DYNAMICS OF POLICYMAKING: STEPPING BACK TO LEAP FORWARD, STEPPING FORWARD TO KEEP BACK[1]

Peter Buisseret[2] Dan Bernhardt[3]

Abstract

We study dynamic policy-making when: today's policy agreement becomes tomorrow's status quo; agents account for the consequences of today's policies for future policy outcomes; and there is uncertainty about who will hold future political power to propose and veto future policy changes. Today's agenda-setter holds back from fully exploiting present opportunities to move policy toward her ideal point whenever future proposer and veto players are likely to be aligned *either* in favor of reform, or against it. Otherwise, agenda-setters advance their short-run interests. Optimal proposals can vary discontinuously and non-monotonically with political fundamentals.

JEL Codes: C2, D72

[1] We thank Scott Ashworth, Odilon Camara, Bård Harstad, Niall Hughes, Gilat Levy, Stephane Wolton, and seminar audiences at Bristol, the LSE, Chicago, Queen Mary, Nottingham, Stanford, NYU, Berkeley, the 2015 Priorat Workshop in Theoretical Political Science and the 2015 CRENoS Workshop on Institutions, Individual Behavior and Economic Outcomes for helpful comments and suggestions.

[2] Harris School of Public Policy, University of Chicago, *Email*: `pbuisseret@uchicago.edu`

[3] Department of Economics, University of Illinois, *Email*: `danber@illinois.edu`

Preprint submitted to (...) *April 7, 2016*

1. Introduction

Policies implemented today partly determine the policies implemented in the future. This *dynamic linkage* in policy-making may arise through information (**?**), preferences (**?**), or institutions (**?**). We study the consequences of a dynamic linkage that arises in contexts where existing policy agreements prevail until they are superseded by a new agreement. This may be a consequence of formal institutional rules, such as mandatory spending programs in the United States (**?**). It may also arise *de facto*: an example is the Barnett formula, used in the United Kingdom to adjust public expenditure across Northern Ireland, Scotland and Wales, which was introduced in 1978 as a temporary expedient, but has been in continuous use, ever since.

A crucial feature of these environments is that the immediate payoff from today's policy becomes the *opportunity cost* of changing future policy. In this paper, we ask: how does this affect the short-term reform strategy of an agent whose long-term preference is to move policy away from an unpalatable status quo? How does this strategy vary with the form and degree of uncertainty over who will hold power in the future? And, how do the answers to these questions depend on agents' ideological preferences in favor of, or against, policy reform?

We explore these questions in a political economy setting with far-sighted agents, building on the seminal framework of Romer and Rosenthal (1979*b*). The novel ingredient that we introduce is that agents face uncertainty about who will hold power in the future both to propose and to accept policies vis-à-vis the endogenous status quo.

Our model features a *proposer* and a *restrainer*. The proposer may be an executive, such as a president or prime minister, or a senior legislative office-holder such as the majority leader in a legislative chamber. The restrainer may be the median legislator in the same or another legislative chamber, or a super-majority when such a rule applies. More generally, it may be any agent that can forestall progress on an initiative, such as a faction within a governing party or a coalition of governing parties (**?**, **?**).

We consider three types of restrainer: a *progressive*, a *centrist* and a *conservative*. The centrist and progressive both want to move policy in the same direction away from the extant status quo, but a progressive wants to move policy further than a centrist. The conservative also wants to shift policy, but in the opposite direction from the progressive and centrist. To ease presentation, we assume that the initial restrainer is a centrist.

Likewise, the proposer may either be a *radical* or a *reactionary*. A radical wants to move policy away from the status quo in the same direction as progressive and centrist restrainers,

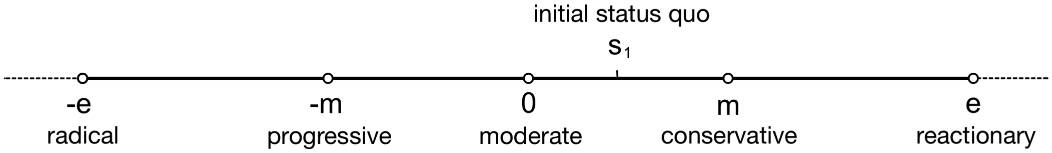


Figure 1: Agents' ideal policies, and the location of the initial status quo

but to a greater extent than both. Similarly, a reactionary wants to shift policy in the same direction as a conservative restrainer, but to a greater extent.

