple members to the Congress. Proof of results and additional results for robustness of the main results are gathered in Appendix.

2 Model



Players know time, $s < t$, and state of the world, w_s, at last transgression.

Figure 1: Deterrence of high crimes, period t

Consider that in each period t there are a President (P) and a Congress (C), and a game between them as in Figure 1. P first chooses whether to act in a way complying with or transgressing the law and norms about the presidency. If P chooses to comply, then both P and C will receive a zero payoff as default, and the period will end there and transition to period $t+1$.

If P chooses to transgress instead, then C will later have a chance to decide whether to punish P by impeachment. If C indeed chooses to impeach P, then P will pay an exogenous cost $c > 0$, while C will receive an exogenous payoff of v, and the period will end there and transition to period $t+1$.

If C instead chooses not to impeach P, then P will receive an exogenous, positive gain $a > 0$ from transgression, while C will receive $v - w_t$, where w_t denotes the net incentive for C to impeach P for transgression, and the period will end there and transition to period $t+1$. Here we let w_t carry a subscript, t, and we interpret it as the state of the world. This approach follows the observation in the literature of comparative and constitutional law: impeachment is an ultimately political solution to a kind of political crisis in which, because of the focal high crimes, popular confidence in the incumbent or the maintenance of constitutional governance is in danger; whether impeachment will happen and how it will works out thus depend on the exact circumstances at the time, be them political, social, economic, or moral (e.g., Baumgartner and Kada, 2003; Tribe and Matz, 2018; Crespo, 2019; Ohnesorge, 2020; Ginsburg et al., 2021; Bowman, 2024; Monaghan et al., 2024).

State of the world, conditional revelation. This dependence on the exact circumstances at the time of a high crime suggests that social knowledge about these circumstances should be generally limited before the high crime is committed. In addition, since the nature of constitutional law is not concerned with being specific, but general, adaptable, and flexible, such law is often incomplete about the subject of high crimes, and thus does not help much in predicting how a high crime will be dealt with in reality (e.g., Story, 1833, p. 287; Huq, 2018, p. 1515; Tribe and Matz, 2018, p. 45; Bowman, 2024, p. 99).[2] As a result, one today has to estimate ex ante whether

[2]On the general incompleteness of law, see, for example, Landes and Posner (1975) and Pistor and Xu (2003); on the theoretical arguments for constitutional law to be relatively silent on high crimes, see, for example, Huq (2018) and Crespo (2019). A prime example here would be Section 4 of Article II of the United States Constitution, which reads and only reads: "[t]he President, Vice President and all civil Officers of the United States, shall be removed from Office on Impeachment for, and Conviction of, Treason, Bribery, or other high Crimes and Misdemeanors." The meaning of this "catchall phrase,"

and how a transgression will be punished, and, when doing so, she has to rely on her understanding of the circumstances, not merely the judicial decisions, if any, around the past cases of transgressions (e.g., Tribe and Matz, 2018; Congressional Research Service, 2023, p. 839; Bowman, 2024).

Considering this, here comes the key feature of our model, which we call the conditional revelation of the state of the world: as in Figure 1, at the beginning of period t, the players do not know the exact value of the state of the world, w_t; instead, if and only if P chooses to transgress, then nature will reveal the value of w_t to them, right before C chooses whether to impeach P. Coming back to the beginning of period t, what the players do know is the last period when a transgression happened before now, denoted by $s < t$, and the state of the world, i.e., the circumstances at that time, w_s.

High vs. low crimes. We see this conditional revelation as the key difference between high crimes, as modeled here, and low, petty crimes, for which the players would always have abundant knowledge about the state of the world ex ante. This is because, for low crimes, there are many everyday settings, in addition to the setting where P is holding a high office, for one to learn about the state of the world relevant to these crimes. In reality, w_t may well contain a component that is known by the players at the beginning of period t. We can thus interpret that component as a "low-crime" component. In Section 3, we discuss the dynamics of transgression and impeachment when low crimes are in focus.

State of the world, evolution. Given that information, how the state of the world has been evolving will be important in the players' estimation of w_t at the beginning of period t. We assume that

$$w_t = \rho w_{t-1} + \epsilon_t, \text{ where } 0 \leq \rho \leq 1, \text{ and } \epsilon_t \sim \mathcal{N}(0, \sigma^2), \text{ i.i.d., with } \sigma \geq 0. \quad (1)$$

Here $0 \leq \rho \leq 1$ guarantees that the influence of historical shocks on the present does not increase over time; without losing intuition, these shocks, which are unknown ex ante, are also simplified to follow a normal distribution, which is independent and identically

"high Crimes and Misdemeanors," is hereby "not defined in the Constitution or in statute," such that we are "left with an area of considerable vagueness" (Black and Bobbitt, 2018, p. 26; Congressional Research Service, 2023, p. 839). For the literature on the meaning and historical background of the term "high Crimes and Misdemeanors," see, for example, Tribe (1998), Black and Bobbitt (2018), Tribe and Matz (2018), Congressional Research Service (2023), and Bowman (2024).

distributed, with mean zero.

Note that this approach to modeling how the world evolves is sufficiently general to cover various world views along two dimensions:

- $0 < \rho \leq 1$ would indicate path-dependence, which may arise for legal, social, political, economic, or historical reasons and may well be endogenous and micro-founded, whereas $\rho = 0$ would indicate path-independence;
- $\sigma = 0$ would represent a deterministic world, whereas $\sigma > 0$ would allow the world to be uncertain, while the uncertainty may, again, arise for various reasons and be micro-founded.

It thus allows us to analyze the role of how the world evolves in deterring high crimes.

Complete the setup. We assume that the players are myopic, i.e., they care only about their payoffs within each period, not looking beyond the current period. This is justifiable if we interpret each period in our model as the finite term of office of the players. In Appendix B, we show that our main results are robust with respect to the players being infinitely forward-looking, when symmetric, pure-strategy Markov-perfect equilibria are considered, where "symmetric" means all Ps adopting the same strategy.

We will use backward induction to solve the model. For simplicity while not losing intuition, we assume that C will not impeach P if indifferent between impeaching and not impeaching; P will comply if indifferent between transgressing and complying.

3 Solution and Implied Dynamics

We now solve the game at period t. In case that P does transgress, C will impeach P if and only if $w_t > 0$. Given that, P's estimate of the probability of punishment is thus $\mathbf{P}[w_t > 0 \mid w_s, s]$. He will thus transgress in the first place if and only if

$$a \cdot (1 - \mathbf{P}\left[w_t > 0 \mid w_s, s\right]) - c \cdot \mathbf{P}[w_t > 0 \mid w_s, s] > 0. \tag{2}$$

We have thus solved the game at period t:

Proposition 1 (Solution to period t)**.** *At period t, P will transgress if and only if the threat to punish is sufficiently uncredible, i.e.,*

$$\mathbf{P}[w_t > 0 \mid w_s, s] < \frac{a}{a + c}; \tag{3}$$

when P transgresses, C will impeach P if and only if $w_t > 0$.

To understand the dynamics of transgression and impeachment, we need to examine how this conditional probability of punishment, $\mathbf{P}[w_t > 0 \mid w_s, s]$, depends on $t - s$ and w_s, i.e., how distant the last transgression has been in time and the revealed state of the world at that time, relative to the critical threshold, $a/(a + c)$. To do that, note that, given how the world evolves, P sees that

$$w_t = \rho w_{t-1} + \epsilon_t = \rho^{t-s} w_s + \sum_{\tau=s+1}^{t} \rho^{t-\tau} \epsilon_\tau. \tag{4}$$

Since all ϵ_τs are independent, identically distributed, and follow $\mathcal{N}(0, \sigma^2)$, P sees that

$$(w_t \mid w_s, s) \sim \mathcal{N}\left(\rho^{t-s} w_s, \sum_{\tau=s+1}^{t} \rho^{t-\tau} \cdot \sigma^2\right). \tag{5}$$

Different ways of how the world evolves, i.e., different values of ρ and σ, would thus render different estimates of the conditional distribution, $(w_t \mid w_s, s)$, and, therefore, of the conditional probability of punishment, $\mathbf{P}[w_t > 0 \mid w_s, s]$.

Path-dependent, uncertain world. We first consider a path-dependent, uncertain world, i.e., $0 < \rho \leq 1$ and $\sigma > 0$; we will discuss alternatives below. Given that, the conditional distribution of the state of the world is

$$(w_t \mid w_s, s) \sim \begin{cases} \mathcal{N}\left(\rho^{t-s} w_s, \frac{1-\rho^{t-s}}{1-\rho}\sigma^2\right) & \text{if } \rho \in (0,1), \\ \mathcal{N}\left(w_s, (t-s)\sigma^2\right) & \text{if } \rho = 1. \end{cases} \tag{6}$$

This implies that the conditional probability of punishment is

$$\mathbf{P}[w_t > 0 \mid w_s, s] = \begin{cases} \Phi\left(\frac{\rho^{t-s} w_s}{\sigma \cdot \sqrt{\frac{1-\rho^{t-s}}{1-\rho}}}\right) & \text{if } \rho \in (0,1), \\ \Phi\left(\frac{w_s}{\sigma\sqrt{t-s}}\right) & \text{if } \rho = 1, \end{cases} \tag{7}$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. We can thus observe the following properties of this conditional probability:

Lemma 1 (Credibility of threat to punish)**.** *In any path-dependent, uncertain world, it would be more likely than not for C to impeach a transgressing P, if and only if C*

impeached P at the last transgression, but, as time goes by with constant deterrence, the threat to punish will look increasingly similar to a 50/50 coin toss. Mathematically, given any $\rho \in (0, 1]$ *and* $\sigma > 0$*, for any* $t > s$*,*

- *if* $w_s \leq 0$*, then* $\mathbf{P}[w_t > 0 \mid w_s, s] \leq 0.5$*, is increasing in* $t - s$*, and converges to* 0.5 *as* $t - s$ *increases;*
- *if* $w_s > 0$*, then* $\mathbf{P}[w_t > 0 \mid w_s, s] > 0.5$*, is decreasing in* $t - s$*, and converges to* 0.5 *as* $t - s$ *increases.*

The intuition of Lemma 1 is as follows. First, why does whether C impeached P at the last transgression, i.e., whether $w_s > 0$, indicate whether it would be more likely than not for C to impeach a transgressing P now? This is because, since the world is path-dependent and no new empirical knowledge has been revealed about the state of the world since the last transgression, any current estimate of the current state of the world has to be centered around the last revealed state, adjusted with the degree of path-dependence, i.e., w_s in the case of $\rho = 1$, and $\rho^{t-s} w_s$ in the case of $\rho \in (0, 1)$.

Second, why is the monotonic convergence to a 50/50 coin toss? In the case of $\rho \in (0, 1)$, this is because the influence of historical shocks dwindles over time, which is set mechanically by a factor of $\rho \in (0, 1)$. As time goes by with constant deterrence, because of the conditional revelation of the state of the world, no new information about the state of the world has been revealed. The conditional distribution of the current state of the world would thus converge to a normal distribution with mean zero and a finite variance, i.e.,

$$(w_t \mid w_s, s) \xrightarrow{d} \mathcal{N}\left(0, \frac{\sigma^2}{1 - \rho}\right) \text{ as } t - s \to \infty, \quad \text{if } \rho \in (0, 1). \tag{8}$$

This makes impeachment look closer and closer to a 50/50 coin toss.

