Political Kludges[†]

By Keiichi Kawai, Ruitian Lang, and Hongyi Li*

This paper explores the origins of policy complexity. It studies a model where policy is difficult to undo because policy elements are entangled with each other. Policy complexity may accumulate as successive policymakers layer new rules upon existing policy. Complexity emerges and persists in balanced democratic polities, when policymakers are ideologically extreme, and when legislative frictions impede policymaking. Complexity begets complexity: simple policies remain simple, whereas complex policies grow more complex. Patience is not always a virtue: farsighted policymakers may engage in obstructionism, deliberately introducing complex policies to hinder future opponents. (*JEL* C73, D72, D73, D78)

The complexity of public policy imposes significant costs on society. The United States Internal Revenue Service estimated that the various costs of tax compliance exceeded $168 billion in 2010, which was 15 percent of total tax receipts for that year.[1] In many areas of policy ranging from the tax code to education to healthcare, such complexity is pervasive and persistent.

This paper studies the evolution of policy complexity. It develops a theory where policy complexity emerges in the course of political conflict. Successive policymakers modify policy in pursuit of their own policy goals; they do so by layering new rules upon existing policy. As layers of rules accumulate, so does policy complexity.

A key aspect of our theory is that new rules take the form of *kludges*: piecemeal modifications that patch over old programs rather than replace them. Kludges serve to remedy flaws in the implementation of existing policy, or even cancel out their impact. As such, they improve on existing policy, but do so in an inelegant and

  [1] This may not be too surprising, given that the US tax code contains more than four million words.

inefficient fashion relative to the alternative—to completely rewrite existing policy, unburdened by legacy concerns.[2]

We model a setting where policy control shifts intermittently between two rival policymakers. While in control, each policymaker may add or delete policy rules to achieve his ideological goal. By *complexity* we mean the measure of rules that make up policy (e.g., the number of lines in the tax code). In this setting, kludges are rules that are added to cancel out the ideological impact of old rules without having to delete those old rules. In other words, kludges allow policymakers to avoid elaborate policy overhauls, at the cost of excessive policy complexity.

This narrative has, so far, not yet addressed why policymakers may favor adding new rules as kludges, rather than deleting and replacing existing rules. Our theory incorporates features of the legislative process that are conducive to kludges. First, policymaking is incremental: rules may only be added or deleted gradually. Interest groups may oppose adding new rules that they dislike or deleting old rules that they like. Such resistance constrains policymakers with limited political capital from undertaking radical overhauls; instead, they make incremental changes.[3]

Further, rules are entangled with one another. Each rule is designed to fit well with the rest of policy—either by legislative intent, or through subsequent administrative and judicial interpretation of enacted legislation. Rules may rely on features of other rules, or fill gaps in other rules, or build upon other rules to modify their effect. Such interdependencies between rules create entanglements that hinder the undoing of existing policy. Deletion of a rule may cripple other dependent rules that rely on the functionality of the deleted rule. Consider the Alternative Minimum Tax (AMT) in the US Tax Code. Many observers deem the AMT to be an unsatisfactory solution to the problems it was intended to solve, but also believe that it will be difficult to undo or significantly edit the AMT because many other aspects of the federal tax system have come to rely on the AMT.

In this paper, such entanglements are modeled as an exogenous constraint on the ability of policymakers to precisely undo existing policy. Essentially, an existing rule cannot be deleted without also deleting other rules that are (randomly) entangled with the targeted rule.

Consequently, each policymaker faces a trade-off between improving his ideological position and reducing policy complexity. He may add new rules that favor his ideological position. Or, he may delete rules that detract from his position. Deletion has the benefit of reducing complexity. However, the deletion process is stymied by entanglement. Unfavorable rules cannot be deleted surgically; they may be entangled with favorable rules that have to be deleted as well, thus slowing

---

[2] One example of policy kludge is the US Affordable Care Act (ACA) of 2010, which introduced mechanisms (including mandates, subsidies, and insurance exchanges) to fill gaps in the existing patchwork of private and public insurance options. A common view of both proponents and opponents was that the ACA is excessively complex compared to alternatives such as a single-payer healthcare system. These alternatives, however, would have required a politically infeasible complete overhaul of healthcare policy.

[3] For recent discussion, see Levy and Razin (2013) and Teles (2013). Further, besides political constraints, cognitive limitations may introduce uncertainty about the impact of large-scale policy changes and thus force policymakers to focus on making small "local" changes to policy (see, e.g., Lindblom 1959; Bendor 1995; Callander 2011a, b).

progress toward—or even moving policy away from—the policymaker's position. Entanglements thus induce a bias toward adding rather than deleting rules.

In this setting, we analyze the long-run evolution of policy under political conflict. The main dynamic effects are not driven by strategic interactions between policymakers. In fact, for most of this paper, we completely ignore strategic interactions by focusing on myopic policymakers. We demonstrate how, and under what circumstances, complexity may emerge and persist from such myopic dynamics.

Our first finding is that initial conditions matter. We show that when complexity is high, all parties are particularly prone to adding further complexity in the form of kludges. Consequently, complexity begets complexity: simple policies remain simple forever, whereas complex policies may become increasingly complex.

When complexity is high, policy evolution switches intermittently between two phases. In one phase, the policymaker in control adds rules to shift policy toward his favored position, increasing complexity as he does so. In the other phase, the policymaker in control has attained his policy goals, and deletes rules to reduce complexity. So, long-run outcomes are determined by the tug-of-war between these two phases, which exert opposite forces on complexity. If the first force is more powerful than the second, then complexity may accumulate and become unbounded over the long-run—in which case we say that policy becomes *kludged*.

Having identified this tension, we characterize conditions under which policy may become kludged. One set of comparative static results relates to political institutions:

- Policy is likely to become kludged if parties hold control for relatively equal periods, i.e., political power is relatively balanced.
- Policy is likely to become kludged if power transitions between parties occur frequently; for example, if electoral terms are short.
- Policy is likely to become kludged if legislative *friction* is high; that is, if the legislative process is slowed by procedural hurdles such as veto points.

The same institutional features that generate kludged policies also serve to reduce ideological polarization in policy outcomes. Indeed, our results highlight a trade-off in the design of political institutions between policy outcomes that are simple but ideologically polarized, and outcomes that are complex but ideologically moderate.

Thus, viewed through the lens of our theory, the American political system is geared toward generating ideologically moderate political outcomes (relative to the preferences of its political parties)—but with the downside of complex, kludged public policy. After all, American electoral competition has historically been relatively balanced and volatile, with control over the presidency and congress switching regularly between the two major parties over the past decades. Further, many hurdles in the American legislative process, such as supermajority voting requirements, a proliferation of veto points, and filibuster rules, hinder the creation of new laws and the undoing of existing laws. Conversely, a planner who prioritizes complexity reduction over ideological moderation should design political institutions that weaken (or even eliminate) political competition and minimize legislative frictions.

Another set of comparative statics relates to political preferences. We show that policy is likely to become kludged as parties' preferences become more polarized—that is, as the ideological distance between parties' favored positions increases, and also as parties' preferences over positions become more intense. These results suggest the following connection between two trends in American politics: an inexorable increase in policy complexity over recent decades (Teles 2013) may have been driven by increasing polarization in political preferences over the same period.[4]

Later in the paper, we move to a setting with forward-looking parties, so that strategic interactions come into play. Here, the main lesson is that patience is not necessarily a virtue. A forward-looking party may engage in obstructionism: he makes policy changes not to achieve his own policy goals, but rather to hinder his opponent's future policy moves. Specifically, in a conflict between ideologically zealous policymakers, parties exhibit *strategic extremism*. Each policymaker pursues policy positions that are even more ideologically extreme than his preferences would naively dictate. This serves to "shift the goalposts" against his opponent, which ensures that policy remains relatively close to the policymaker's preferred positions in the medium run.[5] Such strategic extremism has long-run consequences for policy complexity. In particular, relative to the myopic case, strategic extremism may increase the probability that policy becomes kludged.

Literature Review.—Most closely related is Ely (2011), who studies how inefficiency—in the form of kludged designs—may arise and persist in single-player adaptive processes. Ely (2011) considers a genetic code—a set of genes—that performs well if each gene aligns appropriately with the environment, as mediated by the alignment of a particular "master gene." A code is kludged if the "master gene" is poorly aligned with the external environment. In an evolutionary setting where the genetic code grows increasingly long while being subject to fitness selection over random mutations, Ely (2011) focuses on showing that kludge may persist indefinitely: even though mutations may be arbitrarily large and thus a mutation that unkludges the code while maintaining internal alignment will eventually occur for any fixed code length, the increase in code length over time ensures that such mutations grow increasingly rare, and in fact may never occur. Ely (2011) shares with our paper the central conceit of kludges: that interdependencies between elements make kludges difficult to undo, and that increases in complexity lengthen the process of undoing kludges. But our approach differs in various ways. We highlight three clear distinctions here. First, Ely (2011) introduces a "mechanical" evolutionary force that increases complexity over time, whereas in our model, players endogenously choose whether to increase complexity. This difference reflects our model's central focus on understanding the origins of complexity, whereas complexity in Ely (2011) is

---

[4] McCarty, Poole, and Rosenthal (2006) find that American political parties have become increasingly extreme since the 1970s; Azzimonti (2018) finds that political disagreement between parties has intensified during the same period.

[5] Glaeser, Ponzetto, and Shapiro (2005) present a voting model where politicians may *declare* extreme positions (relative to the voting public) to pander to their base. In contrast, in our model, politicians may *implement* extreme policies (relative to their own preferences).

principally a device to inhibit unkludging.[6] Second, we consider a two-player game between policymakers with conflicting objectives to highlight the role of political competition in producing kludges; in a one-player version of our model, kludges would never persist. Third, moving beyond the focus in Ely (2011) on a single myopic player, we discuss how conflict and strategic motives may lead to kludges.

Like our paper, Gratton et al. (2017) study the dynamics of policy complexity. In their model, policymakers enact legislation purely to bolster their reputation with the public. Reputational incentives to avoid bad legislation are muted if existing policy is already complex, potentially leading to a "complexity trap" that is superficially reminiscent of our model's path dependence result—albeit via a different mechanism.