Thus, the proposer and restrainer are *at best* imperfectly aligned. Even when their interests are nominally aligned, e.g., when they belong to the same political party, the 'effective' decisive agent need not precisely share the proposer's preferences. This may be due to explicit supermajority requirements or to the ability of a determined minority—on the chamber floor, in a legislative committee or a faction within the majority party—to impede a bill's progress.

The timing unfolds as follows. At date one, the proposer offers the restrainer a choice between the status quo, and an alternative policy. If the restrainer adopts the proposer's alternative, it becomes the new status quo. Otherwise, the initial status quo remains in place. Between periods, the identities of both the legislative proposer and the restrainer may change, for example, due to an election. Thus, a reactionary proposer may remain in power or be replaced by a radical proposer, or vice versa. Similarly, the centrist restrainer may retain veto power, or be replaced by a progressive or conservative. Once again, the proposer designs a policy. If approved by the restrainer, it is implemented; otherwise the status quo prevails.

We derive how proposals are affected by uncertainty about who will hold future proposal and veto power, agent policy preferences and their relative concern for future policy outcomes. If agents care only about the present, the optimal proposal takes a simple form: move policy as far as possible in the proposer's favored direction, subject to the constraint that the restrainer prefers the outcome to the status quo (Romer and Rosenthal (1979*b*)).

When agents care about the future, though, optimal proposals may take strikingly different forms. We provide conditions under which a reactionary proposes more initial reform than a radical. This arises when, in the future, a radical proposer and progressive restrainer are likely to hold proposal and veto power. In this case, a radical 'steps back in order to leap forward', fostering the opportunity to make dramatic future reform by showing restraint today. For the same reason, a reactionary 'steps forward in order to keep back', sacrificing ground today with a view to preventing more drastic reform in the future.

The standard explanation for why policymakers hold back from fully exploiting political power is that they fear losing power in the future. This explanation is present in our setting, and is especially relevant in non-democratic contexts. For example, an autocratic elite may concede limited redistribution to stave off a threat of revolution (**?**). When political power is initially fully concentrated in a single agent, any change in the distribution of power is necessarily *unfavorable* to the incumbent.

In democratic contexts, however, future political power may evolve *favorably* or *unfavorably* from an incumbent's perspective. In a presidential system, a party may initially control the legislature but then win the presidency, advancing from *divided* to *unified* control of government. A similar phenomenon arises in parliamentary systems: in 2015, the British Conservative Party won a majority, allowing them to dispense with their former coalition partners, the Liberal Democrats. In such settings, there is a second reason to hold back: partial reform today engenders an *opportunity cost* of implementing more powerful reform in the future.

We show that *alignment* of future proposer and restrainer interests is a force for a proposer to refrain from moving policy in her preferred direction, while *mis-alignment* is a force for a radical to shift reform forward, and for a reactionary to maintain the status quo.

Alignment: Suppose a date-1 radical proposer believes that (1) she will retain proposal power and is likely to face a progressive restrainer, or (2) that she is likely to be replaced as proposer by a reactionary who will face a conservative restrainer. The first scenario reflects a presidential system in which election timing is staggered: one agent may remain in office for sure, whilst the other is subject to potential replacement. The second scenario reflects a parliamentary system, or an 'on-year' in a presidential system, where both agents may change during the course of a single election.