In the case of $\rho = 1$, on the contrary, the influence of historical shocks would remain constant, but its relevance would be dwindling, still, even without any mechanical depreciation of the influence of historical shocks or forgetfulness of players, i.e., $\rho = 1$. This is because, without new revelation, the uncertainty in the evolution of the world would pile up over time and is not bounded from above. The conditional distribution of the state of the world would thus converge to a normal distribution with an infinite variance, i.e.,

$$(w_t \mid w_s, s) \xrightarrow{d} \mathcal{N}(w_s, \infty) \text{ as } t - s \to \infty, \quad \text{if } \rho = 1. \tag{9}$$

This makes impeachment look closer and closer to a 50/50 coin toss, too.

Lemma 1 implies that, to fully characterize the dynamics of impeachment and transgression in a path-dependent, stochastic world, we need to compare the critical threshold for the conditional probability of punishment, $a/(a+c)$, with the limit of this conditional probability under constant deterrence, which is one-half. We observe the following lemma:

Lemma 2 (50/50 coin toss)**.** *A threat to punish that is equivalent to a 50/50 coin toss will not be able to deter P from transgressing, if and only if the potential gain from transgression is greater than the punishment from impeachment. Mathematically,*

$$0.5 < \frac{a}{a+c} \text{ if and only if } a > c. \tag{10}$$

The intuition of Lemma 2 is straightforward. If P considers the threat to punish to be equivalent to a 50/50 coin toss, then he will simply compare the gain from transgression with the punishment, and will transgress if and only if it is greater.

Lemma 2 implies that we need to separate two cases, i.e., whether the potential gain from transgression is greater than the punishment or not.

3.1 Transgression Gain Greater than Punishment

We first focus on the case where the gain is greater, however slightly, i.e., $a > c$. Lemmas 1 and 2 imply the following result:

Proposition 2 (The difficulty in deterring high crimes)**.** *Given any $\rho \in (0,1]$, $\sigma > 0$, and $a > c$,*

1. *if $w_s \leq 0$, then C will not impeach P at period s, and P will transgress at period $s+1$;*

2. *if $0 < w_s < \bar{w}$, where $\bar{w} \equiv (\sigma/\rho) \cdot \Phi^{-1}(a/(a+c)) > 0$, with $\Phi^{-1}(\cdot)$ being the inverse cumulative distribution function of the standard normal distribution, then C will impeach P at period s, but P will still transgress at period $s+1$;*

3. *if $w_s \geq \bar{w}$, then C will impeach P at period s, and there will always exist a period $T \in [s+2, \infty)$, such that P will not transgress from period $s+1$ to $T-1$, but will transgress at period T.*

In this sense, in any path-dependent, uncertain world, given any transgression gain greater than the punishment, transgression cannot be forever deterred.

We prove Proposition 2 in Appendix A.1. The intuition is as follows. Claim 1 is simply an application of Lemma 1 to any period right after an unpunished transgression. About Claim 2, having the last transgression punished, i.e., $w_s > 0$, may not be enough to deter transgression in the following period, since, by Lemma 1, the credibility of the threat to punish may drop too much even in just one period. Therefore, the last revealed state of the world has to be sufficiently against transgression, i.e., $w_s \geq \bar{w}$, to deter another transgression right after the last transgression.

About Claim 3, Lemma 1 implies that as time goes by with constant deterrence, the threat to punish will look closer and closer to a 50/50 coin toss, whereas Lemma 2 suggests a 50/50 coin toss will not be able to deter transgression, as long as the potential gain from transgression is greater than the punishment. Therefore, there must exist a period when the threat to punish will eventually become too uncredible to deter transgression.

The difficulty, formalized. Proposition 2 formalizes the difficulty in deterring high crimes: as long as the world is path-dependent yet uncertain, for any potential gain from transgression that is greater, however slightly, than the punishment, constant deterrence would make the threat to punish increasingly as uncredible as a 50/50 coin toss, eventually failing to deter, and making thus forever deterrence impossible.

Discussion: Alternative world views. A closer look at the other world views would help us better understand the theoretical foundation of this difficulty. Having a path-independent world corresponds to the case of $\rho = 0$. In this case, the current state of the world at each period is independent of past states of the world, i.e.,

$$w_t = \epsilon_t \sim \mathcal{N}\left(0, \sigma^2\right), \text{ i.i.d., with } \sigma \geq 0. \tag{11}$$

Its conditional distribution and the conditional probability of punishment are thus independent of the state of the world at the last transgression, i.e.,

$$(w_t \mid w_s, s) \sim \mathcal{N}\left(0, \sigma^2\right), \text{ with } \sigma \geq 0, \tag{12}$$

and thus

$$\mathbf{P}[w_t > 0 \mid w_s, s] = \begin{cases} 0.5 & \text{if } \sigma > 0, \\ 0 & \text{if } \sigma = 0. \end{cases} \tag{13}$$

P will thus always transgress, or always not transgress, in each period, independent of the last revelation and how distant it has been in history. The difficulty formalized in Proposition 2 thus does not exist.

Having a deterministic world corresponds to the case of $\sigma = 0$. In this case, the current state of the world at each period follows a deterministic process,

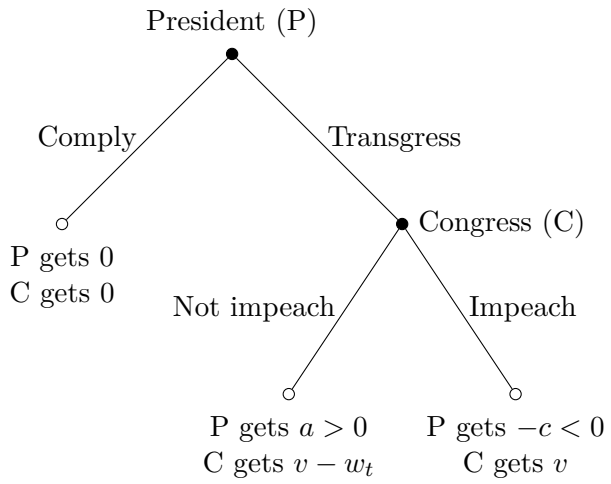
$$w_t = \rho w_{t-1} = \rho^2 w_{t-2} + \rho \epsilon_{t-1} = \cdots = \rho^{t-s} w_s, \text{ where } 0 \leq \rho \leq 1. \tag{14}$$

This implies that the revelation of the state of the world at the last transgression can perfectly indicate the current state, i.e., $w_t > 0$ if and only if $w_s > 0$. Therefore, once a transgression is punished, at each future period, P will understand that any transgression will be punished, so will not transgress. It is thus possible to forever deter transgression, and the difficulty formalized in Proposition 2 does not exist, either.

Discussion: Low crimes. We have shown that, in Proposition 2, it is because the world is path-dependent yet uncertain that lack of new knowledge, brought by constant deterrence, would become increasingly consequential in estimating the current state of the world. What is the role of high crimes, i.e., what would happen if the transgression in focus were a low, petty crime?

Figure 2 shows the structure of the game if the focal transgression were a low crime. Compared with Figure 1, as discussed in Section 2, the only difference is here the players know w_t at the beginning of period t, without the conditional revelation. By backward induction, if and only if $w_t \leq 0$, C will not punish any transgression and P will transgress. Conditional on the current state of the world, w_t, which is part of common knowledge at the beginning of each period t, P's decision is thus independent of the history of transgression and impeachment. Therefore, the difficulty formalized in Proposition 2 would not exist.

Summary. We have thus formalized in Proposition 2 the difficulty in deterring high crimes. We have also established that this difficulty comes from a combination of 1) a lack of new knowledge about the world, which is caused by constant deterrence, and the world being 2) path-dependent yet 3) uncertain. It is because high crimes are in



Players know current state of the world, w_t.

Figure 2: Deterrence of low crimes, period t

focus that constant deterrence would imply a lack of new empirical knowledge about the world. It is because the world is uncertain that this lack of new knowledge would be consequential in estimating the current state of the world and thus the probability of punishment. It is because the world is path-dependent that this lack of new knowledge would become *increasingly* consequential, such that constant deterrence would eventually *cause* the threat to punish to become too uncredible to deter the high crimes.

Numerical illustration. Some visualization may help to illustrate the dynamics described in Proposition 2. Figure 3 plots the typical dynamics by simulating a series of 500 periods. In this simulation, the initial revelation of the state of the world w_1 is set exogenously to be sufficiently against transgression, i.e., $w_1 \geq \bar{w} > 0$, so that P does not transgress at period 2. Starting from there, consistent with Lemma 1, the conditional probability of punishment declines over time with constant deterrence. Consistent with Proposition 2, it is at period 88 that the threat to punish eventually becomes too uncredible to deter transgression, i.e., $\mathbf{P}[w_{88} \mid w_1, 1] < a/(a+c)$. Right at this transgression, the revealed state of the world turns out to be so against transgression, i.e., $w_{88} \geq \bar{w}$, that no transgression happens for another lengthy period, but, consistent with Lemma 1, again, the conditional probability of punishment keeps declining over time as a result of constant deterrence.

Consistent with Proposition 2, it is at period 180 that the threat to punish eventually becomes too uncredible, again, to deter transgression. At this moment, the revealed