A number of other papers from various literatures explore the idea that incremental rule development may be path-dependent. Callander and Hummel (2014) consider a model where successive policymakers with conflicting preferences strategically experiment to find their preferred policy. The first policymaker benefits from a "surprising" experiment outcome because it deters experimentation by the second policymaker, and thus preserves any policy gains by the first policymaker. Ellison and Holden (2013) study a model of endogenous rule development where there are exogenous constraints on the extent to which new rules may "overwrite" old rules. Compared to these models, our paper introduces path dependence through a distinct mechanism—entanglement—and thus produces very different implications.

Our results on strategic extremism are also related to the literature on agenda-setting in politics. Chen and Eraslan (2017) consider a model where competing policymakers take turns to address outstanding policy issues; their key assumption is that an issue that has previously been addressed cannot be revisited by subsequent policymakers. Dziuda and Loeper (2016) and Buisseret and Bernhardt (2017) consider settings where this period's policy outcome is determined by the interaction between competing policymakers, and serves as an endogenous status quo for the next period.[7] Dziuda and Loeper (2016) find that the endogenous status quo assumption may lead to policymakers taking extreme positions in bargaining, leading to policy gridlock; this logic is reminiscent of our strategic extremism results. On the other hand, Buisseret and Bernhardt (2017) find that strategic concerns may restrain the agenda-setter from aggressive policy-setting. These papers do not address policy complexity, which is of course the central focus of the present paper. Further, unlike these other papers, the status-quo effect in the present paper is technological; any changes to policy take time for future policymakers to undo, which drives the dynamics of policy position and complexity.

I. Model

Policy.—The policy is a set of infinitesimal rules. Each rule's ideological *direction* is either positive $(+)$ or negative $(-)$. The policy is summarized as a pair of numbers

---

[6] Relatedly, whereas unkludged codes can be arbitrarily complex in Ely (2011), kludge and complexity are essentially synonymous in our model.

[7] Other papers in this literature include Bernheim, Rangel, and Rayo (2006); Messner and Polborn (2012); and Levy and Razin (2013).

$\mathbf{p} = (p_+, p_-)$, where $p_j \geq 0$ is the mass of rules with direction $j \in \{+, -\}$. The policy's *position* is the difference between the masses of positive and negative rules,

$$p = p_+ - p_-;$$

and the policy's *complexity* is its total mass, denoted as

$$\|\mathbf{p}\| = p_+ + p_-.$$

Policy evolves in continuous time, $t \geq 0$. So we write, for example, $\mathbf{p}(t) = (p_+(t), p_-(t))$; but we will often conveniently suppress the time-dependence of policy variables. We take the initial policy $\mathbf{p}(0)$ as given, i.e., as a primitive of the model.

Players and Preferences.—There are two Parties, $+1$ and -1, generically identified as i. The flow payoff of Party $i \in \{+1, -1\}$ is a function of policy position and complexity:

(1) $$u_i(\mathbf{p}) = -z_i |p - p_i^*| - \|\mathbf{p}\|,$$

where $p_i^* \in \mathbb{R}$ is his positional *ideal*, $|p - p_i^*|$ is the absolute value of $p - p_i^*$, and $z_i > 1$ is his ideological *zeal*. That is, parties dislike policy positions that are distant from their ideals, and dislike complex policies. We assume that $p_{+1}^* > 0$, $p_{-1}^* < 0$; and that $z_{+1} > 1$, $z_{-1} > 1$.[8]

Some descriptive terminology: parties with small (large) $|p_i^*|$ are called *moderates* (*extremists*). Parties with high z_i are *zealous*. A policy with position $p = p_i^*$ is *i-ideal*.

Each Party i maximizes his discounted payoff,

$$\max E\left[\int_0^\infty u_i(\mathbf{p}(t)) e^{-r_i t} dt \right].$$

Most of this paper considers myopic parties: $r_{+1}, r_{-1} \to \infty$. Two features of myopic behavior are convenient. First, strategic interactions vanish: a myopic Party i is unconcerned about what his opponent $-i$ does after i loses control. Second, only the neighborhood of the current policy $\mathbf{p}(t)$ is relevant because only nearby policies can be attained in the near future. In particular, as $r_i \to \infty$, Party i's problem reduces to maximizing the rate of change of his payoff,

(2) $$\max\left\{ \frac{d}{dt} u_i(\mathbf{p}(t)) \right\}.$$

---

[8] The assumption $z_i > 1$ ensures that parties have sufficiently intense preferences over ideological position, and thus face a nontrivial trade-off between adding and deleting rules. Alternatively, one might posit that payoffs are quadratic in position, $u_i(\mathbf{p}) = -z_i(p_i^* - p)^2 - \|\mathbf{p}\|$, so that Party i's positional preferences intensify as policy strays from p_i^*. This alternative formulation is analytically and expositionally less convenient, but produces qualitatively similar results.

Policymaking Technology.—At any time t, one party is in control; label him as $i(t) \in \{+1, -1\}$. Control transitions from i to $-i$ are random and arrive at rate $\lambda_i > 0$. Without loss of generality, Party $+1$ starts the game in control: $i(0) = +1$. We interpret λ_i as i's political *vulnerability*.

At each time t, Party $i(t)$ chooses nonnegative addition rates $\alpha_+(t), \alpha_-(t)$ and deletion rates $\delta_+(t), \delta_-(t)$, which move (p_+, p_-):

$$(3) \qquad \tfrac{d}{dt} p_j(t) = \alpha_j(t) - \delta_j(t) \quad \text{for each } j \in \{+, -\},$$

subject to a *flow constraint*, reflecting the party's limited capacity to make policy changes, where γ^{-1} parametrizes the degree of legislative *friction*:

$$(4) \qquad \alpha_+(t) + \alpha_-(t) + \delta_+(t) + \delta_-(t) \leq \gamma,$$

$$\delta_j(t) = 0 \quad \text{if} \quad p_j(t) = 0 \quad \text{for } j \in \{+, -\},$$

and an *entanglement constraint* on the direction of rule deletion:

$$(5) \qquad \frac{\delta_+(t)}{\delta_-(t)} = \frac{p_+(t)}{p_-(t)}.$$

In words, the entanglement constraint states that deleted rules must have the same proportions, by direction, as existing rules. This specification of the entanglement constraint is quite tight: given (total) deletion rate

$$\delta(t) = \delta_+(t) + \delta_-(t),$$

each of $\delta_+(t)$ and $\delta_-(t)$ are fully determined from (5). Consequently, we may use $\delta(t)$ to summarize the pair of deletion rates $(\delta_+(t), \delta_-(t))$.

Discussion of the Entanglement Constraint.—The entanglement constraint (5) captures, in reduced form, the notion of dependencies between rules. The premise is that parties cannot surgically target specific rules for deletion: a party who targets rule π for deletion has to also delete other rules that are entangled with π.

The specific form of (5) is a tractable depiction of severe entanglement, whereby each rule is entangled with many other rules. If policy is severely entangled, then deletions will mostly be "indirect" (i.e., of rules entangled with targeted rules) rather than "direct" (i.e., of targeted rules). Consequently, parties will have little control over the directions of deleted rules, especially if they have limited knowledge about which rules are entangled with each other. The overall composition of deleted rules will match the composition of the policy as a whole, rather than the direction of those rules targeted for deletion. This notion is captured succinctly by our entanglement constraint (5).

In Appendix A, we argue that our formulation of the entanglement constraint is quite natural. We present two alternative approaches to model the notion of policy entanglements, and show that both models generate our entanglement constraint under assumptions that reflect severe entanglement.

Appendix A, Part 1 considers a linear network. Rules are totally ordered along a line. A party who seeks to delete a rule π first has to delete all the rules above π in the ordering. This model produces (5) as a limiting outcome when the number of rules is large.

Appendix A, Part 2 considers a random network where any two rules are connected with some small probability. Dependencies are captured by the network structure: when a policymaker targets a rule π for deletion, he also has to delete all of π's neighbors. This model produces (5) at the limit where each rule has infinitely many neighbors. Away from this limit, so that entanglement is not severe, a looser version of the entanglement constraint is obtained; we show that our results continue to hold there as well.

Policy Simplicity and Efficiency.—The following terminology will be helpful. Let the policy's *positive-simplicity* be the ratio of position to complexity, $p/\|\mathbf{p}\| \in [-1, 1]$. (Conversely, *negative-simplicity* is defined as $-p/\|\mathbf{p}\|$.) So, a policy that is very j-simple (j-simplicity close to one) consists mostly of direction-j rules.[9]

Correspondingly, let the policy's *simplicity* be $|p|/\|\mathbf{p}\|$, where $|p| = |p_+ - p_-|$ is the absolute value of position. That is, policy is simple if most rules have the same direction. Restated slightly, policy is simple if complexity $\|\mathbf{p}\|$ is low relative to $|p|$. At the extreme, if all rules have the same direction, then $|p| = \|\mathbf{p}\|$, and we say that policy is *perfectly simple*.

Notice that any policy $\mathbf{p}$ that is not perfectly simple, so that $|p| < \|\mathbf{p}\|$, is inefficient in the following sense: an alternative policy that achieves the same position p—but has lower complexity—can be constructed by deleting equal masses of positive and negative rules from $\mathbf{p}$. Indeed, both parties dislike complexity and thus are strictly better off under this alternative policy than under $\mathbf{p}$.

II. Myopic Dynamics

This section considers myopic parties, who maximize their objective (2) subject to the flow and entanglement constraints: (4) and (5).

A. *Short-Run Dynamics*

We start by characterizing each party's optimal addition and deletion choices at each instant, which determine how policy $\mathbf{p}$ evolves in the short run. This sets the stage for Section IIB to discuss long-run outcomes.