If a date-1 radical proposer believes that either (1) a *friendly alignment* between herself and a progressive restrainer or (2) a *hostile alignment* between a reactionary proposer and a conservative restrainer is likely, then she holds back from moving policy as far as possible in her preferred direction. This restraint derives from a future proposer's ability to exploit the discontent of an aligned restrainer with the status quo to move policy further in her desired direction. This discontent is endogenous to the choice of date-1 proposal, which becomes the date-2 status quo: more reform today reduces what a radical can achieve in the future with a progressive restrainer, and it increases the counter-reform that a reactionary can accomplish with a conservative restrainer. So, too, a date-1 reactionary who believes that a *hostile alignment* between a radical and a progressive is likely may implement a partial reform in order to partially acquiesce to a progressive restrainer's desire for change. By doing so, she reduces the ability of a future radical to exploit the progressive, forestalling even greater movement away from the initial status quo.

Mis-alignment: Suppose a date-1 radical proposer believes that (1) she will retain proposal power but is likely to face a conservative restrainer, or (2) she is likely to be replaced by a reactionary proposer who will face a progressive restrainer. Both scenarios can arise in a presidential election where the proposer's identity (the upper chamber) is fixed and the restrainer (the president) may change, or vice versa. A radical proposer is mis-aligned with a conservative restrainer, and a reactionary proposer is mis-aligned with a progressive. In this instance, a date-2 proposer cannot move policy in her preferred direction—gridlock means that the status quo will prevail. When a proposer anticipates future gridlock, she wants to accelerate her agenda rather than hold back.

Our model can make sense of situations in which policy advocacy and opposition cannot be explained by the respective groups' and individuals' contemporaneous policy interests. A powerful illustration of 'stepping forward to keep back' can be found in the Second Reform Act of 1867. That a British Conservative Government would implement legislation extending voting rights to the British working class was long seen as paradoxical. However, **?** argues that "[t]he Act was certainly conservative in that it was an early concession to public opinion" (**?**, 147), while **?** argues that its enactment effectively postponed further reform for nearly 20 years.

A second example from British political history illustrates the phenomenon of 'stepping back to leap forward'. In 1969, the British Labour government attempted to reform the House of Lords—the upper chamber of which membership was partly hereditary—by restricting the voting rights of hereditary peers and weakening their capacity to delay legislation approved in the House of Commons. It was defeated, in part, by a coalition of left-wing abolitionists within the Labour party, led by Michael Foot, who "was anxious that any reform (rather than outright abolition) would merely serve to imbue the House of Lords with greater legitimacy and longevity..." (**?**, 191).

Do politically-minded agents possess the foresight to make these kinds of calculations? Evidence of this foresight is illustrated by a contemporary example from American politics, where the opportunity cost of short-run reform played a prominent role. The 2009 *American Clean Energy and Security Act* was designed to "curb the heat-trapping gases scientists have linked to climate change"[4] by creating a cap and trade system. The legislation was opposed by TheClean.org, "a grassroots coalition... devoted to moving the U.S... to an economy based on renewable energy," which argued:

> Since President Obama is likely to sign the bill with great fanfare, what will the public take away from this? Will they see it as a "win"—that the problem is solved? If so, what will that mean for pushing for the needed steps later? How will the public be mobilized to push their Representatives when the official and media message is that this is "landmark" legislation?
>
> '*Why We Cannot Support This Bill*' (`http://goo.gl/zZ3U3r`)

The effects of primitives on initial proposals hinge on how they affect the future alignment and mis-alignment of proposers and restrainers. An increased likelihood of alignment—both friendly and hostile—strengthens the incentives of both the radical and the reactionary to initially hold back from moving policy in their preferred directions; while an increased likelihood of mis-alignment strengthens their incentives to accelerate their agendas. Moreover, agents are risk averse, so they trade off these forces at different rates.

Thus, if a radical is likely to hold future proposal power, raising the probability of a progressive restrainer by reducing the likelihood of a conservative one, raises the net likelihood of alignment, and reduces the net likelihood of mis-alignment. In turn, this encourages a radical to pursue less initial reform of the status quo, in order to facilitate a future leap forward, and it induces a reactionary to increase initial reform of the status quo, in order to forestall more rampant future reform. If, instead, a reactionary is likely to hold proposer power, then the transfer in probability toward a progressive restrainer serves to reduce the net likelihood of alignment, and to raise the net likelihood of mis-alignment. In turn, this encourages a radical to bring forward reform and a reactionary to hold the line on the status quo.