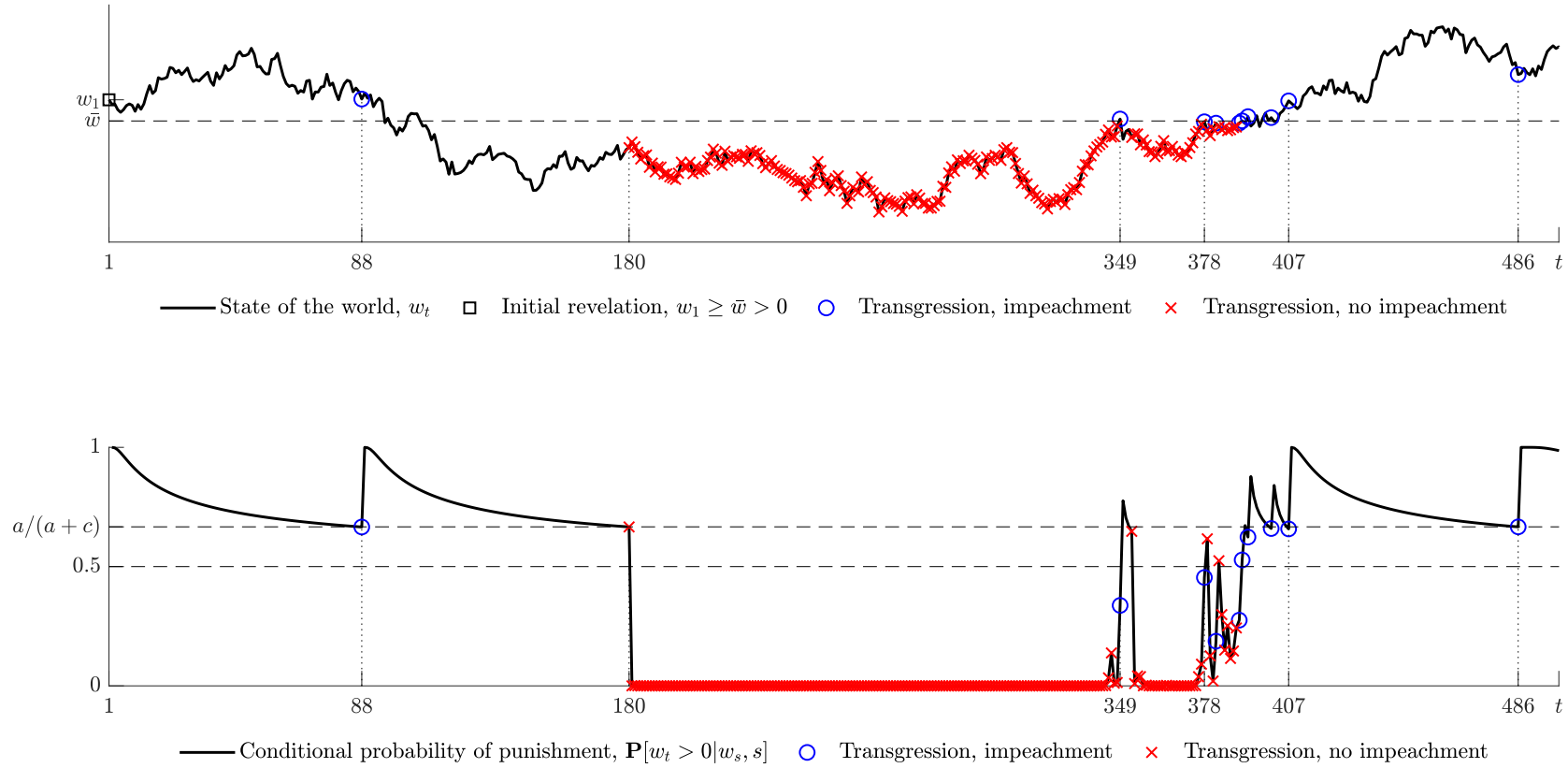


Figure 3: High crimes in a path-dependent, stochastic world, gain greater than punishment, typical dynamics

state of the world turns out to be too low even for C to impeach P, i.e., $w_{180} \leq 0$. Consistent with Proposition 2, this leads to another transgression right at period 181, and the revealed state of the world then still prevents C from impeaching P. This leads to another transgression at period 182. The same keeps happening at each period, until period 349.

From period 349 onward is a phase of frequent transgressions. There are a few periods, for example, period 378, when the transgressing P is impeached, but the revelation of the state of the world is not able to generate consolidated constant deterrence. It is only at period 407 that the world has moved, since the last transgression, quite far against transgression, leading to a lengthy period of deterrence until period 486. The revelation at period 486 turns out to be even more against transgression, so we can expect an even longer period of constant deterrence. That said, by Proposition 2, another transgression will eventually happen.

Stochastic cycles and critical junctures. This illustration suggests that, since success in deterrence does not generate new empirical knowledge about the world, and since such knowledge is revealed only when transgression happens, stochastic cycles between constant deterrence and consecutive transgressions may emerge, and any transgression can be viewed as a critical juncture for future dynamics. If the revealed state of the world is quite against transgression, it will secure punishment not only for the transgression at the moment, but also for potential transgressions in many coming periods, thereby deterring them for quite a while. If the revelation is not quite against transgression, instead, and, especially, if it prevents C from impeaching P at the moment, it may lead the system into a lengthy period of consecutive transgressions. Indeed, Proposition 2 and its proof in Appendix A.1 imply the following corollary:

Corollary 1 (Transgression, a critical juncture)**.** *The higher the revealed state of the world at a transgression, the longer the following period of successful deterrence, i.e., in Proposition 2, P will not transgress at period $t+1$ if and only if $w_s \geq \bar{w}$, and T is increasing in w_s.*

This is especially the case when a transgression happens after a long period of constant deterrence. In such cases, what we know about the world, i.e., the last revelation of the state of the world, would not be very indicative of the current state of the world. Mathematically, given $w_s \geq \bar{w} > 0$, i.e., given deterrence at period $s+1$, the coefficient of variation of the conditional distribution of the state of the world at period t, with

constant deterrence since period $s+1$, is

$$CV = \begin{cases} \frac{\sqrt{\frac{1-\rho^{t-s}}{1-\rho}}}{\rho^{t-s} w_s} \cdot \sigma & \text{if } \rho \in (0,1), \\ \frac{\sqrt{t-s}}{w_s} \cdot \sigma & \text{if } \rho = 1, \end{cases} \tag{15}$$

which is increasing in $t-s$. In this sense, the stake of the revelation of the state of the world in case of a current transgression would become especially significant.

3.2 Transgression Gain Not Greater than Punishment

Now we turn to the case where the potential gain from transgression is not greater than the punishment, i.e., $a \leq c$. Lemmas 1 and 2 imply the following result:

Proposition 3 (Forever deterrence possible under severe punishment)**.** *Given any* $\rho \in (0,1]$*,* $\sigma > 0$*, and* $a \leq c$*,*

1. *if* $w_s < \bar{w}$*, where* $\bar{w} \equiv (\sigma/\rho) \cdot \Phi^{-1}(a/(a+c)) \leq 0$*, then C will not impeach P at period s, and P will transgress at period* $s+1$*;*

2. *if* $\bar{w} \leq w_s \leq 0$*, then C will not impeach P at period s, but P will never transgress from period* $s+1$ *onward;*

3. *if* $w_s > 0$*, then C will impeach P at period s, and P will never transgress from period* $s+1$ *onward.*

In this sense, in any path-dependent, uncertain world, given any transgression gain not greater than the punishment, once deterred, transgression will be forever deterred.

We prove Proposition 3 in Appendix A.2. The intuition is as follows. About Claim 3, if the last revealed state of the world were against transgression, i.e., if $w_s > 0$, then Proposition 1 would imply impeachment at the last transgression. Lemma 1 would also imply that, as long as no more transgression has happened since the last transgression, although the threat to punish at any future period will decline over time, but in the limit it is still at least as credible as a 50/50 coin toss. On that, Lemma 2 implies that, since the punishment is sufficiently severe, i.e., $a \leq c$, a 50/50 coin toss would be sufficiently credible to deter any transgression. Transgression would thus be forever deterred.

About Claim 2, if the last revealed state of the world were for transgression, i.e., if $w_s \leq 0$, then Proposition 1 would imply no impeachment at the last transgression.

That said, Lemma 1 also implies that, if the last revealed state were not too friendly to transgression, either, i.e.,. if $w_s \geq \bar{w}$, the conditional probability of punishment would rise so much in one period that, at the period right after the last transgression, the threat to punish would become sufficiently credible to deter transgression. From then onward, by Lemma 1, the threat to punish would become more and more credible over time with successful deterrence. Transgression would thus be forever deterred, too.

Following the argument above, Claim 1 is thus the case where the last revealed state of the world was indeed very friendly to transgression, such that P will transgress at the period right after the last transgression. In addition to that, Proposition 1 implies no impeachment at the last transgression.

Numerical illustrations. Figures 4 and 5 illustrate the dynamics described in Proposition 3. In both figures, the initial revelation of the state of the world is sufficiently friendly to transgression, i.e., $w_1 < \bar{w}$, such that P will transgress at period 2, and following revelations of the state of the world keep P transgressing at each period for a while.

In Figure 4, it is at period 78 that the revealed state of the world turns out to be friendly, but not too friendly, to transgression, i.e., $\bar{w} \leq w_{78} \leq 0$. Consistent with Proposition 3, although C does not impeach P at period 78, P is deterred from transgressing at period 79. Once deterred, consistent with Lemma 1, the threat to punish becomes even more credible and converges to a 50/50 coin toss, as time goes by with constant deterrence. As a result, consistent with Proposition 3, transgression is forever deterred from there on.

In Figure 5, it is at period 242 that the revealed state of the world turns out to be against transgression, i.e., $w_{242} > 0$. Consistent with Proposition 3, C impeaches P at period 242, and P does not transgress at period 243. Once deterred, consistent with Lemma 1, the threat to punish becomes less credible over time with constant deterrence, but in the limit it is still at least as good as a 50/50 coin toss. Consistent with Proposition 3, this leads to forever deterrence from there on.

Severity of punishment as solution to the difficulty in deterring high crimes. Proposition 3 suggests that the difficulty in Proposition 2 will be solved, if the punishment for transgression is sufficiently severe. This is because, in the limit over time with constant deterrence, without new empirical knowledge about the world since the last transgression, apprehension is only as certain as a 50/50 coin toss. Therefore, in