It is also instructive to rewrite the law of motion (3) in terms of complexity $\|\mathbf{p}\|$ and position p:

(6a)
$$\frac{d}{dt}\|\mathbf{p}(t)\| = \alpha_+(t) + \alpha_-(t) - \delta(t),$$

(6b)
$$\frac{d}{dt}p(t) = \alpha_+(t) - \alpha_-(t) - \delta(t)\frac{p(t)}{\|\mathbf{p}(t)\|}.$$

---

[9] Indeed, j-simplicity $\left(j\frac{p}{\|\mathbf{p}\|} = \frac{p_j - p_{-j}}{p_j + p_{-j}}\right)$ is just the difference between the proportion of direction-j and the proportion of direction-$(-j)$ rules in the policy.
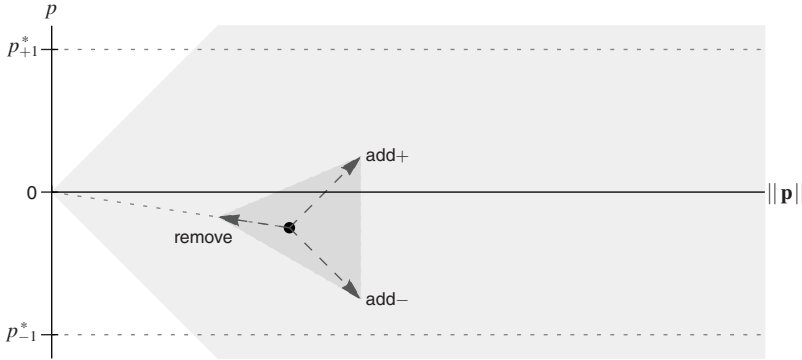
FIGURE 1. $\left(\frac{d}{dt}\|\mathbf{p}\|, \frac{d}{dt}p\right)$ UNDER ADDITION AND DELETION

Figure 1 illustrates how complexity and position change as rules are added or removed. Complexity $\|\mathbf{p}\|$ increases at unit rate when adding rules in either direction, and decreases at unit rate when deleting rules. Position p increases (decreases) at unit rate when adding positive rules (negative rules). The effect of deletion on position p is more subtle. Under deletion, position shifts at rate $-p(t)/\|\mathbf{p}(t)\|$: equal in magnitude to the policy's simplicity, and with *opposite* sign. For example, with a relatively positive-simple policy, deletion would shift position "downward" relatively quickly, as many more positive rules than negative rules are deleted. This has a straightforward geometric interpretation, highlighted in Figure 1: deletion moves $\mathbf{p}$ toward the empty policy $(0,0)$. Note that $|p(t)|/\|\mathbf{p}(t)\| \leq 1$: the rate at which position shifts under deletion is (weakly) lower than under addition. Only for perfectly simple policies does deletion shift position as rapidly as addition (in the corresponding direction).

For concrete exposition, focus on Party $+1$.[10] Figure 2 illustrates Party $+1$'s optimal strategy by depicting, as a function of complexity $\|\mathbf{p}\|$ and position p, the direction in which policy evolves.

Start with policies that lie "below" $+1$'s ideal $\left(p < p_{+1}^*\right)$. Here, Party $+1$'s payoff function (1) simplifies to

$$u_{+1}(\mathbf{p}) = z_{+1}(p - p_{+1}^*) - \|\mathbf{p}\|;$$

Party $+1$'s payoff improves as complexity $\|\mathbf{p}\|$ decreases, and as position p increases toward p_{+1}^*. Combining this observation with the laws of motion (6a) and (6b) yields

$$(7) \qquad \frac{d}{dt}u_{+1}(\mathbf{p}) = \alpha_+(z_{+1} - 1) + \alpha_-(-z_{+1} - 1) + \delta\left(-z_{+1}\frac{p}{\|\mathbf{p}\|} + 1\right).$$

This representation clarifies the pros and cons of adding versus deleting elements. Party $+1$ has two partially conflicting goals: to increase position p toward his ideal,
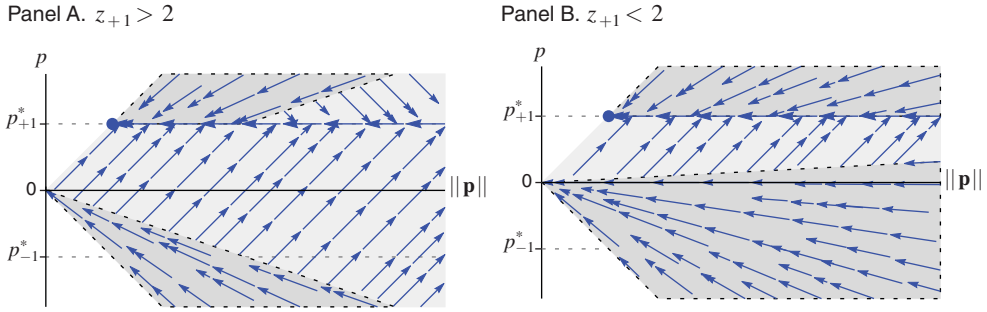
FIGURE 2. PARTY +1'S OPTIMAL STRATEGY

and to reduce complexity. Clearly, adding negative rules is never optimal for $+1$: complexity increases and position moves "downward," away from p_{+1}^*. So, the relevant trade-off for $+1$ is between adding (positive) rules and deleting rules. Deletion reduces complexity. But, relative to positive addition, deletion slows or even reverses the shift in position towards $+1$'s ideal, especially if policy is highly positive-simple (so that deleted rules are mostly positive).

Given this trade-off, Party $+1$ optimally deletes rules if and only if policy is sufficiently negative-simple, so that deleted rules are mostly negative (and thus are "bad" for $+1$); specifically, if and only if $-p/\|\mathbf{p}\| > 1 - 2/z_{+1}$. This deletion region, shaded grey in Figure 2, panels A and B, shrinks as z_{+1} increases: a zealous party prioritizes positional gains over complexity reduction, and thus favors positive addition over deletion. On the other hand, wherever policy is sufficiently positive-simple (and lies below p_{+1}^*), Party $+1$ optimally adds positive rules.

The case where position lies above $+1$'s ideal $\left(p > p_{+1}^*\right)$ is identical, except that directions are reversed: $+1$ has to move "downward" to get closer to his ideal. Here, Party $+1$ optimally deletes rules if and only if $\mathbf{p}$ is sufficiently positive-simple $\left(p/\|\mathbf{p}\| > 1 - 2/z_{-1}\right)$, and optimally adds negative rules otherwise.

To summarize our discussion above, the following proposition specifies Party i's optimal choice at nonideal positions $\left(p \neq p_i^*\right)$. It states that i deletes rules if a sufficiently large proportion of rules are "bad" for him, and adds rules otherwise. Formal proofs are given in online Appendix B.

PROPOSITION 1a: *Suppose that Party i is in control and that policy $\mathbf{p}$ is not i-ideal $\left(p \neq p_i^*\right)$. Let $j = \text{sgn}(p - p_i^*)$ be the direction from Party i's ideal p_i^* to the policy's position p; that is, j is the direction of "bad" rules:*

(i) *If policy is sufficiently j-simple $\left(j\frac{p}{\|\mathbf{p}\|} > 1 - \frac{2}{z_i}\right)$, then Party i deletes rules: $\delta = \gamma$.*

(ii) *Otherwise, if $j\frac{p}{\|\mathbf{p}\|} < 1 - \frac{2}{z_i}$, then Party i adds direction-j rules: $\alpha_j = \gamma$.*

The final case consists of policies at Party $+1$'s ideal $\left(p = p_{+1}^*\right)$. Here, $+1$ has achieved his ideal position, and thus seeks to reduce complexity as *quickly* as

possible while shifting position as *slowly* as possible. Thus, he optimally deletes rules if he is not too zealous and if policy is not too simple, so that deletion does not shift position away from his ideal too quickly. Otherwise, he instead chooses an appropriate combination of addition and deletion to maintain position at his ideal while reducing complexity.[11]

PROPOSITION 1b: *Suppose Party i is in control and policy* $\mathbf{p}$ *is i-ideal,* $p = p_i^*$:

(i) *If policy is sufficiently simple* $\left(\frac{|p|}{\|\mathbf{p}\|} > 1 - \frac{2}{z_i} \right)$, *then Party i reduces complexity while staying on his ideal:*

$$(\alpha_j, \alpha_{-j}, \delta) = \gamma \cdot \left(\frac{|p|}{\|\mathbf{p}\| + |p|}, 0, \frac{\|\mathbf{p}\|}{\|\mathbf{p}\| + |p|} \right), \quad \text{so that}$$

$$\frac{d}{dt} \|\mathbf{p}(t)\| = -\gamma \frac{\|\mathbf{p}\| - p}{\|\mathbf{p}\| + p} \quad \text{and} \quad \frac{d}{dt} p(t) = 0.$$

(ii) *Otherwise, if* $\frac{|p|}{\|\mathbf{p}\|} < 1 - \frac{2}{z_i}$, *then Party i deletes rules:* $\delta = \gamma$.

B. *Path Dependence and Kludge*

We now consider long-run policy dynamics. To move beyond the short run, we have to account for how political competition—as captured by the (random) switches of control between parties—affects the evolution of policy.

We will make two points about long-run outcomes. First, policy complexity is path-dependent: the starting point of policy strongly influences the long-run distribution of complexity. Second, there is a tight long-run relationship between complexity and the distribution of policy positions.

The following terminology hints at the outcomes we will be analyzing. We say that policy becomes *kludged* if $\lim_{t\to\infty} \|\mathbf{p}(t)\| = \infty$. We will focus on two statistics for the long-run distribution of complexity: the probability κ that policy becomes kludged, and a binary indicator K for the possibility of kludge,

$$\kappa = \Pr\left[\lim_{t\to\infty} \|\mathbf{p}(t)\| = \infty \right] \quad \text{and} \quad K = \begin{cases} 0 & : \kappa = 0 \\ 1 & : \kappa > 0 \end{cases}.$$

As a preliminary, observe that any starting policy eventually becomes *regular*, i.e., positioned at or between the parties' ideals: $p \in [p_{-1}^*, p_{+1}^*]$. This is unsurprising. If policy position lies outside the ideals, then both parties will act to shift policy position in the same direction—toward the positional interval $[p_{-1}^*, p_{+1}^*]$, where both ideals lie. Only for regular policies does positional conflict arise between the parties' preferences, leading to interesting dynamics.