Our benchmark analysis assumes that the proposer is dynamically sophisticated, but that the restrainer evaluates alternatives to the status quo based on her current payoff. We conclude by considering an initial centrist restrainer who is dynamically sophisticated. If a future radical-progressive pairing is likely, then there are policies further from a centrist's ideal than the status quo—policies that are closer to the radical's ideal point—that

[4] Broder, John (2009-06-26). "House Passes Bill to Address Threat of Climate Change". New York Times. This example was originally cited in **?**.

a dynamically-sophisticated centrist restrainer will accept over the status quo, even though it is worse for her in the short-run. In turn, an initial radical proposer may exploit this sophistication by accelerating reform, since she no longer needs to hold back in order to move policy closer to her ideal. In this instance, a centrist may be hurt by her own sophistication.

If, instead, a reactionary-conservative pairing is likely, an initial radical proposer may propose a policy that is *further* from her ideal than the status quo, and a dynamically sophisticated centrist restrainer may accept in order to forestall an even larger future shift in the direction of a reactionary. Here, both the radical and centrist benefit from the centrist's dynamic sophistication. A practical example is found in Schroeder's defense of his package of rightist social and economic reform, 'Agenda 2010', in which he argued: "Either we modernize ourselves, and by that I mean as a social market economy, or others will modernize us, and by that I mean unchecked market forces which will simply brush aside the social element".[5]

The paper's outline is as follows. After we review the literature, we present our base model. We first analyze scenarios in which proposers never hold back, always exploiting a centrist restrainer to some extent. We then analyze the full model in which the identity of both the proposer and the restrainer may change over time. Finally, we consider how the primitives of the political environment affect proposals, and analyze settings with a dynamically-sophisticated centrist restrainer. A conclusion follows. Proofs are in an appendix.

Related Literature. Our work relates to several important papers in which agents bargain over policies, and there is a reversion point which is either fixed, or evolves as a function of earlier agreement or disagreement. A pioneering contribution is Romer and Rosenthal (1979*b*), in which a proposer with fixed identity makes a proposal (or sequence of proposals) to a group of voters, which is pitted against a fixed, exogenous status quo.

? introduces an endogenous status quo to a spatial legislative bargaining setting with a fixed distribution of agent preferences. He recovers a 'dynamic median voter theorem': policies may move to the left or right in any period, but they gradually converge to the median voter's ideal policy. However, when, as in our model, societal preferences may evolve over time in ways that cannot be perfectly anticipated, convergence to either the present or to the anticipated future pivotal voter's ideal policy need not occur. Some of the strategic considerations that we identify also appear in **?**, who consider the design of an optimal

[5] Gerhard Schroeder, 'Agenda 2010—The Key to Germany's Economic Success', *Social Europe*, 23 April 2012, `http://goo.gl/yCuxgd`

voting rule (e.g., the optimal size of a supermajority) in a legislative setting when citizen preferences may change over time, and preferences of legislators and citizens can diverge. **?** consider optimal voting rules when a society can delay a reform policy to resolve uncertainty about its benefits.

? introduce an endogenous status quo to a 'divide the dollar' setting. Recent work includes **?**, **?**, **?**, **?**, **?** and **?**. Some authors study policy environments with an endogenous status quo that admit both spatial and distributive interpretations (**?**, **?**) and others explicitly include both dimensions (**?**, **?**).

Our work also relates to a literature on the political economy of reform. In these models, uncertainty about economic fundamentals, agent preferences, or future prospects for holding power affect an incumbent's actions in office. **?** argue that parties with an avowed historical opposition to particular reforms—e.g., market liberalization—are often the most likely to implement these policies while in office because their relative hostility to such policies ensures that only they can credibly claim that they are indeed necessary. Such reversals can also occur in our setting. However, our explanation is based not on asymmetric information about the policy context, but rather on a fear that a failure to implement reform now will make the inevitable actions of a successor even more drastic.