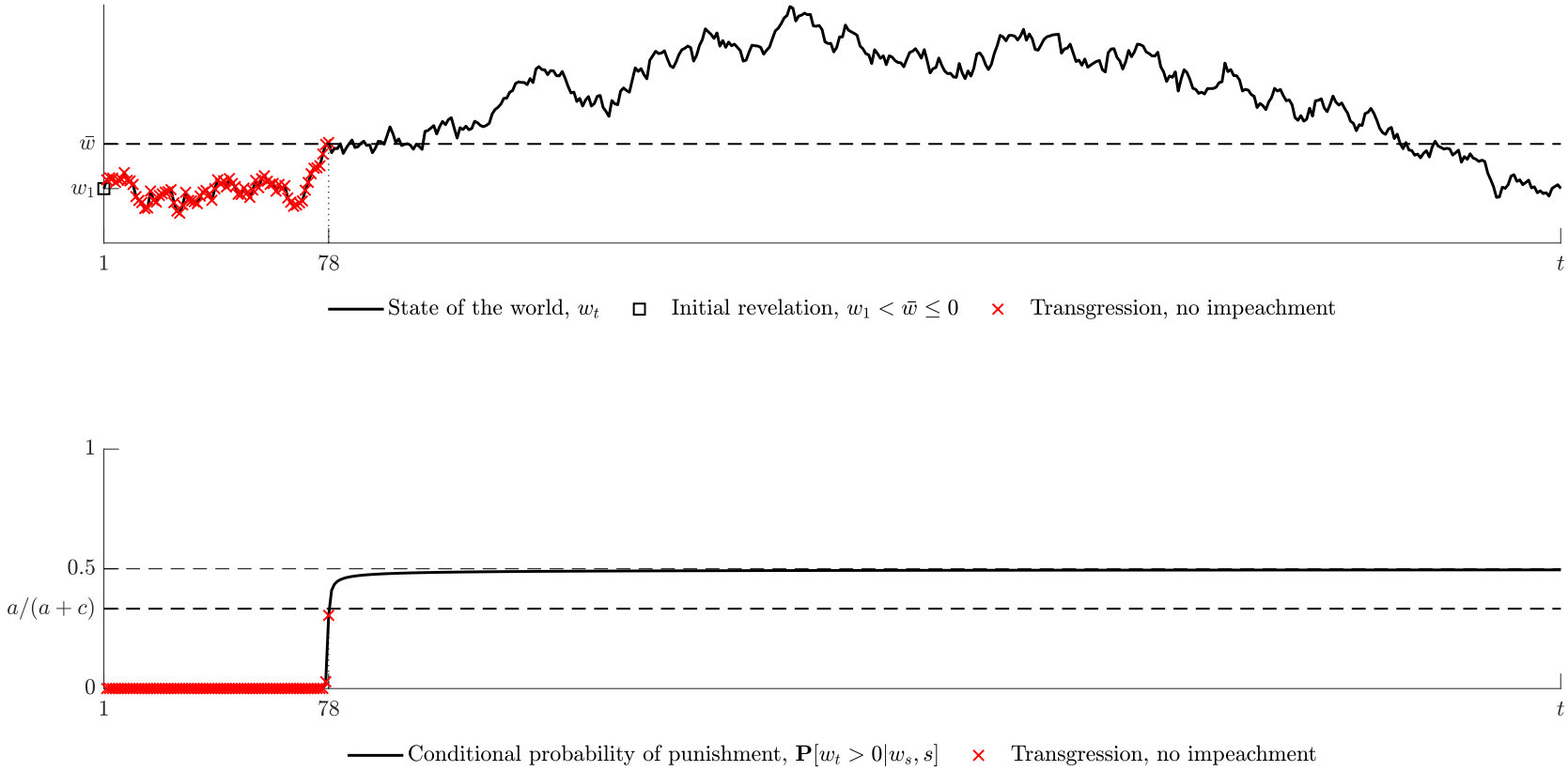


Figure 4: High crimes in a path-dependent, stochastic world, gain smaller than punishment, typical dynamics 1/2

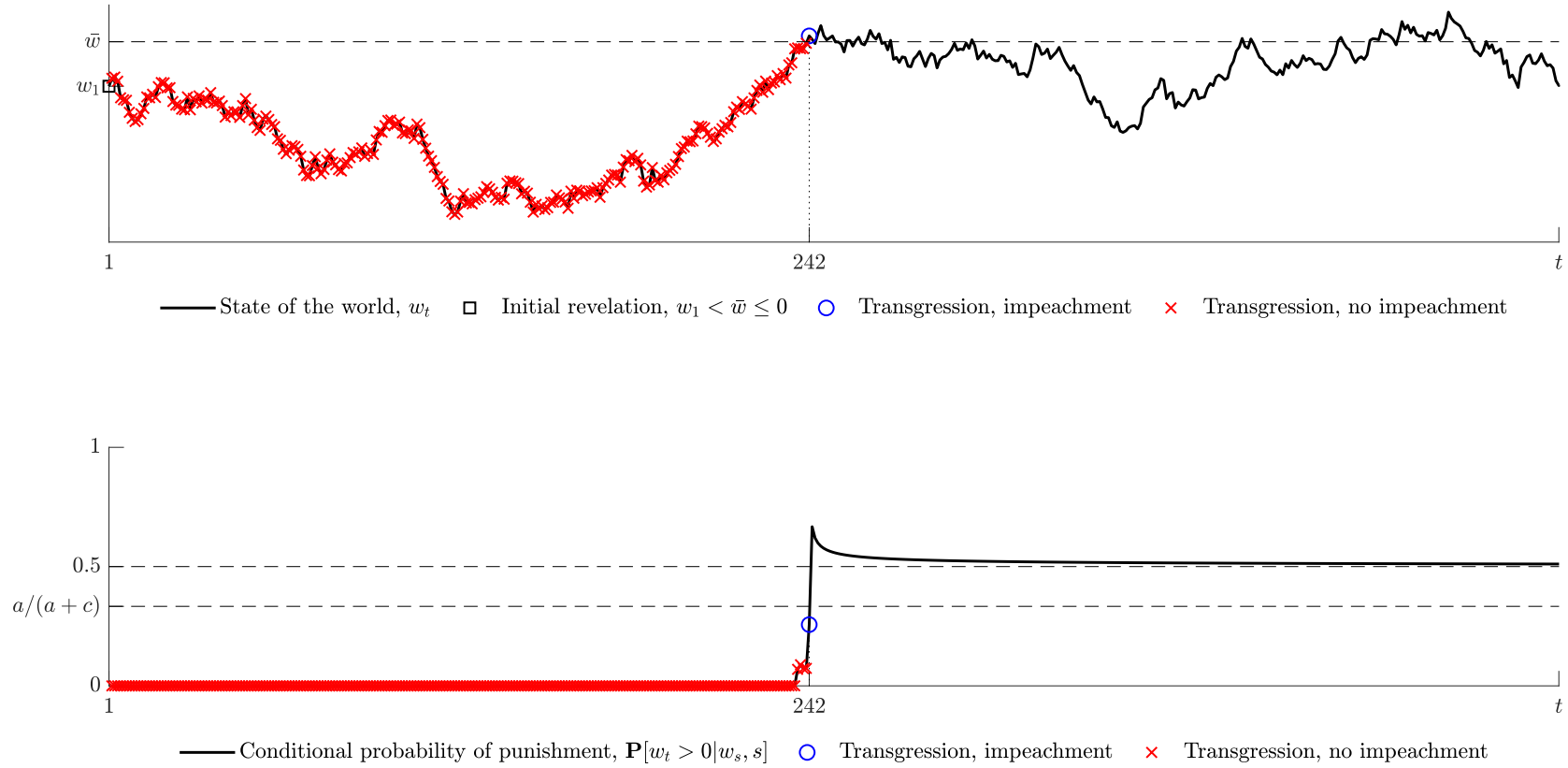


Figure 5: High crimes in a path-dependent, stochastic world, gain smaller than punishment, typical dynamics 2/2

this limit, on and only on severity of punishment can deterrence rely.

Examining the process approaching this limit further reveals the role of knowledge about the world, or the lack of it, in the dynamics, and how severity of punishment changes it. In Proposition 2, i.e., when the transgression gain is greater than the punishment, the deterrence-required level of credibility of the threat to punish, $a/(a+c)$, is quite high. Starting from a successful deterrence, the lack of new empirical knowledge, which is brought by constant deterrence given high crimes are in focus, makes the threat to punish become less and less credible and closer and closer to a 50/50 coin toss. Such a level of credibility is too low for deterrence, i.e., $0.5 < a/(a+c)$.

In Proposition 3, i.e., when the punishment is sufficiently severe, the deterrence-required level of credibility of the threat to punish is quite low. Starting from a successful deterrence, even if the lack of new empirical knowledge under constant deterrence makes the treat to punish become less and less credible over time, eventually the limit, a 50/50 coin toss, is still above the required level for deterrence, i.e., $0.5 \geq a/(a+c)$. The lack of new empirical knowledge under constant deterrence is thus the key to the difficulty in deterring high crimes, and to its solution, while the severity of punishment is pivotal in the eventual effect of it on deterrence.

3.3 Can Constant Deterrence Itself Reveal the Institution?

Proposition 3 shows that, in a path-dependent yet uncertain world, the exogenous parameter that determines whether the difficulty of deterring high crimes in Proposition 2 is the relative size of P's gain and punishment from transgression, an important feature of the institution surrounding the players. A question naturally follows: if someone cannot observe the institution directly, can she infer it through reading the dynamics of transgression and impeachment? Propositions 2 and 3 provide an answer about constant deterrence:

Corollary 2 (Constant deterrence uninformative about institution)**.** *Not knowing the exogenous parameters of the model, only observing a transgressing P being impeached at period $s < t$ and constant deterrence from then to the current period t, one cannot identify whether or not the institution will make the constant deterrence forever, i.e., whether $a \leq c$ or $a > c$; she can conclude $a > c$ once another transgression happens.*

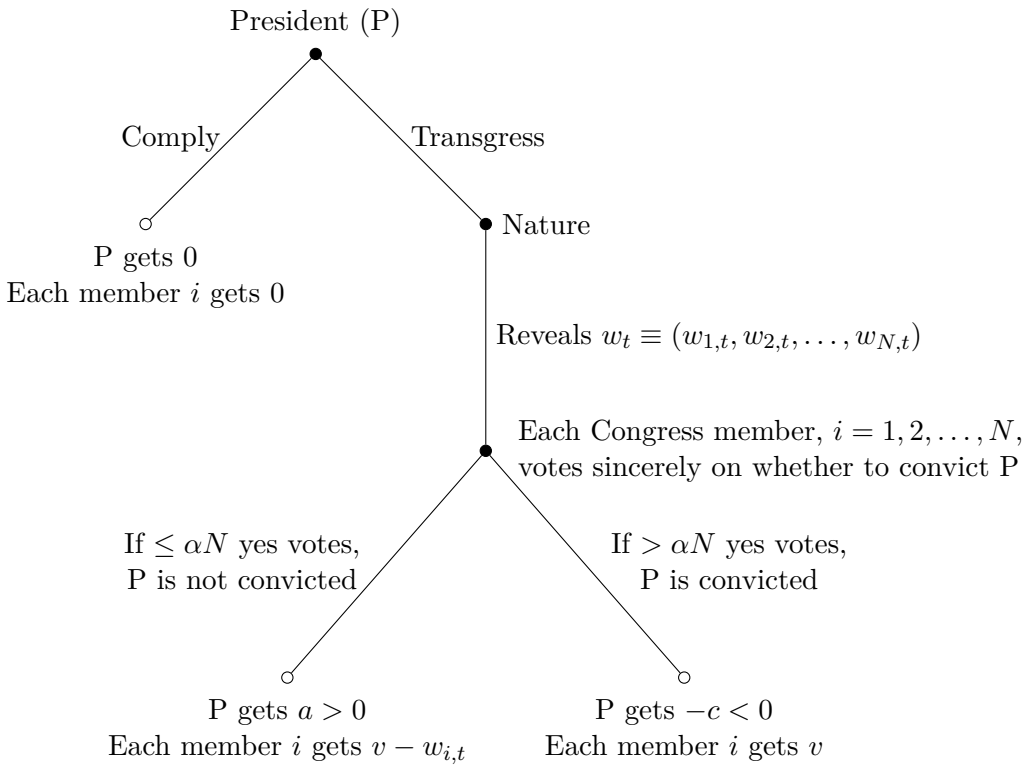
Corollary 2 suggests that only observing an ongoing period of constant deterrence will not tell us much about the underlying institution, i.e., between the cases of $a > c$

and $a \leq c$, which applies to the current game. Just to illustrate the idea, for example, both periods 88–97 in Figure 3 and periods 242–251 in Figure 5 are ten periods that begin with a transgressing P being impeached, followed by constant deterrence for nine periods. That said, in Figure 3 a transgression will eventually happen at period 180, whereas in Figure 5 transgression is forever deterred from period 243 onward. Therefore, not knowing the exogenous parameters of the model, anyone who only observes such institutional performance of ten periods cannot tell whether she is in Figure 3, i.e., the case of $a > c$, or Figure 5, i.e., the case of $a \leq c$, and only another transgression, as at period 180 in Figure 3, can inform her that she is in the case of $a > c$. In this sense, not only does constant deterrence generate a lack of new knowledge about the state of the world, as contained in the setup of our model, but it cannot tell much about the institution by itself; in order to learn about the institution, one has to wait for the institution to be under stress, i.e., when another transgression happens.