---

[11] At the perfectly simple *i*-ideal policy, where $p_{\mathrm{sgn}(i)} = p_i^*$, Party *i* cannot reduce complexity any further without moving away from his ideal, and policy stagnates: $\frac{d}{dt} p = \frac{d}{dt} \|\mathbf{p}\| = 0$.
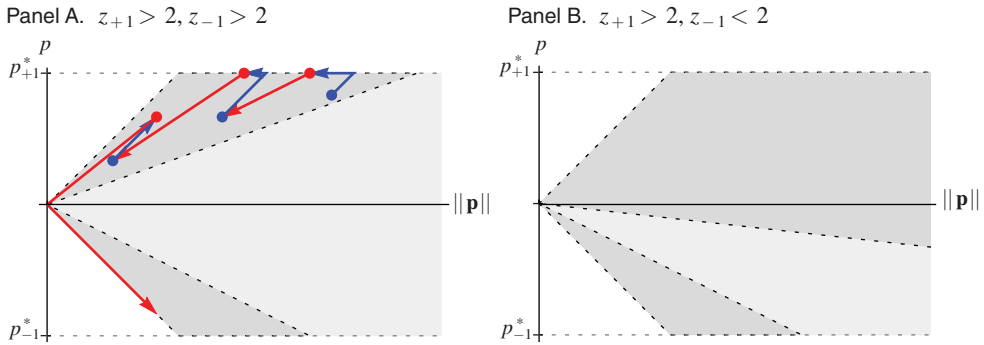
Panel A. $z_{+1} > 2, z_{-1} > 2$ Panel B. $z_{+1} > 2, z_{-1} < 2$

FIGURE 3. BASIN $\mathcal{B}$ (*shaded gray region*)

REMARK 1:

(i) *Suppose that policy* $\mathbf{p}(t)$ *is regular, i.e.,* $p(t) \in [p_{-1}^*, p_{+1}^*]$. *Then policy* (*surely*) *remains regular forever.*

(ii) *Suppose that policy* $\mathbf{p}(t)$ *is not regular. Then policy* (*surely*) *becomes regular at some finite random time* $\tau > t$, *and remains regular thereafter.*

Remark 1 permits us to restrict attention to regular policies. We do so henceforth.

Our first result highlights one aspect of path dependence: *simple policies remain simple*. Let the *basin* $\mathcal{B}$ be the set of regular policies at which at least one party chooses to delete rules (c.f. Propositions 1a and 1b):

$$(8) \quad \mathcal{B} = \left\{ \mathbf{p} : \left(-\frac{p}{\|\mathbf{p}\|} \geq 1 - \frac{2}{z_{+1}} \text{ or } \frac{p}{\|\mathbf{p}\|} \geq 1 - \frac{2}{z_{-1}} \right) \text{ and } p \in [p_{-1}^*, p_{+1}^*] \right\}.$$

Policies that are trapped within the basin can never escape, and tend to grow simpler over time. (See Figure 3.)

PROPOSITION 2: *Suppose that policy lies within the basin:* $\mathbf{p}(t) \in \mathcal{B}$:

(i) *Policy* (*surely*) *remains within the basin forever:* $\mathbf{p}(t') \in \mathcal{B}$ *for all* $t' \geq t$.

(ii) *Policy* (*almost surely*) *becomes perfectly simple at some finite random time* $\tau \geq t$.

(iii) *A perfectly simple policy* (*surely*) *remains perfectly simple forever.*

If both parties are sufficiently zealous $(z_{+1} > 2$ and $z_{-1} > 2)$, then the basin $\mathcal{B}$ consists of relatively simple policies. In this case, the intuition for Proposition 2 can be cleanly stated: *sufficiently simple policies grow* (*weakly*) *monotonically simpler*. This is because neither party benefits from reducing simplicity by "contaminating" a sufficiently simple policy with new rules of the minority type. (See Figure 3, panel A.)

To fix ideas, consider a (regular) policy that is highly positive-simple. From Party $+1$'s perspective, the policy's position is either at or below his ideal $(p \leq p_{+1}^*)$. He

adds positive rules if $p < p^*_{+1}$. He reduces complexity, while maintaining position, if $p = p^*_{+1}$. In either case, simplicity increases. From Party -1's perspective, the policy's position is above his ideal $(p > p^*_{-1})$. He faces a trade-off between adding negative rules and deleting rules. But, because policy is mostly positive-simple, deletion (of mostly positive rules) is optimal for -1. This leaves policy simplicity unchanged.

Given these incentives, simple policies grow progressively simpler as control changes hands between the two parties. In fact, any policy within $\mathcal{B}$ eventually becomes perfectly simple, and remains so. In other words, $\mathcal{B}$ serves as a basin of attraction for the set of perfectly simple policies.[12]

As the parties become less zealous, the basin expands to include less-simple policies. If either party is insufficiently zealous $(z_{+1} \leq 2$ or $z_{-1} \leq 2)$, the basin even contains all policies below or above the $\|\mathbf{p}\|$-axis $(p \leq 0$ or $p \geq 0)$. (See Figure 3, panel B.) In this case, the basin becomes infinite in extent. Consequently, any starting policy inevitably becomes captured within $\mathcal{B}$. Long-run dynamics are mundane in this case.

REMARK 2: *Suppose that $z_{+1} \leq 2$ or $z_{-1} \leq 2$. Then any policy $\mathbf{p}(t)$ (almost surely) becomes perfectly simple at some finite random time $\tau \geq t$, and remains perfectly simple thereafter.*

Hereafter, our analysis will focus on the case where both parties are sufficiently zealous $(z_{+1} > 2$ and $z_{-1} > 2)$.

Outside the basin $\mathcal{B}$, how does policy complexity evolve? In particular, does policy always move into the basin and remain perfectly simple forever? Or, conversely, does complexity increase unboundedly?

Propositions 1a and 1b tell us that outside $\mathcal{B}$, each Party i adds rules toward his ideal, and focuses on reducing complexity when at his ideal. This leads, in equilibrium, to the following laws of motion for position and complexity. For regular $\mathbf{p}(t) \notin \mathcal{B}$, position p moves towards (and stops at) p^*_i while Party i is in control:

$$(9a) \qquad \frac{d}{dt}p(t) = \gamma \cdot \begin{cases} 1 & : p(t) \in [p^*_{-1}, p^*_{+1}) \quad \text{and} \quad i(t) = +1 \\ -1 & : p(t) \in (p^*_{-1}, p^*_{+1}] \quad \text{and} \quad i(t) = -1. \\ 0 & : p(t) = p^*_{i(t)} \end{cases}$$

Whereas, complexity $\|\mathbf{p}\|$ decreases when position is *at* either ideal, and increases when position is *between* ideals;[13] see Figure 4:

$$(9b) \quad \frac{d}{dt}\|\mathbf{p}(t)\| = \gamma \cdot \begin{cases} 1 & : p(t) \in (p^*_{-1}, p^*_{+1}) \text{ or } p(t) = p^*_{-i(t)} \\ -\dfrac{\|\mathbf{p}(t)\| - |p(t)|}{\|\mathbf{p}(t)\| + |p(t)|} & : p(t) = p^*_{i(t)} \end{cases}.$$

---

[12] As we will see shortly, this statement is somewhat imprecise. Depending on parameter values, the basin of attraction for the set of perfectly simple policies is either $\mathcal{B}$ or the larger set of all regular policies.

[13] The exception to this rule is the case where Party i is in control and policy is at $-i$'s ideal: $p = p^*_{-i}$, in which case complexity increases $(\frac{d}{dt}\|\mathbf{p}\| = 1)$. But this case may essentially be ignored because policy spends zero time in this region of the state space: if Party i takes control at $-i$'s ideal p^*_{-i} outside $\mathcal{B}$, he adds j-rules and instantaneously moves policy away from p^*_{-i}.
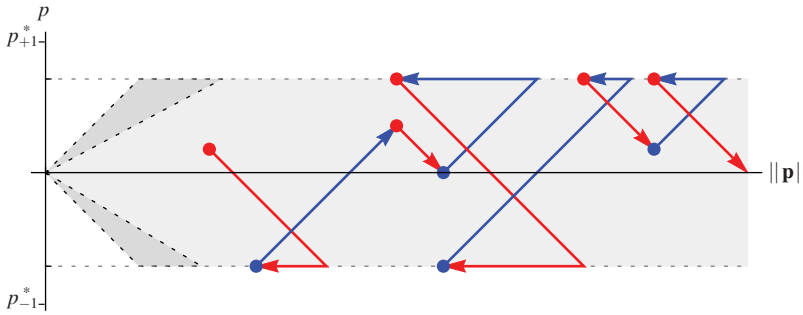
FIGURE 4. INCREASING COMPLEXITY OUTSIDE THE BASIN

This short-run relationship between complexity and position, as expressed by (9b), extends naturally into the long-run. To state the long-run result precisely, first note that outside the basin $\mathcal{B}$, the long-run behavior of position p can be described in terms of its steady-state distribution.

LEMMA 1: *Let $q(t) \in [p^*_{-1}, p^*_{+1}]$ be the random process that obeys, for all $t \geq 0$, the law of motion specified by (9a). Then the Markov process $(q(t), i(t))$ is uniquely ergodic, i.e., has a unique invariant (steady-state) distribution.*

Let $F(\cdot)$ be the steady-state marginal distribution of $q(t)$ from Lemma 1. Define

$$\mu = \int_{[p^*_{-1}, p^*_{+1}]} v(q)\, dF(q) \quad \text{where} \quad v(q) \equiv \begin{cases} 1 & : p^*_{-1} < q < p^*_{+1} \\ -1 & : q = p^*_{+1} \text{ or } p^*_{-1} \end{cases}.$$

We may interpret μ as the long-run average drift of complexity $\|\mathbf{p}\|$ outside the basin $\mathcal{B}$ (given the normalization $\gamma = 1$). Alternatively, and equivalently, $-\mu$ represents the long-run average frequency of ideal positions: it captures how much time p spends *at* ideals instead of *between* ideals. Our next result builds on this equivalence and points out that kludge is possible if and only if ideal positions are achieved infrequently, so that complexity drifts upward in the long-run.[14]

PROPOSITION 3: *Suppose that both parties are sufficiently zealous ($z_{+1} > 2$ and $z_{-1} > 2$), and that the starting policy $\mathbf{p}(0)$ is regular and not in the basin $\mathcal{B}$:*

(i) *If $\mu > 0$, then $K = 1$ and policy (almost surely) becomes kludged or perfectly simple.*

---

[14] One might wonder whether our interpretation of μ as the long-run average drift of complexity is inaccurate, given that the rate at which complexity $\|\mathbf{p}\|$ decreases at ideals, $\left|\frac{d}{dt}\|\mathbf{p}(t)\|\right| = \frac{\|\mathbf{p}\| - |p|}{\|\mathbf{p}\| + |p|}$, is smaller than the rate at which complexity increases between ideals, $\left|\frac{d}{dt}\|\mathbf{p}(t)\|\right| = 1$. However, this difference vanishes at the high-complexity limit: $\frac{\|\mathbf{p}\| - |p|}{\|\mathbf{p}\| + |p|} \to 1$ as $\|\mathbf{p}\| \to \infty$. Indeed, it turns out that the long-run statistics that we calculate are determined by the dynamics of policy at this high-complexity limit. Our interpretation of μ as the drift of complexity reflects this insight.