Other work on the political economy of reform focuses on other channels through which a dynamic linkage in policy-making can arise. Such work includes **?**, **?**, **?**, **?**, and over the long-run, **?**. In **?**, an initial proposer can work on one of several issues. When future power may change hands, a proposer may use her choice of issue to manipulate a successor's priorities. Unlike in our environment, however, once an issue is addressed it cannot be revisited. Hence, a proposer always implements her static optimum on any issue that she addreses.

2. Model

We consider a two-date economy, with dates 1 and 2. The policy space is $\mathbb{R}$. There are two agents: a *decisive restrainer* ("the restrainer"), and a *legislative proposer* ("the proposer"), whose date-t ideal policies are r_t and p_t, respectively. The legislative proposer may be interpreted as the executive or a senior legislative office-holder. The restrainer could be the median legislator or the 'effective' pivotal legislator in cases where a super-majority requirement applies. In other settings, the restrainer could be the median legislator in a governing party or coalition, or the median legislator in the majority party.

The date-t payoff of an agent with ideal policy i from a date-t policy $y_t \in \mathbb{R}$ is $u_i(y_t) = -(y_t - i)^2$. Initially, there is a status quo $s_1 > 0$ that is inherited from a previous legislative cycle. The proposer is either a *reactionary*, or a *radical*, with ideal policies e and $-e$, respectively, where $e > s_1$. Initially, the restrainer is a *centrist*, with an ideal policy that we normalize to zero. Symmetry of agents' ideal policies eases analysis, but is not needed for our results.

The timing is as follows. At date 1, the proposer first chooses a policy $y_1 \in \mathbb{R}$ that the restrainer may *accept* or *reject*. If accepted, the proposal is implemented; otherwise the status quo s_1 is implemented. The policy implemented at date 1 serves as the status quo s_2 at date 2.

Between dates 1 and 2, an election takes place that may change the identity of the proposer, the restrainer, or both. For example, in a parliamentary system, both agents may change in the same election; in a presidential system in which election timing is staggered, one agent may remain in office for sure, whilst the other is subject to potential replacement. In contexts where proposals originate in the legislature, the change in restrainer could be due to a change in president. At date 2, the restrainer may remain a centrist or be replaced by either a *conservative* restrainer with ideal policy $m > s_1$, or by a *progressive* restrainer with ideal policy $-m$. We use $\Pr(r_2)$ to denote the probability of a type r_2 restrainer at date 2. Likewise, the proposer may remain a reactionary (radical) or be replaced by a radical (reactionary). We let α denote the probability that the date-2 proposer is a radical, and let $\beta = 1 - \alpha$ denote the probability that the proposer is a reactionary. For simplicity, we initially assume that the probability distributions over these transitions are independent, but we later introduce correlation in these distributions.

At date 2, the proposer chooses a policy $y_2 \in \mathbb{R}$, which the restrainer may *accept* or *reject*. If the proposal is accepted, it is implemented; otherwise the date-2 status quo s_2 is implemented. The game then ends.

That the proposer and restrainer are likely to be imperfectly aligned (i.e., $e \neq m$) mirrors real-world settings. In the United States, it is rare for a single party to control the House, Senate and presidency; and even when the same party controls each branch, a supermajority may be required in the Senate. Moreover, preferences may vary across the three branches, for example, if agents face different electoral constituencies—e.g., national versus local electorates. In parliamentary systems where a single party is likely to hold both a legislative majority and the executive, a party leader who acts as a proposer must still win the support of a majority within her own governing party. This problem can be especially severe when parties must work together in a coalition government.[6] Institutional rules may also render the 'effective' restrainer different from the median of the legislative chamber in which the party holds a majority. This would be the case if proposals initiate in a lower chamber, but are subject to veto by an agent in the upper chamber.