4 Congress of Multiple Members

So far we have assumed that C is a single player. Now consider an extension as in Figure 6. The main difference from Section 2 is that now the Congress consists of $N \geq 3$ members, each denoted by $i = 1, 2, \ldots, N$. With this Congress of multiple members, after P transgresses and Nature reveals the state of the world, each Congress member will vote sincerely, i.e., considering oneself to be pivotal when voting, on whether to convict P. Standard in the literature (e.g., Gehlbach, 2013, p. 13–14; Dziuda and Loeper, 2016, p. 1154), sincere voting is equivalent to assuming that strategic voters play weakly undominated voting strategies between the two voting options, or stage-undominated strategies. P will then be convicted if and only if there are greater than αN yes votes, where $\alpha \in [0, 1)$ is exogenous and represents the voting rule for conviction.

All the other features of the extension follow the model in Section 2, with accommodations necessary for a Congress of multiple members. About the state of the world, we assume that each Congress member faces her own net incentive to convict a transgressing P, $w_{i,t}$. The state of the world is thus a vector now, $w_t \equiv (w_{1,t}, w_{2,t}, \ldots, w_{N,t})$. We assume that each component of the state of the world evolves as modeled in Section 2, i.e., for any i, $w_{i,t} = \rho_i w_{i,t-1} + \epsilon_{i,t}$, where $0 \leq \rho_i \leq 1$ and $\epsilon_{i,t} \sim \mathcal{N}(0, \sigma_i^2)$, independent and identically distributed, with $\sigma_i \geq 0$. We further assume that all $w_{i,t}$s are mutually independent across all i, so that the Congress members can be interpreted as equally



Players know time, $s < t$, and state of the world, w_s, at last transgression.

Figure 6: Congress of $N \geq 3$ members, with voting rule $\alpha \in [0, 1)$, period t

powerful voting blocks.[3] As in Section 3, we focus on the path-dependent yet uncertain world, i.e., we further assume that, for any i, $0 < \rho_i \leq 1$ and $\sigma_i > 0$.

Since it is still the high crimes that are in focus, as in Section 2, we still have conditional revelation of the state of the world. The payoff structure is the same as in Section 2. As in Section 3, when analyzing this extension, we still assume that all the players are myopic, P will not transgress if indifferent between transgressing and complying, and each Congress member will vote not to convict P if indifferent between voting to convict or not. We still denote the period of the last transgression as s.

Given this setting, as in Proposition 1, at each period t, by backward induction, each Congress member i will vote to convict a transgressing P if and only if $w_{i,t} > 0$, while P will transgress if and only if the threat of conviction is sufficiently uncredible,

[3]Here we could have assumed that, in addition to the independent component, each $w_{i,t}$ may also include a common component. In that case, one can imagine that if the weight of this common component is sufficiently great relative to the weight of the independent component, all these Congress members will act similarly as the single-player Congress in Sections 2 and 3. Here we thus consider only the independent component.

i.e.,

$$\mathbf{P}[W_t > \alpha N \mid w_s, s] < a/(a+c), \tag{16}$$

where W_t is the number of convicting votes against a transgressing P.

Limit over time with constant deterrence. About this threat of conviction, since all $w_{i,t}$s are mutually independent across all i, Lemma 1 still applies. In particular, as time goes by with constant deterrence, the threat from each Congress member i to vote to convict P, conditional on the time and state of the world at the last transgression, s and w_s, will converge to a 50/50 coin toss, i.e.,

$$\mathbf{P}[w_{i,t} > 0 \mid w_s, s] \to 0.5 \text{ as } t - s \to \infty. \tag{17}$$

Therefore, in this limit, the conditional distribution of the number of votes to convict a transgressing P will converge to a Binomial distribution, with the number of Bernoulli trials being N and the success probability being 0.5, i.e.,

$$(W_t \mid w_s, s) \xrightarrow{d} W \sim \mathcal{B}(N, 0.5) \text{ as } t - s \to \infty. \tag{18}$$

By the Weak Law of Large Numbers, this implies that, in this limit over time with constant deterrence, as the size of the Congress, N, increases, it will be almost certain that the proportion of the Congress members voting to convict a transgressing P will be close to one half, i.e.,

$$\frac{W}{N} \xrightarrow{p} 0.5 \text{ as } N \to \infty. \tag{19}$$

Implication for voting rule. Applying the voting rule to this limit, we have the following proposition:

Proposition 4 (Voting rules)**.** *For any transgression gain and punishment for P, given any super-majority rule for conviction in impeachment, in the limit over time with constant deterrence, as the size of the Congress increases, it will be almost certain that the Congress will not convict any transgressing P and, therefore, P will transgress; given any minority rule, in the limit over time with constant deterrence, as the size of the Congress increases, it will be almost certain that the Congress will convict any*

transgressing P and, therefore, P will comply. Mathematically, for any $a > 0$ and $c > 0$,

$$\begin{aligned} &\textit{if } 0.5 < \alpha < 1, \textit{ then } \mathbf{P}\left[W > \alpha N\right] \to 0 < \frac{a}{a+c} \textit{ as } N \to \infty; \\ &\textit{if } 0 \leq \alpha < 0.5, \textit{ then } \mathbf{P}\left[W > \alpha N\right] \to 1 \geq \frac{a}{a+c} \textit{ as } N \to \infty. \end{aligned} \tag{20}$$

Proposition 4 suggests that any super-majority rule will impose another difficulty in deterring high crimes, in the similar style to Proposition 2, but even to a greater extent: when the Congress consists of many members, it is almost certain that transgression cannot be forever deterred, even if the transgression gain is smaller than the punishment, i.e., even if $a \leq c$. This is because, in this limit over time with constant deterrence, each Congress member will vote to convict a transgressing P just like a 50/50 coin toss, and the total number of convicting votes will be unlikely to overcome the super-majority rule. A minority rule, on the contrary, can solve the difficulty, since, when the Congress consists of many members, in the limit over time with constant deterrence, it is almost certain that a transgressing P will be convicted.

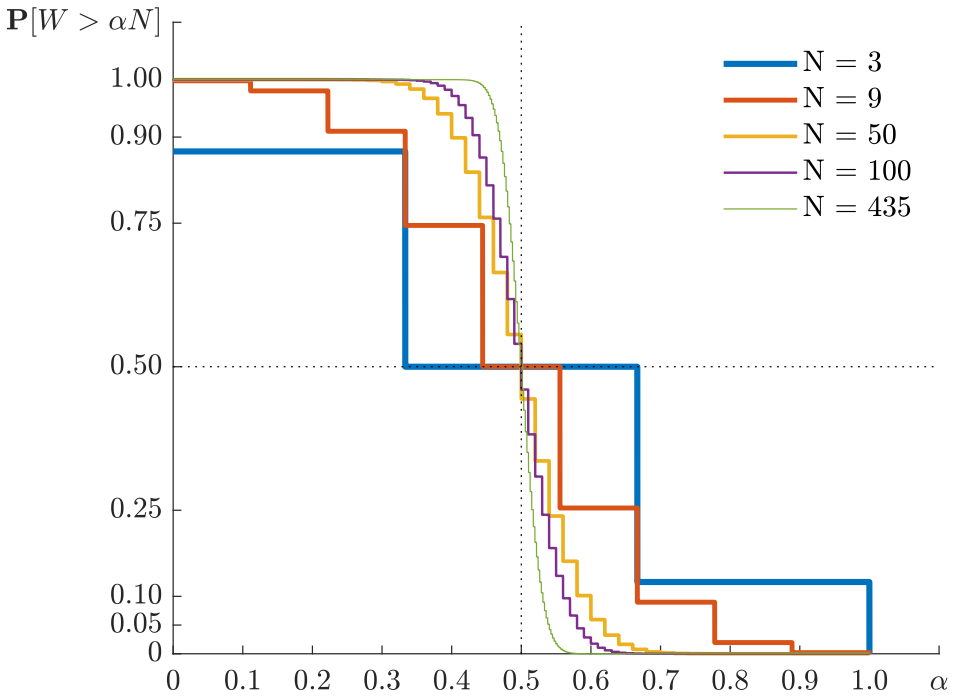


Figure 7: Probability for a transgressing P to be convicted in the limit over time with constant deterrence, i.e., $\mathbf{P}\left[W > \alpha N\right]$, across different sizes of Congress, N

Numerical illustration. Figure 7 plots the probability for a transgressing P to be convicted in the limit over time with constant deterrence, i.e., $\mathbf{P}[W > \alpha N]$, as a function of the voting rule, α, across different sizes of the Congress. We first observe that, across all sizes of the Congress, N, in the limit over time, the probability of conviction under a simple majority rule, i.e., $\alpha = 0.5$, is always 0.5. This is because, by $W \sim \mathcal{B}(N, 0.5)$, we always have $\mathbf{P}[W > 0.5N] = 0.5$. This is to say that, under a simple majority rule, the threat to punish is as credible as a 50/50 coin toss.

Second, more importantly, consistent with Proposition 4, for any voting rule that is more demanding than the simple majority rule, i.e., any super-majority rule, in the limit over time with constant deterrence, the probability of conviction is lower than 0.5, and decreases toward zero as the size of the Congress increases. For example, for a Congress of only three members or independent voting blocks, a 2/3-majority rule, i.e., $\alpha = 0.67$, will already give a probability of conviction that is only slightly higher than 0.1. For a Congress of nine members, i.e., the size of the US Supreme Court, that probability will be lower than 0.1. For a Congress of 50 members, i.e., the number of the US states, the same probability is already lower than 0.05. When the size of the Congress is equal to the size of the US Senate, i.e., 100, or that of the US House of Representatives, 435, the same probability will be almost zero.

Also consistent with Proposition 4, the opposite is true for any minority rule. Not to labor the reader with the exact numbers, for a Congress of 50, 100, or 435 members, in the limit over time with constant deterrence, a 1/3-minority rule, i.e., $\alpha = 0.33$, will almost guarantee the conviction of any transgressing P.

This illustration thus highlights the implication from Proposition 4: if we recognize the difficulty that constant deterrence of high crimes will bring a lack of new empirical knowledge about the world, and thus lead to uncredible threat to convict such transgressions, then, when designing the voting rule in impeachment, making it less demanding for conviction will help to solve the difficulty, while making it more demanding will make the difficulty even worse. The epistemic trap in constant deterrence of high crimes is still the key, and institutional designers may want to recognize the weakness of institutional arrangements that suffer from it, and pay more attention to arrangements that make use of it.

Appendix

A Proof of Results

A.1 Proof of Proposition 2

Proof. To prove Claim 1, suppose $w_s \leq 0$. By Proposition 1, C will impeach P at period s. By $w_s \leq 0$, $a > c$, and Lemmas 1 and 2, we have

$$\mathbf{P}[w_{s+1} > 0 \mid w_s, s] \leq 0.5 < \frac{a}{a+c}. \tag{A.1}$$

By Proposition 1, this implies P will transgress at period $s+1$. Claim 1 is proved.