(*ii*) *If* $\mu < 0$, *then* $K = 0$ *and policy* (*almost surely*) *becomes perfectly simple.*

Proposition 3 is a limited result in some respects. It does not fully characterize the probability κ that kludge occurs, and solves instead for the less-informative binary statistic K. This limitation arises because the policy process $(\mathbf{p}(t), i(t))$ is non-ergodic (i.e., path-dependent), so that long-run distributional outcomes such as κ are difficult to directly characterize.

In other respects, Proposition 3 is quite powerful. It links K to the long-run properties of the (modified) position process $(q(t), i(t))$, which is ergodic and permits a closed-form solution for the long-run distribution—derived in Lemma B.1b in online Appendix B. This enables a rich set of sharp comparative statics results about how K changes with model primitives, which we explore in Section IIC.

Let's return to our discussion of path dependence. Proposition 2 showed that simple policies grow simpler over time. If $\mu > 0$, then the (loose) converse holds as well: complex policies tend to grow more complex. This is illustrated crudely by Proposition 3, which shows that any policy outside the basin $\mathcal{B}$ of relatively simple policies may become kludged. The following result is a cleaner formulation of the same basic points. Even outside the basin, if the initial policy has low (high) complexity, then the probability that policy eventually becomes kludged is low (high).

PROPOSITION 4: *Fix a starting policy that is neutral and not perfectly simple:* $p(0) = 0$ *and* $\|\mathbf{p}(0)\| > 0$. *Suppose that* $z_{+1} > 2$ *and* $z_{-1} > 2$, *and that* $\mu > 0$, *so that* $K = 1$. *Then,*

$$\kappa \to 0 \quad as \quad \|\mathbf{p}(0)\| \to 0,$$

$$\kappa \to 1 \quad as \quad \|\mathbf{p}(0)\| \to \infty.$$

Such dependence on initial conditions has policy implications. Because complexity begets further complexity, one-time interventions to reduce complexity may produce long-run gains that are underestimated in static analyses. As a concrete example, a simplification of the tax code obviously reduces the costs of tax compliance, and has an additional potential benefit: it may potentially prevent the tax code from growing ever more complex, or at least may slow the growth of said complexity.

C. *Comparative Statics: The Politics of Kludges*

For convenient exposition, relabel some of the model's primitives as follows. Define

$$\text{volatility:} \quad \lambda = \sqrt{\lambda_{+1}\lambda_{-1}},$$

$$\text{distance:} \quad \Delta_p^* = p_{+1}^* - p_{-1}^*,$$

$$\text{imbalance:} \quad \Lambda = \max\left\{\frac{\lambda_{+1}}{\lambda_{-1}}, \frac{\lambda_{-1}}{\lambda_{+1}}\right\},$$

where λ, the (geometric) average of control change arrival rates, captures the volatility of political control; where Δ_p^*, the difference between parties' ideals, captures the ideological distance between parties; and where Λ captures the degree of power imbalance between parties. Further, label the following increasing function of Λ as

$$\tilde{\Lambda} = \begin{cases} 1 & \text{if } \Lambda = 1 \\ \dfrac{\log \frac{3 - \Lambda^{-1}}{3 - \Lambda}}{\sqrt{\Lambda} - \sqrt{\Lambda^{-1}}} & \text{if } 1 < \Lambda < 3. \\ +\infty & \text{if } \Lambda \geq 3 \end{cases}$$

Propositions 5a and 5b encapsulate our comparative statics for complexity. We first state the results, then discuss the intuition.

PROPOSITION 5a: *Suppose that both parties are sufficiently zealous* $(z_{+1} > 2$ *and* $z_{-1} > 2)$, *and that the starting policy* $\mathbf{p}(0)$ *is regular and not in the basin* $\mathcal{B}$:

(i) *If* $\Delta_p^* \lambda \gamma^{-1} > \tilde{\Lambda}$, *then* $K = 1$ *and policy (almost surely) becomes kludged or perfectly simple.*

(ii) *If* $\Delta_p^* \lambda \gamma^{-1} < \tilde{\Lambda}$, *then* $K = 0$ *and policy (almost surely) becomes perfectly simple.*

PROPOSITION 5b: *Suppose that the conditions of Proposition 5a. 1 are satisfied, so that* $K = 1$. *Then* κ *is increasing in* z_{+1} *and* z_{-1}.

Proposition 5a follows closely from Proposition 3. It specifies conditions under which the long-run frequency of ideal positions is low (or high) enough that (outside $\mathcal{B}$) complexity drifts upward (or downward), i.e., $\mu > 0$ (or $\mu < 0$). This equivalence helps us to understand the comparative statics specified in Proposition 5a: (i) As imbalance Λ in political power becomes large, the more powerful party attains his ideal position more frequently (Figure 5). So policy drifts downward, and the complexity statistic K decreases; (ii) As legislative friction γ^{-1} increases, each party makes more time to reach his ideal. Thus ideal positions are attained less frequently, and K increases; (iii) As political volatility λ increases, each party spends less time in control. So ideal positions are attained less frequently, and K increases;[15] and (iv) As ideological distance Δ_p^* increases, each party has more distance to cover to reach his ideal. Ideal positions are attained less frequently, and K increases.

Proposition 5b states that as parties becomes more zealous, kludge becomes more likely. The logic of this result differs from those of Proposition 5a. A change in zealousness has no effect on the frequency of ideal positions and thus no effect on the drift of complexity. Instead, an increase in zealousness shrinks the basin $\mathcal{B}$. Consequently, policy becomes less likely to enter the basin and get trapped; rather, policy is more likely to "escape" and become kludged.

---

[15] In fact, an increase in volatility Λ is equivalent to an increase in friction γ^{-1} (and a corresponding time dilation); both changes result in less policy change for each control interval.
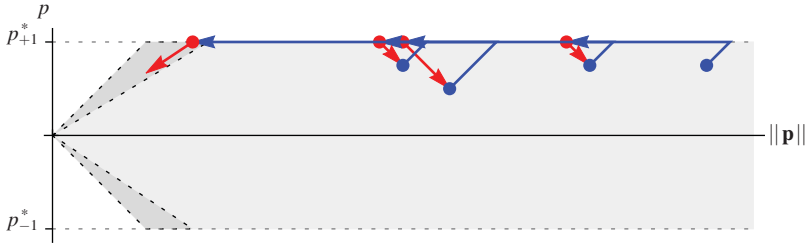
FIGURE 5. LARGE POWER DIFFERENTIAL LEADS TO LESS KLUDGE

A second set of comparative statics relate to policy polarization; that is, the long-run extent to which policy deviates from a "neutral" position. To measure policy polarization, let's adopt the normalization $p_{+1}^* = -p_{-1}^*$, so that the midpoint $p = 0$ between ideals is neutral. Let H be the steady-state marginal distribution of $|p(t)| \in [0, p_{+1}^*]$ under the law of motion (9a). In fact, H is a natural long-run measure of policy polarization: recall that policy eventually becomes either kludged or perfectly simple, and that $p(t)$ obeys the law of motion (9a) in either case. Accordingly, we say that polarization *increases* if $H(\cdot)$ increases in the sense of first-order stochastic dominance.

PROPOSITION 6: *Suppose that $p_{+1}^* = -p_{-1}^*$. Then, policy polarization is:*

  (*i*) *increasing in imbalance Λ,*

  (*ii*) *constant in zealousness z_{+1} and z_{-1}.*

Further fixing $\Lambda = 1$ (political power is balanced), policy polarization is:

  (*i*) *decreasing in friction γ^{-1},*

  (*ii*) *decreasing in volatility λ,*

  (*iii*) *increasing in ideological distance $\Delta_p^* = 2p_{+1}^*$.*

Our comparative statics for complexity and our comparative statics for polarization are closely related. To start, consider those primitives corresponding to technological aspects of political and legislative processes $(\Lambda, \gamma^{-1}, \lambda)$. As recorded in Table 1, any change to one of these primitives that decreases polarization also increases complexity.

Because our measure of deviation $|p|$ is maximized at either ideal, an increase in polarization corresponds (informally speaking) to an increase in $-\mu$, the fraction of time spent at ideal positions relative to between-ideal positions—and thus in a decrease in long-run complexity, as measured by K. That is, holding the ideological distance Δ_p^* fixed, *polarization reduces complexity.*

Table 1—Comparative Statics for Long-Run Outcomes

	Parameter	Symbol	Complexity	Polarization
Institutions	Volatility	λ	↗	↘
	Friction	γ^{-1}	↗	↘
	Imbalance	Λ	↘	↗
Preferences	Distance	Δ_p^*	↗	↗
	Zealousness	z_{+1}, z_{-1}	↗	→

The comparative static effects of changes to political preferences $(\Delta_p^*, z_{+1}, z_{-1})$ differ from those of changes to political processes:

(i) A decrease in ideological distance Δ_p^* reduces both polarization and complexity. Such a change, by reducing the time taken to travel between ideals, ensures that policy spends more time at than between ideals, and thus lessens complexity. But a decrease in Δ_p^* also forces policy to move within a narrower range of positions, and thus mechanically decreases polarization.

(ii) A decrease in zealousness z_{+1}, z_{-1}, while decreasing complexity (Proposition 5b), has no effect on polarization. After all, changes in zealousness preserve the law of motion (9a) and thus also the long-run distribution of position p.

Policy Implications.—Our comparative statics provide some prescriptions for the design of political institutions. A patient planner who seeks to reduce long-run complexity should remove legislative impediments to rulemaking such as supermajority rules and vetoes (i.e., increase γ). She should reduce political volatility, perhaps by increasing the length of election cycles (i.e., reduce λ). Perhaps controversially, rather than balancing power between political parties, she should instead design political institutions that favor one party over others (i.e., increase Λ); stated crudely, she should support autocracies over democracies. However, our model suggests that such changes to the political process may not be costless, even in the long run: decreasing complexity in this fashion may come with an increase in polarization.

On the other hand, a planner can avoid the complexity-polarization trade-off by manipulating political preferences. We prefer to think of such preference manipulation in terms of cultural change: by fostering a moderate political culture and curbing the extremist tendencies of political parties (reducing Δ_p^*), a polity may reduce both complexity and polarization in policy.

III. Strategic Extremism

This section considers strategic behavior by non-myopic parties. We will show that ideologically zealous (high z_i) parties may engage in *strategic extremism*: i.e., move toward positions that are more extreme than their ideals. Such strategic extremism may increase policy complexity in the long run.

Loosely speaking, we consider the limit where both parties are infinitely zealous, but nonetheless care infinitesimally about complexity. This simplification renders

the problem particularly tractable by reducing the associated two-dimensional optimal control problem (over $\|\mathbf{p}\|$ and p) to a one-dimensional problem (over p). Importantly, as we will argue, this simplification preserves the essential dynamic forces in our model, and thus allows us to cleanly highlight the impact of strategic interactions on long-run outcomes.

Start by introducing purely positional preferences: $u_i(\mathbf{p}(t)) = -|p_i^* - p(t)|$, so Party i maximizes

$$E\left[-\int_0^\infty |p_i^* - p(t)| e^{-r_i t} dt\right].$$

We restrict attention to strategies where each Party i has a favored position p_i^{**} called his *target* and—subject to the flow constraint (4) and entanglement constraint (5)—moves toward it as quickly as possible:

$$(10) \qquad \frac{d}{dt} p(t) = \gamma \cdot \begin{cases} 1 & : p(t) < p_i^{**} \\ -1 & : p(t) > p_i^{**}. \\ 0 & : p(t) = p_i^{**} \end{cases}$$

Note that (10) does not specify how complexity evolves, and thus does not uniquely define a strategy. For example, if $p_i^{**} > 0$ and $\mathbf{p}(t)$ contains only negative rules, then (10) may be satisfied either by adding positive rules or by deleting (negative) rules.

Let's mechanically introduce an infinitesimal distaste for complexity. Suppose that each party minimizes $\frac{d}{dt}\|\mathbf{p}\| = \alpha_+ + \alpha_- - \delta$ given the constraint (10). With this additional assumption, the strategy is uniquely defined. The strategy adds complexity everywhere—except at perfectly simple policies with opposite sign to p_i^{**} and at p_i^{**} itself, where the strategy reduces complexity as quickly as possible. That is,

$$(11) \quad \frac{d}{dt}\|\mathbf{p}(t)\| = \gamma \cdot \begin{cases} 1 & : p \neq p_i^{**} \quad \text{and} \quad |p| < \|\mathbf{p}\| \\ \mathrm{sgn}(p)\,\mathrm{sgn}(p_i^{**} - p) & : p \neq p_i^{**} \quad \text{and} \quad |p| = \|\mathbf{p}\| \\ -\dfrac{\|\mathbf{p}\| - p}{\|\mathbf{p}\| + p} & : p = p_i^{**} \quad \text{and} \quad |p| \leq \|\mathbf{p}\| \\ 0 & : p = p_i^{**} \quad \text{and} \quad |p| = \|\mathbf{p}\| \end{cases}$$

We say that a strategy is *focused Markov* if it obeys constraints (10) and (11). Figure 6 depicts focused strategies for Party $+1$. We show in online Appendix B that a (subgame perfect) equilibrium in focused Markov strategies always exists (Lemma B.6f). Hereafter, we restrict attention to these *focused Markov equilibria*.

Focused Markov strategies are very similar to optimal myopic strategies, especially for highly zealous parties. Indeed, returning to Figure 2, we see that a myopic strategy is identical to a focused Markov strategy which targets the party's ideal $(p_i^{**} = p_i^*)$ everywhere except the slivers of policies containing mostly "bad" rules $(j(p/\|\mathbf{p}\|) < 1 - 2/z_i$, where $j = \mathrm{sgn}(p - p_i^*))$. These slivers narrow as zealousness z_i increases; at the limit of infinite zealousness, $z_i \to \infty$, the difference between the optimal myopic strategy and the focused Markov strategy vanishes.

Panel A. Myopic strategy: $p_{+1}^{**} = p_{+1}^*$ Panel B. Strategic extremism: $p_{+1}^{**} > p_{+1}^*$
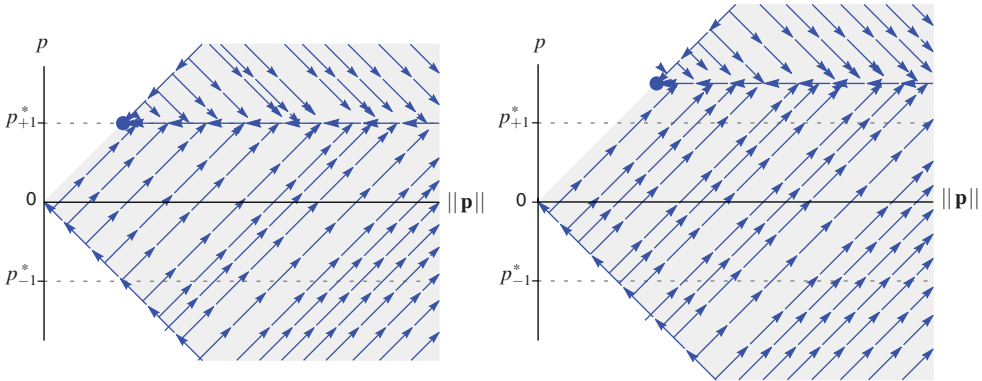

FIGURE 6. FOCUSED STRATEGIES FOR PARTY $+1$

There, focused Markov strategies generalize the myopic strategy by allowing the party to target a non-ideal position.

Consequently, with focused Markov strategies, the dynamics of complexity from the myopic setting of Section II are essentially preserved, with the twist that parties' targets act as "endogenous ideals." Complexity increases when policy lies *between* targets, and decreases when policy is *at* either target. Thus, a version of Proposition 5a holds in this setting, with targets taking the place of ideals: the average drift μ of complexity—and thus the long-run complexity statistic K—increases with the distance between players' targets, denoted as

$$\Delta_p^{**} = \left| p_{+1}^{**} - p_{-1}^{**} \right|.$$

PROPOSITION 7: *Suppose that both parties play focused Markov strategies, and that the starting policy is not perfectly simple* $\left(|p(0)| < \| \mathbf{p}(0) \| \right)$:

(i) *If* $\Delta_p^{**} \lambda > \gamma \tilde{\Lambda}$, *then* $K = 1$.

(ii) *If* $\Delta_p^{**} \lambda < \gamma \tilde{\Lambda}$, *then* $K = 0$.

Given that target locations affect long-run complexity, we seek to understand how strategic considerations affect the parties' equilibrium target choices. Myopic play serves as a benchmark: strategic considerations are absent there, and parties choose their ideals as targets, i.e., $\Delta_p^{**} = \Delta_p^*$. (See Figure 6 panel A.)

We say that Party i engages in *strategic extremism* if he chooses a target that is more extreme than his ideal: that is, if his ideal p_i^* lies between his target and his opponent's target.

PROPOSITION 8: *In any focused Markov equilibrium, both parties always engage in (weak) strategic extremism:* $p_{+1}^{**} \geq p_{+1}^*$ *and* $p_{-1}^{**} \leq p_{-1}^*$. (*See Figure* 6 *panel B*.)

Combined, Propositions 7 and 8 deliver the main lesson of this section. Forward-looking parties take extreme positions, pushing their targets further apart than myopic parties would:

$$p_{-1}^{**} \leq p_{-1}^{*} < 0 < p_{+1}^{*} \leq p_{+1}^{**},$$

so

$$\Delta_p^{**} \geq \Delta_p^{*}.$$

This in turn increases (weakly, and sometimes strictly) the average drift μ of complexity, and correspondingly the long-run complexity statistic K. To summarize: strategic behavior takes the form of strategic extremism, which increases long-run policy complexity.

To understand the logic of strategic extremism, consider the trade-off that Party $+1$ faces between targeting his ideal position p_{+1}^{*} versus a more extreme position $p_{+1}^{**} > p_{+1}^{*}$. Intuitively, an extreme target "shifts the goalposts" upward. This initial upward shift puts policy position further from Party $+1$'s ideal p_{+1}^{*} in the short run. But, after Party -1 takes control and moves policy below p_{+1}^{*}, the initial shift becomes advantageous for $+1$ because policy is now *closer* to p_{+1}^{*} than it would have been under the counterfactual where no initial shift occurred. This advantage is maintained even after subsequent control changes, at least until policy reaches the other target p_{-1}^{**} or moves back above p_{+1}^{*}. Strategic extremism is optimal if the later benefits outweigh the earlier costs.[16]

Because strategic extremism entails short-run costs and medium-run benefits, it arises only if players are sufficiently patient. (Indeed, we already know from Section II that myopic parties do not engage in strategic extremism.) The following proposition presents a version of this intuition. Given targets p_{+1}^{**} and p_{-1}^{**}, define a binary indicator for strategic extremism:

$$\phi = \begin{cases} 1 & \text{if } \Delta_p^{**} > \Delta_p^{*} \\ 0 & \text{if } \Delta_p^{**} = \Delta_p^{*}. \end{cases}$$

PROPOSITION 9: *Fix p_{+1}^{*}, p_{-1}^{*}, and $\lambda_{+1} \neq \lambda_{-1}$, so that parties have unequal durations. Restricting attention to focused Markov equilibria, there exists $\overline{\gamma} > 0$ such that the following hold if friction is high $(\gamma < \overline{\gamma})$:*

(i) *There is a unique pair of equilibrium targets, $(p_{+1}^{**}, p_{-1}^{**})$.*

(ii) *ϕ is weakly decreasing in the parties' discount rates r_{+1} and r_{-1}.*

---

[16] Under stronger assumptions, we may also quantify the extent of strategic extremism by each party, $\Delta_i^{**} = |p_i^{**} - p_i^{*}|$. For example, in online Appendix B (Proposition B.1), we calculate Δ_i^{**} at the asymptotic limit where Δ_p^{*} and γ^{-1} are large. We show that a more vulnerable party (high λ_i) engages in more strategic extremism (high Δ_i^{**}).