To prove Claim 2, suppose $w_s > 0$. First, at period s, by $w_s > 0$ and Proposition 1, C will impeach P. Second, at period $s+1$, by Proposition 1 and the evolution of the state of the world, P will transgress at period if and only if

$$\frac{a}{a+c} > \mathbf{P}[w_{s+1} > 0 \mid w_s, s] = \begin{cases} \Phi\left(\frac{\rho w_s}{\sigma}\right) & \text{if } \rho \in (0,1), \\ \Phi\left(\frac{w_s}{\sigma}\right) & \text{if } \rho = 1, \end{cases} \tag{A.2}$$

i.e., if and only if

$$w_s < \frac{\sigma}{\rho} \cdot \Phi^{-1}\left(\frac{a}{a+c}\right) \equiv \bar{w}. \tag{A.3}$$

Note $\bar{w} > 0$ because of $a > c$. Claim 2 is thus proved.

To prove Claim 3, suppose $w_s \geq \bar{w}$. By $\bar{w} > 0$, we have $w_s > 0$; by that and Proposition 1, C will impeach P at period s. By the argument above, P will not transgress at period $s+1$. Now consider periods onward. By $w_s > 0$ and Lemma 1, as time goes by with constant deterrence, the conditional probability of punishment decreases and converges to 0.5; by Lemma 2, $0.5 < a/(a+c)$. Therefore, there must exist a period T, when $\mathbf{P}[w_t > 0 \mid w_s, s]$ is so close to 0.5 for the first time that $\mathbf{P}[w_t > 0 \mid w_s, s] < a/(a+c)$ holds. Indeed, in the case of $\rho = 1$, we have

$$T = s + \left\lfloor \left(\frac{w_s}{\sigma \cdot \Phi^{-1}\left(\frac{a}{a+c}\right)} \right)^2 \right\rfloor + 1, \tag{A.4}$$

where $\lfloor \cdot \rfloor$ is the floor function, and, by $w_s \geq \bar{w}$, we have

$$\left(\frac{w_s}{\sigma \cdot \Phi^{-1}\left(\frac{a}{a+c}\right)} \right)^2 \geq 1 \tag{A.5}$$

and, thus, $T \geq s+2$; in the case of $\rho \in (0,1)$, we have

$$T = s + \left\lfloor \frac{\ln \left(\frac{\left(\frac{\sigma \cdot \Phi^{-1}\left(\frac{a}{a+c}\right)}{w_s} \right) \cdot \left(\sqrt{\left(\frac{\sigma \cdot \Phi^{-1}\left(\frac{a}{a+c}\right)}{w_s} \right)^2 + 4(1-\rho)} - \left(\frac{\sigma \cdot \Phi^{-1}\left(\frac{a}{a+c}\right)}{w_s} \right) \right)}{2(1-\rho)} \right)}{\ln \rho} \right\rfloor + 1, \tag{A.6}$$

and, by $w_s \geq \bar{w}$, we have

$$\frac{\sigma \cdot \Phi^{-1}\left(\frac{a}{a+c}\right)}{w_s} \leq \rho, \tag{A.7}$$

and, thus,

$$\frac{\ln \left(\frac{\left(\frac{\sigma \cdot \Phi^{-1}\left(\frac{a}{a+c}\right)}{w_s} \right) \cdot \left(\sqrt{\left(\frac{\sigma \cdot \Phi^{-1}\left(\frac{a}{a+c}\right)}{w_s} \right)^2 + 4(1-\rho)} - \left(\frac{\sigma \cdot \Phi^{-1}\left(\frac{a}{a+c}\right)}{w_s} \right) \right)}{2(1-\rho)} \right)}{\ln \rho} \geq 1, \tag{A.8}$$

which implies $T \geq s+2$, too. P will thus not transgress until period $T-1$ but will transgress at period T. Claim 3 is thus proved, and so is the proposition. □

A.2 Proof of Proposition 3

Proof. About Claim 3, by $w_s > 0$ and Proposition 1, C will impeach P at period s. By Lemma 1, the conditional probability of punishment at period $s+1$ is greater than 0.5; by $a \leq c$ and Lemma 2, even a 50/50 coin toss is able to deter transgression. Therefore, P will not transgress at period $s+1$. Since no transgression has happened since period $s+1$, period s was still the last time when the state of the world was revealed. The same argument applied to period $s+1$ thus applies from period $s+2$ onward, and P

will thus not transgress from period $s+2$ onward. Claim 3 is thus proved.

About Claim 2, by $w_s \leq 0$ and Proposition 1, C will not impeach P at period s. About period $s+1$, by Proposition 1, as in the proof of Proposition 2, P will transgress if and only if

$$w_s < \frac{\sigma}{\rho} \cdot \Phi^{-1}\left(\frac{a}{a+c}\right) \equiv \bar{w}, \tag{A.9}$$

though, by $a \leq c$, $\bar{w} \leq 0$. Therefore, if $\bar{w} \leq w_s \leq 0$, then the threat to punish is sufficiently credible to deter P from transgressing. About period $s+2$, by the fact that P will not transgress at period $s+1$, period s is still the last time when the state of the world was revealed. By $w_s \leq 0$ and Lemma 1, that implies that the threat to punish at period $s+2$ is even more credible than that at period $s+1$, so it is sufficiently credible to deter P from transgressing at period $s+2$. The same argument applies from period $s+2$ onward. Claim 2 is thus proved.

About Claim 1, by $w_s < \bar{w}$ and the argument above, we have $w_s < 0$. By that and Proposition 1, C will not impeach P at period s. By the argument above, also, P will transgress at period $s+1$. Claim 1 is thus proved, and so is the proposition. □

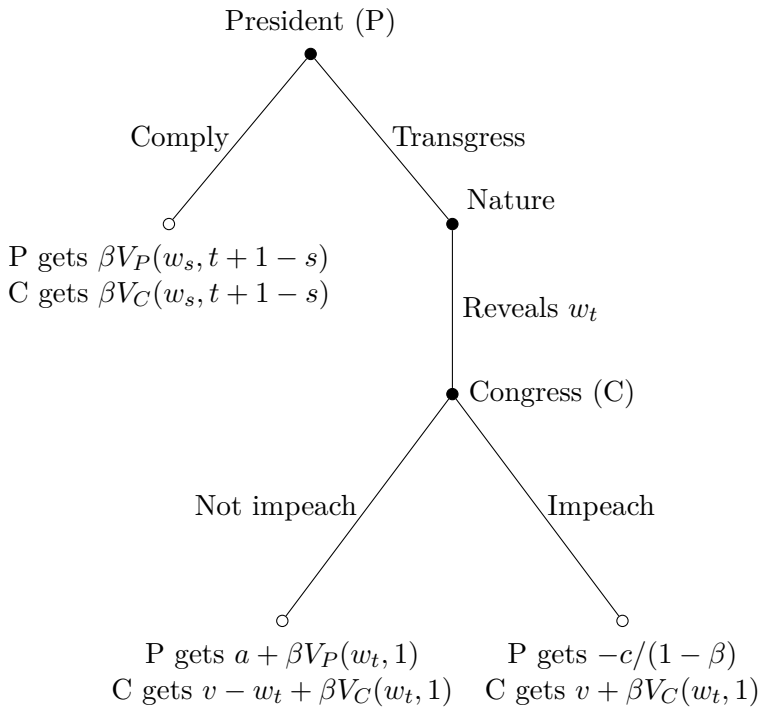
B Results with Forward-looking Players

In the main text we have assumed that the players are myopic. In this section, we show that all main results are robust with respect to forward-looking players. We now assume that, once a P is impeached, a new P will replace him immediately, and $-c$ is the payoff per period for the exiting P; all the players are forward-looking with an infinite horizon, and the discount factor is $\beta \in (0,1)$.

We use the solution concept of symmetric, pure-strategy Markov-perfect equilibria (SMPE), where "symmetric" means that all Ps are restricted to adopting the same strategy. The payoff-relevant variables, i.e., state variables, are thus only the last revelation of the state of the world, i.e., at the last regression, w_s, and how long time it has been since the last transgression, $t-s$. We thus write the net present values that the players enjoy in equilibrium at the beginning of period t as $V_i(w_s, t-s)$, $i = C, P$.

Figure B.1 then illustrates the structure of the subgame starting from t, taking the continuation value in equilibrium for the future period as given. At the beginning of period t, the payoff-relevant variables, w_s and $t-s$, are known to all the players.

As in Section 2, for simplicity, we still assume that C will not impeach P if indifferent between impeaching and not impeaching, and P will not transgress if indifferent between



Players know time, $s < t$, and state of the world, w_s, at last transgression.

Figure B.1: Forward-looking players, subgame at period t

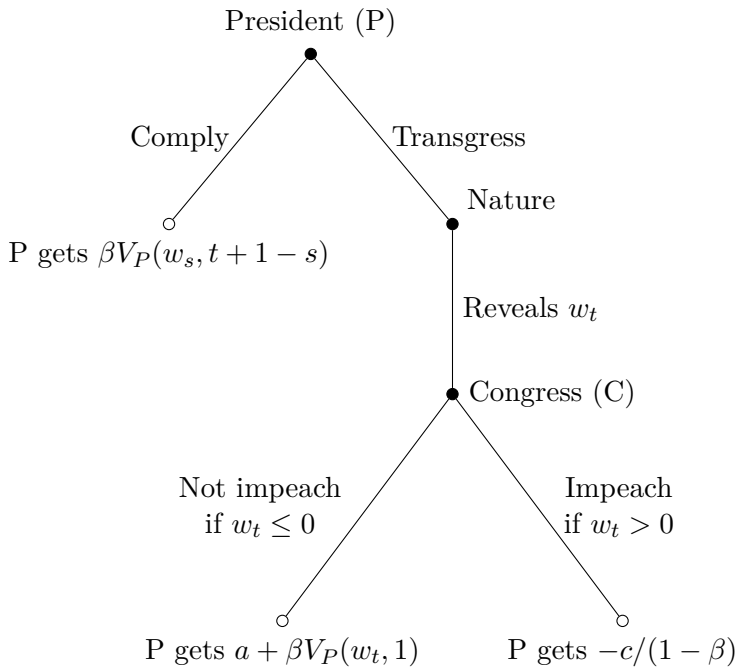
transgressing and complying.

We first establish the existence and uniqueness of the equilibrium. By Figure B.1, it is easy to start by pinning down C's strategy in any equilibrium:

Lemma B.1 (Equilibrium strategy of forward-looking C)**.** *In any SMPE, C's strategy is to impeach any transgressing P at period t if and only if $w_t > 0$.*

Lemma B.1 helps us to reduce the subgame at period t, as in Figure B.1, into a decision problem for P, as in Figure B.2. P's strategy in any SMPE is thus to transgress at period t if and only if

$$\begin{aligned}&\mathbf{E}\left[a + \beta V_P(w_t, 1) \mid w_t \leq 0, w_s, s\right] \cdot \mathbf{P}[w_t \leq 0 \mid w_s, s] - \frac{c}{1-\beta} \cdot \mathbf{P}[w_t > 0 \mid w_s, s] \\ &> \beta \cdot V_P(w_s, t+1-s), \end{aligned} \tag{B.1}$$



Players know time, $s < t$, and state of the world, w_s, at last transgression.