(iii) *For sufficiently small discount rates, strategic extremism occurs*: $\phi = 1$.

So, given that strategic extremism occurs only if players are sufficiently patient, kludge may occur with patient players despite being impossible under myopic players. From the perspective of a planner who dislikes complexity, *patience is not necessarily a virtue*.

Reducing Complexity by Trigger Strategies.—We have shown that strategic extremism emerges in focused Markov equilibria, which rule out the use of punishment schemes to enforce cooperation between parties. For an different perspective, we now describe a class of trigger-strategy equilibria where patient parties cooperate to reduce complexity and avoid strategic extremism.

A *focused trigger-strategy equilibrium* is characterized by a *common target* position $\tilde{p}^{**}$ and a pair of *punishment target* positions $\hat{p}_{-1}^{**}$ and $\hat{p}_{+1}^{**}$. On the equilibrium path, both parties focus on the common target $\tilde{p}^{**}$. If either player deviates, each party changes focus to his punishment target $\hat{p}_i^{**}$. In other words: each Party i always obeys constraints (10) and (11), but changes targets from $p_i^{**} = \tilde{p}^{**}$ to $p_i^{**} = \hat{p}_i^{**}$ immediately following any deviation.

It is easy to see that regardless of the initial policy $\mathbf{p}(0)$, a focused trigger-strategy equilibrium achieves the common target position $\tilde{p}^{**}$ in finite time and stays there forever. Once the common target is achieved, complexity $\|\mathbf{p}\|$ decreases monotonically and $\lim_{t\to\infty} \|\mathbf{p}(t)\| = \tilde{p}^{**}$, i.e., policy becomes asymptotically perfectly simple.[17]

The following proposition states that given sufficiently patient parties, and given conditions under which focused Markov equilibria exhibit strategic extremism, there exist focused trigger-strategy equilibria that enforce regular positions, i.e., positions that lie between the parties' ideals. Intuitively, cooperation can be achieved if both parties "compromise" on a regular position as the common target, then punish non-cooperation by reverting to (mutually harmful) strategic extremism.

PROPOSITION 10: *Suppose $\lambda_{+1} \neq \lambda_{-1}$ and $\gamma < \overline{\gamma}$, as in Proposition 9. Suppose both parties have common discount rate r. For some $\underline{r} > 0$ and for all $r \leq \underline{r}$, there exists a focused trigger-strategy equilibrium where (i) the common target is regular, $\tilde{p}^{**} \in [p_{-1}^*, p_{+1}^*]$; and (ii) the punishment targets exhibit strategic extremism, $\hat{p}_{-1}^{**} \leq p_{-1}^*$ and $\hat{p}_{+1}^{**} \geq p_{+1}^*$ (with strict inequality for at least one punishment target).*

Taken together, Propositions 9 and 10 suggest that patience may be a double-edged sword. Whether complexity increases or decreases with patience may depend on

---

[17] If the initial policy $\mathbf{p}(0)$ is perfectly simple, then policy remains perfectly simple forever. Otherwise, policy never becomes perfectly simple because complexity reduction slows asymptotically to zero, $\frac{d}{dt}\|\mathbf{p}(t)\| = -\frac{\|\mathbf{p}\| - \tilde{p}^{**}}{\|\mathbf{p}\| + \tilde{p}^{**}} \to 0$, while policy stays at the common target position $p = \tilde{p}^{**}$. However, a straightforward modification to the focused trigger-strategy equilibrium allows policy to become perfectly simple in finite time; see the discussion following Proposition 10's proof in online Appendix B, and the accompanying Figure B.1.

whether patient parties can successfully build a (self-enforcing) collaborative relationship.

IV. Concluding Remarks

Throughout this paper, we have emphasized the applications of our model to public policy. However, we view our model as also being relevant to other settings in which the design of complicated contracts or policies involves political or ideological disagreement; for example, in the politics of organizational design, or in the decentralized development of open-source software. In particular, the insights we derive in the model can be straightforwardly reinterpreted for an organizational context. For example, our results on long-run kludge suggest that political conflict between different factions within an organization may give rise to persistently inefficient bureaucratic routines and procedures within the organization.

In our model, the structure and density of entanglement—as captured by the entanglement constraint—is specified exogenously. This captures crudely the premise that entanglements between elements of complicated systems are difficult to anticipate, and arise inevitably during the design process.[18] A more nuanced approach would be to partially endogenize entanglement; for example, by allowing parties to reduce or increase the entanglement of new rules from a "baseline" level, perhaps at a cost. Such a setting may produce additional insights. For example, policymakers may deliberately enact highly-entangled rules, so as to obstruct their rivals from undoing those rules in the future.

Political power fluctuates exogenously in our model: in particular, the rates at which random control transitions occur are independent of past and present policy positions. This simplification allows us to cleanly highlight the key forces that animate policy complexity.[19] Future work may consider richer settings where equilibrium policy positions may affect the present and future allocation of political power.

APPENDIX A: THE ENTANGLEMENT CONSTRAINT

This appendix presents two distinct microfoundations for the entanglement constraint (5). The second formulation (Part 2) allows us to parametrize the degree of entangledness, and derive comparative statics.

Part 1: Linear Network.—Here, the policy is a set of *rules* endowed with a total order $\succ$. We say that π depends on π' if $\pi \succ \pi'$. We start by describing the policymaking technology in a discrete setting where each rule has small but positive mass ϵ, and time proceeds in discrete intervals of length ϵ/γ. (We will interpret γ later.) This discrete setting serves to build intuition for the role of dependencies in our analysis. We subsequently focus on the limit $\epsilon \to 0$, where

---

[18] Readers who have written and debugged computer programs will surely sympathize with this premise.
[19] Indeed, this simplification is common to models of dynamic policymaking under political conflict; see, for example, Buisseret and Bernhardt (2017), Dziuda and Loeper (2016), and Chen and Eraslan (2017).

the policymaking technology simplifies to a tractable continuous formulation. As before, p_j denotes the mass of j-rules in $\mathbf{p}$, and $\|\mathbf{p}\| = p_+ + p_-$.

A new rule π added at time t is uniformly randomly allocated a position in the order $\succ$. That is, at the moment of addition, π is equiprobably kth in the ordering for all $k \in \{1, 2, \ldots, |\mathbf{p}(t)|\}$, where $|\mathbf{p}(t)|$ is the number of rules in $\mathbf{p}(t)$ (including π). Ordering is pairwise persistent: if $\pi \succ \pi'$ at time t, then $\pi \succ \pi'$ for all future times τ that $\pi, \pi' \in \mathbf{p}(\tau)$. For simplicity of exposition, consider a single party who is always in control. At the start of each interval, the policymaker may choose any of the following actions, which is then realized at the end of the interval:

(i) Add a new j-rule in either direction $j \in \{+, -\}$.

(ii) Delete the $\succ$-maximal rule.

At time t, the party observes the direction of each rule in $\mathbf{p}(t)$ and the history of all added and deleted rules up until time t, but does not observe the ordering between rules. So, if he chooses "delete" at the start of a time interval, then he observes which rule was $\succ$-maximal (and thus was deleted) only at the end of the interval.

In general, one might expect the party's beliefs about the dependency ordering $\succ$ to evolve in a complicated fashion. Conveniently, our technical assumptions allow us to abstract from the details of (beliefs about) the ordering.

REMARK A.1: *At any time-t history, from the policymaker's perspective, every permutation of the dependency ordering over $\mathbf{p}(t)$ is equally likely.*

So, rules are indistinguishable beyond their direction: all positive rules look alike, and all negative rules look alike. Consequently, the policymaker's beliefs are summarized by the masses (p_+, p_-) of positive and negative rules.

Now, we calculate how p_+ and p_- change over a single time interval under addition and deletion. Let Δp_j denote the change in p_j over a single time interval.

If the party adds a j-rule, then (remembering that $\Delta t = \epsilon/\gamma$)

$$\Delta p_j = \gamma \Delta t \quad \text{and} \quad \Delta p_{-j} = 0.$$

If the party deletes the $\succ$-maximal rule, it is equally likely to be any of the existing rules in $\mathbf{p}(t)$. So, deletion preserves (in expectation) the ratio of positive to negative rules in the policy:

$$E[\Delta p_j] = -\gamma \frac{p_j}{\|\mathbf{p}\|} \Delta t \quad \text{for each } j \in \{+, -\}.$$

More generally, if the party mixes over positive rule addition, negative rule addition, deletion, and doing nothing, then he can achieve (in expectation) any convex combination of addition and deletion outcomes:

(A1a) $$E[\Delta p_j] = \left(\alpha_j - \frac{p_j}{\|\mathbf{p}\|} \delta \right) \Delta t \quad \text{for each } j \in \{+, -\},$$

for any

(A1b) $\alpha_+ \geq 0, \alpha_- \geq 0, \delta \geq 0$ such that $\alpha_+ + \alpha_- + \delta \leq \gamma.$

Now, focus on the limit $\epsilon \to 0$, so that each rule becomes infinitesimally small and time is continuous. Here, the laws of motion (A1a)–(A1b) can be expressed in differential form. The party chooses addition and deletion rates $\alpha_+(t) \geq 0$, $\alpha_-(t) \geq 0, \delta(t) \geq 0$, which determine the velocity of (p_+, p_-):

(A2a) $\dfrac{d}{dt} p_j(t) = \alpha_j(t) - \dfrac{p_j(t)}{\|\mathbf{p}(t)\|} \delta(t)$ for each $j \in \{+, -\}$,

subject to a flow constraint

(A2b) $\alpha_+(t) + \alpha_-(t) + \delta(t) \leq \gamma.$

Together, equations (A2a) and (A2b) are equivalent to the law of motion (3) and entanglement constraint (5).