Figure B.2: Forward-looking players, subgame at period t reduced

i.e.,

$$\begin{aligned} & a \cdot \mathbf{P}[w_t \leq 0 \mid w_s, s] - \frac{c}{1-\beta} \cdot \mathbf{P}[w_t > 0 \mid w_s, s] \\ & + \beta \cdot \mathbf{E}\left[V_P(w_t, 1) \mid w_t \leq 0, w_s, s\right] \cdot \mathbf{P}[w_t \leq 0 \mid w_s, s] \\ & > \beta \cdot V_P(w_s, t+1-s), \end{aligned} \tag{B.2}$$

The Bellman equation for P at the beginning of period t is thus

$$\begin{aligned} V_P(w_s, t-s) = \max \big\{ & a \cdot \mathbf{P}[w_t \leq 0 \mid w_s, s] - \frac{c}{1-\beta} \cdot \mathbf{P}[w_t > 0 \mid w_s, s] \\ & + \beta \cdot \mathbf{E}\left[V_P(w_t, 1) \mid w_t \leq 0, w_s, s\right] \cdot \mathbf{P}[w_t \leq 0 \mid w_s, s], \\ & \beta \cdot V_P(w_s, t+1-s) \big\}. \end{aligned} \tag{B.3}$$

We can then have the following result:

Proposition B.1 (Equilibrium with forward-looking players)**.** *There exists a unique SMPE, in which C will impeach any transgressing P at period t if and only if* $w_t > 0$*,*

and P's strategy solves the Bellman equation,

$$V_P(w_s, t-s) = \max\big\{a \cdot \mathbf{P}[w_t \leq 0 \mid w_s, s] - \frac{c}{1-\beta} \cdot \mathbf{P}[w_t > 0 \mid w_s, s] \\ + \beta \cdot \mathbf{E}\left[V_P(w_t, 1) \mid w_t \leq 0, w_s, s\right] \cdot \mathbf{P}[w_t \leq 0 \mid w_s, s], \\ \beta \cdot V_P(w_s, t+1-s)\big\}. \tag{B.4}$$

Proof. First, by Lemma B.1, we have C's strategy in any SMPE, if exists. Second, observe that any $V_P(\cdot, \cdot)$ in equilibrium, if exists, must be bounded by $[0, a/(1-\beta)]$. This is because, an always-complying strategy will generate a zero net present value for P, while the best possible scenario for P would not be better than an always-transgressing strategy without any punishment, which would generate a net present value of $a/(1-\beta)$ for P.

Now denote the set of functions from $\mathbb{R} \times \mathbb{N}^+$ to $\mathbb{R}$ that are bounded by $[0, a/(1-\beta)]$ as

$$S \equiv \left\{\nu : \mathbb{R} \times \mathbb{N}^+ \to \mathbb{R}, \nu(\cdot, \cdot) \in \left[0, \frac{a}{1-\beta}\right]\right\}, \tag{B.5}$$

and endow it with the sup-norm metrics,

$$\rho(\nu, \mu) = \sup_{(x,y) \in \mathbb{R} \times \mathbb{N}^+} |\nu(x, y) - \mu(x, y)|. \tag{B.6}$$

Also denote the operator from S to S that is based on the Bellman equation as

$$G(\nu)(w_s, t-s) \equiv \max\big\{a \cdot \mathbf{P}[w_t \leq 0 \mid w_s, s] - \frac{c}{1-\beta} \cdot \mathbf{P}[w_t > 0 \mid w_s, s] \\ + \beta \cdot \mathbf{E}\left[\nu(w_t, 1) \mid w_t \leq 0, w_s, s\right] \cdot \mathbf{P}[w_t \leq 0 \mid w_s, s], \\ \beta \cdot \nu(w_s, t+1-s)\big\}. \tag{B.7}$$

We want to show that G satisfies Blackwell (1965)'s sufficient conditions for a contraction of modules $\beta \in (0, 1)$. First, suppose that $\nu \in S$ and $\mu \in S$ satisfy that, for any $(w_s, t-s) \in \mathbb{R} \times \mathbb{N}^+$, $\nu(w_s, t-s) \leq \mu(w_s, t-s)$. We then have, by

$\nu(w_s, t-s) \leq \mu(w_s, t-s)$,

$$\begin{aligned}
G(\nu)(w_s, t-s) &= \max\big\{a \cdot \mathbf{P}[w_t \leq 0 \mid w_s, s] - \frac{c}{1-\beta} \cdot \mathbf{P}[w_t > 0 \mid w_s, s] \\
&\qquad + \beta \cdot \mathbf{E}\left[\nu(w_t, 1) \mid w_t \leq 0, w_s, s\right] \cdot \mathbf{P}[w_t \leq 0 \mid w_s, s], \\
&\qquad \beta \cdot \nu(w_s, t+1-s)\big\} \\
&\leq \max\big\{a \cdot \mathbf{P}[w_t \leq 0 \mid w_s, s] - \frac{c}{1-\beta} \cdot \mathbf{P}[w_t > 0 \mid w_s, s] \\
&\qquad + \beta \cdot \mathbf{E}\left[\mu(w_t, 1) \mid w_t \leq 0, w_s, s\right] \cdot \mathbf{P}[w_t \leq 0 \mid w_s, s], \\
&\qquad \beta \cdot \mu(w_s, t+1-s)\big\} \\
&= G(\mu)(w_s, t-s). \qquad \text{(B.8)}
\end{aligned}$$

The monotonicity condition in Blackwell (1965) thus holds.

Second, for any $\Delta \geq 0$ and $(w_s, t-s) \in \mathbb{R} \times \mathbb{N}^+$, we have, by $\mathbf{P}[w_t \leq 0 \mid w_s, s] \in [0, 1]$,

$$\begin{aligned}
G(\nu+\Delta)(w_s, t-s) &\equiv \max\big\{a \cdot \mathbf{P}[w_t \leq 0 \mid w_s, s] - \frac{c}{1-\beta} \cdot \mathbf{P}[w_t > 0 \mid w_s, s] \\
&\qquad + \beta \cdot \mathbf{E}\left[\nu(w_t, 1) + \Delta \mid w_t \leq 0, w_s, s\right] \cdot \mathbf{P}[w_t \leq 0 \mid w_s, s], \\
&\qquad \beta \cdot (\nu(w_s, t+1-s) + \Delta)\big\} \\
&= \max\big\{a \cdot \mathbf{P}[w_t \leq 0 \mid w_s, s] - \frac{c}{1-\beta} \cdot \mathbf{P}[w_t > 0 \mid w_s, s] \\
&\qquad + \beta \cdot \mathbf{E}\left[\nu(w_t, 1) \mid w_t \leq 0, w_s, s\right] \cdot \mathbf{P}[w_t \leq 0 \mid w_s, s] \\
&\qquad + \beta\Delta \cdot \mathbf{P}[w_t \leq 0 \mid w_s, s], \\
&\qquad \beta\nu(w_s, t+1-s) + \beta\Delta\big\} \\
&\leq \max\big\{a \cdot \mathbf{P}[w_t \leq 0 \mid w_s, s] - \frac{c}{1-\beta} \cdot \mathbf{P}[w_t > 0 \mid w_s, s] \\
&\qquad + \beta \cdot \mathbf{E}\left[\nu(w_t, 1) \mid w_t \leq 0, w_s, s\right] \cdot \mathbf{P}[w_t \leq 0 \mid w_s, s] \\
&\qquad + \beta\Delta, \beta\nu(w_s, t+1-s) + \beta\Delta\big\} \\
&= \max\big\{a \cdot \mathbf{P}[w_t \leq 0 \mid w_s, s] - \frac{c}{1-\beta} \cdot \mathbf{P}[w_t > 0 \mid w_s, s] \\
&\qquad + \beta \cdot \mathbf{E}\left[\nu(w_t, 1) \mid w_t \leq 0, w_s, s\right] \cdot \mathbf{P}[w_t \leq 0 \mid w_s, s], \\
&\qquad \beta\nu(w_s, t+1-s)\big\} + \beta\Delta \\
&= G(\nu)(w_s, t-s) + \beta\Delta, \qquad \text{(B.9)}
\end{aligned}$$

where $\beta \in (0, 1)$. The discounting condition in Blackwell (1965) thus holds, too.

Therefore, G satisfies Blackwell (1965)'s sufficient conditions for a contraction of

modules β. By Banach (1922)'s contraction mapping theorem, G thus admits a unique fixed point, which is the value function $V_P(\cdot,\cdot)$ in the unique SMPE. The proposition is thus proved. □

With the existence and uniqueness of the SMPE established, we first show a result parallel to Proposition 2:

Proposition B.2 (The difficulty in deterring high crimes, forward-looking players)**.** *Given any $a > c/(1-\beta)$, there does not exist a finite level of the state of the world at a transgression at period s, $w_s < \infty$, such that, in the unique SMPE, P will always comply from period $s+1$ onward.*

Proof. Suppose that there exist such $w_s < \infty$ that, in the unique SMPE, P will comply in any period $t \geq s+1$. Since P will comply in any period $t \geq s+1$, then we have, for any $t \geq s+1$,

$$\begin{aligned} & a \cdot \mathbf{P}[w_t \leq 0 \mid w_s, s] - \frac{c}{1-\beta} \cdot \mathbf{P}[w_t > 0 \mid w_s, s] \\ & + \beta \cdot \mathbf{E}\left[V_P(w_t, 1) \mid w_t \leq 0, w_s, s\right] \cdot \mathbf{P}[w_t \leq 0 \mid w_s, s] \\ & \leq \beta \cdot V_P(w_s, t+1-s), \end{aligned} \tag{B.10}$$

where

$$V_P(w_s, t+1-s) = 0. \tag{B.11}$$

This condition thus becomes

$$\begin{aligned} & a \cdot \mathbf{P}[w_t \leq 0 \mid w_s, s] - \frac{c}{1-\beta} \cdot \mathbf{P}[w_t > 0 \mid w_s, s] \\ & + \beta \cdot \mathbf{E}\left[V_P(w_t, 1) \mid w_t \leq 0, w_s, s\right] \cdot \mathbf{P}[w_t \leq 0 \mid w_s, s] \leq 0, \end{aligned} \tag{B.12}$$

i.e.,

$$\begin{aligned} & a - \left(a + \frac{c}{1-\beta}\right) \cdot \mathbf{P}[w_t > 0 \mid w_s, s] \\ & + \beta \cdot \mathbf{E}\left[V_P(w_t, 1) \mid w_t \leq 0, w_s, s\right] \cdot \mathbf{P}[w_t \leq 0 \mid w_s, s] \leq 0. \end{aligned} \tag{B.13}$$

Note that, by Lemma 1, $\mathbf{P}[w_t > 0 \mid w_s, s] \to 0.5$ as $t - s \to \infty$; we also know $V_P(\cdot,\cdot) \in [0, a/(1-\beta)]$. Therefore, by $a > c/(1-\beta)$, as $t - s \to \infty$, the left-hand side of Inequation (B.13) is strictly positive, while the right-hand side is zero. This

inequation thus cannot hold as $t - s \to \infty$, contradicting to what we have supposed. The proposition is thus proved by contradiction. □

The proof follows the idea in the proof of Proposition 2: as time goes by with constant deterrence, the threat to punish will become increasing similar to a 50/50 coin toss, which, given a sufficient great gain from transgression, would be too uncredible to deter P from transgressing.

We can also show a result parallel to Proposition 3:

Proposition B.3 (Forever deterrence possible under severe punishment, forward looking players)**.** *Given any $a < c$, in the unique SMPE, if transgression happens at period s with $w_s > 0$, then P will always comply from period $s + 1$ onward.*

Proof. At any period $t \geq s + 1$, given the continuation strategy in consideration, we have

$$V_P(w_s, t + 1 - s) = 0. \tag{B.14}$$

and, by Lemma 1,

$$\mathbf{P}[w_t > 0 \mid w_s, s] > 0.5. \tag{B.15}$$

Therefore, further by $a < c$ and $V_P(\cdot, \cdot) \in [0, a/(1 - \beta)]$, we have

$$\begin{aligned}
&a \cdot \mathbf{P}[w_t \leq 0 \mid w_s, s] - \frac{c}{1-\beta} \cdot \mathbf{P}[w_t > 0 \mid w_s, s] \\
&+ \beta \cdot \mathbf{E}\left[V_P(w_t, 1) \mid w_t \leq 0, w_s, s\right] \cdot \mathbf{P}[w_t \leq 0 \mid w_s, s] \\
&\leq a \cdot \mathbf{P}[w_t \leq 0 \mid w_s, s] - \frac{c}{1-\beta} \cdot \mathbf{P}[w_t > 0 \mid w_s, s] \\
&+ \beta \cdot \frac{a}{1-\beta} \cdot \mathbf{P}[w_t \leq 0 \mid w_s, s] \\
&= \frac{a}{1-\beta} \cdot \mathbf{P}[w_t \leq 0 \mid w_s, s] - \frac{c}{1-\beta} \cdot \mathbf{P}[w_t > 0 \mid w_s, s] \\
&= \frac{1}{1-\beta} \cdot (a - (a + c) \cdot \mathbf{P}[w_t > 0 \mid w_s, s]) \\
&< \frac{1}{1-\beta} \cdot (a - 0.5\,(a + c)) \\
&< 0 = \beta \cdot V_P(w_s, t + 1 - s).
\end{aligned} \tag{B.16}$$

P thus has no incentive to have a single deviation from complying to transgressing. The strategy in consideration is thus Markov perfect, and the proposition is proved. □

The proof follows the idea in the proof of Proposition 3: once a transgressing P is punished, the threat to punish will be at least as credible as a 50/50 coin toss, which, given a sufficient severe punishment for transgression, would be sufficiently credible to deter P from transgressing.

References

Acemoğlu, Kamer Daron, Nicolás Ajzenman, Cevat Giray Aksoy, Martin Fiszbein, and Carlos Molina. Forthcoming. (Successful) democracies breed their own support. *Review of Economic Studies* .

Acemoğlu, Kamer Daron, Georgy Egorov, and Konstantin Sonin. 2013. A political theory of populism. *Quarterly Journal of Economics* **128**: 771–805.

Acemoğlu, Kamer Daron, and James A. Robinson. 2022. Non-modernization: Power–culture trajectories and the dynamics of political institutions. *Annual Review of Political Science* **25**: 323–339.

Banach, Stefan. 1922. Sur les opérations dans les ensembles abstraits et leur application aux équations intégrales. *Fundamenta Mathematicae* **1**: 133–181.

Baumgartner, Jody C., and Naoko Kada (eds.). 2003. *Checking Executive Power: Presidential Impeachment in Comparative Perspective*. Westport: Praeger Publishers.

Becker, Gary S. 1968. Crime and punishment: An economic approach. *Journal of Political Economy* **76**: 169–217.

Besley, Timothy, and Torsten Persson. 2019. Democratic values and institutions. *American Economic Review: Insights* **1**: 59–76.

Bidner, Chris, and Patrick Francois. 2013. The emergence of political accountability. *Quarterly Journal of Economics* **128**: 1397–1448.

Black, Charles L., Jr., and Philip Bobbitt. 2018. *Impeachment: A Handbook, New Edition*. New Haven: Yale University Press.

Blackwell, David Harold. 1965. Discounted dynamic programming. *Annals of Mathematical Statistics* **36**: 226–235.

Bowman, Frank O., III. 2024. *High Crimes and Misdemeanors: A History of Impeachment for the Age of Trump*. Cambridge: Cambridge University Press, 2nd edition.

Callander, Steven, and Gregory J. Martin. 2017. Dynamic policymaking with decay. *American Journal of Political Science* **61**: 50–67.

Chalfin, Aaron, and Justin McCrary. 2017. Criminal deterrence: A review of the literature. *Journal of Economic Literature* **55**: 5–48.

Chiopris, Caterina, Monika Nalepa, and Georg Vanberg. 2021. A wolf in sheep's clothing: Citizen uncertainty and democratic backsliding. Working paper, Harvard University.

Congressional Research Service. 2023. *The Constitution of the United States of America: Analysis and Interpretation: Analysis of Cases Decided by the Supreme Court of the United States to June 30, 2022*. Washington: United States Government Publishing Office.

Conte, Marc N., and David L. Kelly. 2021. Understanding the improbable: A survey of fat tails in environmental economics. *Annual Review of Resource Economics* **13**: 289–310.

Crespo, Andrew Manuel. 2019. Impeachment as punishment. *Harvard Law and Policy Review* **13**: 579–592.

Dziuda, Wioletta, and Antoine Loeper. 2016. Dynamic collective choice with endogenous status quo. *Journal of Political Economy* **124**: 1148–1186.

Fearon, James D. 2011. Self-enforcing democracy. *Quarterly Journal of Economics* **126**: 1661–1708.

Gehlbach, Scott. 2013. *Formal Models of Domestic Politics*. New York: Cambridge University Press.

Ginsburg, Tom, Aziz Z. Huq, and David Landau. 2021. The comparative constitutional law of presidential impeachment. *University of Chicago Law Review* **88**: 81–164.

Gratton, Gabriele, and Barton E. Lee. 2024. Liberty, security, and accountability: The rise and fall of illiberal democracies. *Review of Economic Studies* **91**: 340–371.

Grillo, Edoardo, Zhaotian Luo, Monika Nalepa, and Carlo Prato. 2024. Theories of democratic backsliding. *Annual Review of Political Science* **27**: 381–400.

Grillo, Edoardo, and Carlo Prato. 2023. Reference points and democratic backsliding. *American Journal of Political Science* **67**: 71–88.

Howell, William G., and Stephane Wolton. 2018. The politician's province. *Quarterly Journal of Political Science* **13**: 119–146.

Huang, Zhangkai, Lixing Li, Guangrong Ma, and Lixin Colin Xu. 2017. Hayek, local information, and commanding heights: Decentralizing state-owned enterprises in China. *American Economic Review* **107**: 2455–2478.

Huq, Aziz Z. 2018. Legal or political checks on apex criminality: An essay on constitutional design. *UCLA Law Review* **65**: 1506–1530.

Landes, William M., and Richard A. Posner. 1975. The independent judiciary in an interest-group perspective. *Journal of Law and Economics* **18**: 875–901.

Li, Weijia, Gérard Roland, and Yang Xie. 2022. Erosion of state power, corruption control and fiscal capacity. *Economic Journal* **132**: 1542–1565.

Luo, Zhaotian, and Adam Przeworski. 2023. Democracy and its vulnerabilities: Dynamics of democratic backsliding. *Quarterly Journal of Political Science* **18**: 105–130.

Madison, James, Jr. 1845. *Debates on the Adoption of the Federal Constitution, with a Diary of the Debates of the Congress of the Confederation, Volume V*. Washington: United States Congress. Revised and arranged by Jonathan Elliot.

Matějka, Filip, and Guido Tabellini. 2021. Electoral competition with rationally inattentive voters. *Journal of the European Economic Association* **19**: 1899–1935.

Miller, Michael K. 2021. A republic, if you can keep it: Breakdown and erosion in modern democracies. *Journal of Politics* **83**: 198–213.

Monaghan, Chris, Matthew Flinders, and Aziz Z. Huq (eds.). 2024. *Impeachment in a Global Context: Law, Politics, and Comparative Practice*. New York: Routledge.

Myerson, Roger B. 2008. The autocrat's credibility problem and foundations of the constitutional state. *American Political Science Review* **102**: 125–139.

Myerson, Roger B. 2019. The dilemma of presidential impeachment. Perspectives on Economics and Civilization, accessible at `https://econciv.wordpress.com/2019/12/15/the-dilemma-of-presidential-impeachment/` on May 15, 2024.

Ohnesorge, John. 2020. Comparing impeachment regimes. *Duke Journal of Comparative and International Law* **31**: 259–299.

Pistor, Katharina, and Chenggang Xu. 2003. Incomplete law. *New York University Journal of International Law and Politics* **35**: 931–1013.

Powell, Robert. 1990. *Nuclear Deterrence Theory: The Search for Credibility*. Cambridge: Cambridge University Press.

Przeworski, Adam. 2006. Self-enforcing democracy. In Wittman, Donald A., and Barry R. Weingast (eds.) *The Oxford Handbook of Political Economy*. New York: Oxford University Press, 312–328.

Schelling, Thomas C. 1960. *The Strategy of Conflict*. Cambridge: Harvard University Press.

Stern, Nicholas. 2013. The structure of economic modeling of the potential impacts of climate change: Grafting gross underestimation of risk onto already narrow science models. *Journal of Economic Literature* **51**: 838–859.

Story, Joseph. 1833. *Commentaries on the Constitution of the United States*. Cambridge: Hilliard, Gray, and Company.

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

Svolik, Milan W. 2020. When polarization trumps civic virtue: Partisan conflict and the subversion of democracy by incumbents. *Quarterly Journal of Political Science* **15**: 3–31.

Tirole, Jean. 1988. *The Theory of Industrial Organization*. Cambridge: MIT press.

Tribe, Laurence H. 1998. Defining "high Crimes and Misdemeanors": Basic principles. *George Washington Law Review* **67**: 712–734.

Tribe, Laurence H., and Joshua Matz. 2018. *To End a Presidency: The Power of Impeachment*. New York: Basic Books.

von Hayek, Friedrich August. 1945. The use of knowledge in society. *American Economic Review* **35**: 519–530.

Weingast, Barry R. 1997. The political foundations of democracy and the rule of the law. *American Political Science Review* **91**: 245–263.

Weitzman, Martin L. 1974. Prices vs. quantities. *Review of Economic Studies* **41**: 477–491.

Weitzman, Martin L. 2009. On modeling and interpreting the economics of catastrophic climate change. *Review of Economics and Statistics* **91**: 1–19.

Weitzman, Martin L. 2011. Fat-tailed uncertainty in the economics of catastrophic climate change. *Review of Environmental Economics and Policy* **5**: 275–292.