Part 2: Random Network.—We start with an informal description of the model.

A policy is a continuum of infinitesimal rules. We adopt the convenient expositional convention that each rule has (infinitesimal) mass ϵ. Rules are linked to form an undirected network. Whenever a new rule π is created, it randomly forms a link to each existing rule with (infinitesimal) probability $\rho\epsilon$. So, each new rule π forms ρ links (in expectation) per unit mass of existing rules. We interpret ρ as the degree of *entangledness*. Once formed, links between pairs of rule persist until one or both rules in the pair are deleted.

As in Part 1 of this Appendix, consider a single party who is always in control. The party can add rules in either direction, but cannot precisely target a given rule for deletion. Specifically, if the party targets a rule π to be deleted, the direct neighbors of π will also be simultaneously deleted. As before, the maximum rate of addition and deletion (of all rules, including the neighbors of rules targeted for deletion) is mass γ per unit time.

When formalizing this description, we distinguish between rules added at different times when describing the policy. Say that a rule has vintage-τ if it was added at time τ. Define $p_j(t, \tau)$ to be the "quantity" of vintage-τ j-rules that remain at time $t \geq \tau$. Specify the law of motion of $p_j(t, \tau)$ to be

(A3) $p_j(t, \tau) = \alpha_j(\tau) - \displaystyle\int_\tau^t \delta_j(\tilde{t}, \tau) \, d\tilde{t},$

where $\alpha_j(\tau)$ is the time-τ addition rate for j-rules and $\delta_j(\tilde{t}, \tau)$ is the time-$\tilde{t}$ deletion rate for j-rules of vintage τ. That is, the time-t quantity of vintage-τ rules equals the time-τ addition rate, less the total quantity of vintage-τ rules deleted up until time t.

Let $\hat{\delta}_+(t, \tau)$ and $\hat{\delta}_-(t, \tau)$ be the time-t rate at which the party targets vintage-τ rules for deletion. Let the network structure be characterized by $\rho(j, \tau, \tilde{j}, \tilde{\tau})$, which represents the density of connections between j-rules of vintage-τ and $\tilde{j}$-rules of vintage-$\tilde{\tau}$. To capture the idea that immediate neighbors of deleted rules must also

be deleted, we specify that the vintage-τ deletion rate accounts both for directly targeted rules, and for neighbors of targeted rules from other vintages:

$$\delta_j(t, \tau) = \hat{\delta}_j(t, \tau) + p_j(t, \tau) \sum_{\tilde{j}} \int_0^t \rho(j, \tau, \tilde{j}, \tilde{\tau}) \, \hat{\delta}_{\tilde{j}}(t, \tilde{\tau}) \, d\tilde{\tau}.$$

Define the mass of j-rules at time t to be the total quantity of rules, integrated over all vintages: $p_j(t) = \int_0^t p_j(t, \tau) \, d\tau$. Applying (A3), we get

$$p_j(t) = \int_0^t \alpha_j(\tau) \, d\tau - \int_0^t \delta_j(\tilde{t}) \, d\tilde{t}, \quad \text{where} \quad \delta_j(t) = \int_0^t \delta_j(t, \tau) \, d\tau.$$

In differential form, this replicates (3) from Section I:

$$\frac{d}{dt} p_j(t) = \alpha_j(t) - \delta_j(t).$$

Naturally, we interpret $\delta_j(t)$ to be the rate at which j-rules are being deleted.

We specify that at each time t, the party chooses addition rates $\alpha_+(t)$ and $\alpha_-(t)$, and vintage-specific deletion rates $\hat{\delta}_+(t, \tau)$ and $\hat{\delta}_-(t, \tau)$, subject to the familiar flow constraint (4) on the overall rate of addition and deletion:

(A4) $$\alpha_+(t) + \alpha_-(t) + \delta_+(t) + \delta_-(t) \le \gamma.$$

Our key simplifying assumption is that the density of links across vintages and directions of rules is completely homogenous: $\rho(\tau, \tilde{\tau}, j, \tilde{j}) \equiv \rho > 0$. In that case, some algebra reveals that the deletion rate is

(A5) $$\delta_j(t) = \hat{\delta}_j(t) + \rho p_j(t) \big(\hat{\delta}_+(t) + \hat{\delta}_-(t) \big), \quad \text{where} \quad \hat{\delta}_j(t) = \int_0^t \hat{\delta}_j(t, \tau) \, d\tau.$$

Thus, the deletion rates $\delta_+(t)$ and $\delta_-(t)$ are determined entirely by the targeted deletion rates $\hat{\delta}_+(t)$ and $\hat{\delta}_-(t)$. In other words, it does not matter which rules to target; the only relevant decision is how many rules to delete.

Further, inspection of (A5) indicates that the party can, by appropriately choosing deletion rates $\hat{\delta}_+(t)$ and $\hat{\delta}_-(t)$, achieve any combination of deletion rates $\delta_+(t)$ and $\delta_-(t)$ satisfying

(A6) $$\frac{\delta_+(t)}{\delta_-(t)} \in \left[\frac{1 + \rho p_+(t)}{\rho p_-(t)}, \frac{\rho p_+(t)}{1 + \rho p_-(t)} \right].$$

Accordingly, we may restate the laws of motion as follows:

> *At each time t, the party chooses addition rates $\alpha_+(t)$ and $\alpha_-(t)$ and deletion rates $\delta_+(t)$ and $\delta_-(t)$, subject to the flow constraint (A4) and the entanglement constraint (A6).*

Notice that (A6) is a relaxed version of Section I's entanglement constraint, (5). At the limit $\rho \to \infty$, where the network density becomes large, (A6) tightens into (5).

Our results from Section II continue to hold in this setting, even with finite entangledness ρ. As before, suppose that both parties are myopic. Define the basin $\mathcal{B}$, as before, to be the set of regular policies where at least one party deletes rules. Here:

$$\mathcal{B} = \left\{ \mathbf{p} : \left(\frac{\rho p_+}{1 + \rho \|\mathbf{p}\|} < \frac{1}{z_{+1}} \quad \text{or} \quad \frac{\rho p_-}{1 + \rho \|\mathbf{p}\|} > \frac{1}{z_{-1}} \right) \quad \text{and} \quad p \in \left[p_{-1}^*, p_{+1}^* \right] \right\}.$$

The basin $\mathcal{B}$ expands as entangledness ρ decreases. This is intuitive: as the entanglement constraint loosens, the ability of each party to target rules for deletion improves, and thus deletion becomes optimal over a larger range of policies. Proposition 2 continues to hold in this setting: any policy in $\mathcal{B}$ remains forever in $\mathcal{B}$.

Outside the basin $\mathcal{B}$, the laws of motion (9a) and (9b) continue to hold. Consequently, all of our results about kludge—Proposition 3, Propositions 4, 5a, 5b, and 6—are preserved. Further, we may show that kludge increases with entangledness ρ.

PROPOSITION A.1: *Suppose $K = 1$. Then κ is decreasing in ρ.*

The proof is almost identical to that of Proposition 5b, and thus is omitted. An increase in entangledness ρ shrinks the basin $\mathcal{B}$. Consequently, policy becomes less likely to enter the basin and get trapped; rather, policy is more likely to "escape" and become kludged.

REFERENCES

Azzimonti, Marina. 2018. "Partisan Conflict and Private Investment." *Journal of Monetary Economics* 93: 114–31.

Bendor, Jonathan. 1995. "A Model of Muddling Through." *American Political Science Review* 89 (4): 819–40.

Bernheim, B. Douglas, Antonio Rangel, and Luis Rayo. 2006. "The Power of the Last Word in Legislative Policy Making." *Econometrica* 74 (5): 1161–90.

Buisseret, Peter, and Dan Bernhardt. 2017. "Dynamics of Policymaking: Stepping Back to Leap Forward, Stepping Forward to Keep Back." *American Journal of Political Science* 61 (4): 820–35.

Callander, Steven. 2011a. "Searching and Learning by Trial and Error." *American Economic Review* 101 (6): 2277–2308.

Callander, Steven. 2011b. "Searching for Good Policies." *American Political Science Review* 105 (4): 643–62.

Callander, Steven, and Patrick Hummel. 2014. "Preemptive Policy Experimentation." *Econometrica* 82 (4): 1509–28.

Chen, Ying, and Hülya Eraslan. 2017. "Dynamic Agenda Setting." *American Economic Journal: Microeconomics* 9 (2): 1–32.

Dziuda, Wioletta, and Antoine Loeper. 2016. "Dynamic Collective Choice with Endogenous Status Quo." *Journal of Political Economy* 124 (4): 1148–86.

Ellison, Glenn, and Richard Holden. 2013. "A Theory of Rule Development." *Journal of Law Economics, and Organization* 30 (4): 649–82.

Ely, Jeffrey C. 2011. "Kludged." *American Economic Journal: Microeconomics* 3 (3): 210–31.

Glaeser, Edward L., Giacomo A. M. Ponzetto, and Jesse M. Shapiro. 2005. "Strategic Extremism: Why Republicans and Democrats Divide on Religious Values." *Quarterly Journal of Economics* 120 (4): 1283–1330.

Gratton, Gabriele, Luigi Guiso, Claudio Michelacci, and Massimo Morelli. 2017. "From Weber to Kafka: Political Activism and the Emergence of an Inefficient Bureaucracy." https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2984776.

Levy, Gilat, and Ronny Razin. 2013. "Dynamic legislative decision making when interest groups control the agenda." *Journal of Economic Theory* 148 (5): 1862–90.

Lindblom, Charles E. 1959. "The Science of 'Muddling Through'." *Public Administration Review* 19 (2): 79–88.

McCarty, Nolan, Keith T. Poole, and Howard Rosenthal. 2006. *Polarized America: The Dance of Ideology and Unequal Riches*. Cambridge, MA: MIT Press.

Messner, Matthias, and Mattias K. Polborn. 2012. "The option to wait in collective decisions and optimal majority rules." *Journal of Public Economics* 96 (5–6): 524–40.

Teles, Steven M. 2013. "Kludgeocracy in America." *National Affairs* 17: 97–114.