Initially, we do not impose an ordering on the ideology of the proposers and the relatively polarized restrainers, i.e., on the ordering of e and m. However, dynamic trade-offs arise almost exclusively in settings where at least one proposer is more ideologically extreme than the corresponding restrainer. Thus, we focus the bulk of our analysis on settings in which the proposer is relatively more 'extreme' than her most closely aligned restrainer, i.e., when $e > m$.[7]

The payoff of an agent with ideal policy i is $(1-\delta)u_i(y_1) + \delta u_i(y_2)$. The weight $\delta \in (0,1)$ captures the degree to which agents value policy made in the next term relative to the current term. A policymaker may place less emphasis on the current term (i.e., δ is close to one) if an election will soon take place, since there will be an imminent opportunity to revise policy after the election. The most natural literal interpretation of our two date formulation is that the policy in place at the end of the second term is subsequently locked in over a sufficiently long horizon that future opportunities to change it are largely discounted by relatively impatient politicians. In practice, it is often politically and practically infeasible for lawmakers to implement frequent major innovations to a policy area (e.g., health insurance).

Throughout, we assume that the legislative proposer is 'dynamically sophisticated': she recognizes that political competition is not a one-shot game and fully accounts for the future consequences of her proposal. To simplify exposition, our benchmark setting assumes that the restrainer evaluates policy solely according to her status quo payoff. This lets us focus on the dynamic concerns of the legislative proposer. Later, we consider a restrainer who is also dynamically sophisticated, highlighting which features of equilibrium do and do not change.

We assume that the distributions over the future holders of proposal and veto power are independent and exogenous. Positive correlation strengthens incentives for an agent to hold back from initially moving policy toward her ideal policy; while negative correlation weakens those incentives. Positive correlation is likely in a parliamentary system, where the forces

[6] In 2010, the Conservative and Liberal Democrat coalition government in the UK endured several high-profile disagreements between the leadership of each party on key policies; see, e.g., "Cameron faces serious Cabinet split over arming Syrian rebels", *The Independent*, June 5, 2013, `http://goo.gl/tZSuwn`.

[7] Allowing for the possibility of a centrist proposer adds little additional insight. In Appendix A, we generalize the model to allow for arbitrary numbers of restrainer and proposer types, and a general recognition rule.

that make a reactionary proposer more likely, also make a conservative restrainer from the same party more likely. In contrast, negative correlation may be likely in an American context where the president faces a mid-term election in which her party is expected to perform badly. In that case, the restrainer's ideology is likely to move away from the proposer's.

We also assume the exogeneity of the distributions over future proposal and veto power. Policy reforms have indirect effects on preferences. That is, given agents' tastes, they affect their *induced* preference trade-offs over future reforms vis-à-vis the induced status quo. Our analysis focuses on this effect. Policy reforms may also change agents' underlying *primitive* preferences. For example, allowing occupants of state-housing to purchase their homes may alter their preferences over different redistributive policies.[8] Translated into our framework, there are settings in which the distribution over proposal and veto power is itself a function of today's policy choices. We make two observations. First, our framework lets us avoid conflating the two channels whilst still uncovering a bevy of subtle trade-offs. Second, the underlying demographics of a society typically change slowly, taking several legislative cycles to evolve.[9]

3. When Won't Politicians Hold Back?

We first identify settings in which politicians *never* want to hold back from moving policy toward their ideal policies, past the policy preferred by a centrist restrainer. These include (1) a static setting where there are no future opportunities to revise policies; (2) the restrainer is always a centrist; or (3) restrainers have more extreme ideologies than proposers, i.e., $m > e$.

Static setting. A static setting is strategically equivalent to date 2 of a dynamic environment, so we drop time subscripts, and refer to the status quo as s, and the ideal points of the proposer and restrainer as p and r. When future opportunities to change policy are absent or fully discounted, then a proposer wants to move policy as close as possible to her ideal point, subject to receiving approval from the restrainer (Romer and Rosenthal (1979*b*)). The restrainer will accept any policy that is closer to her ideal policy r than the status quo s. Suppose that a radical with ideal policy $-e$ holds proposal power.

[8] An example is Margaret Thatcher's controversial 'right-to-buy' policy, in the 1980s.

[9] For example, Glaeser and Shleifer (2005) document the process by which James Curley, an Irish Bostonion politician, attempted to supplant the predominantly English Bostonion population with the Irish, a process that succeeded over the course of fifty years.

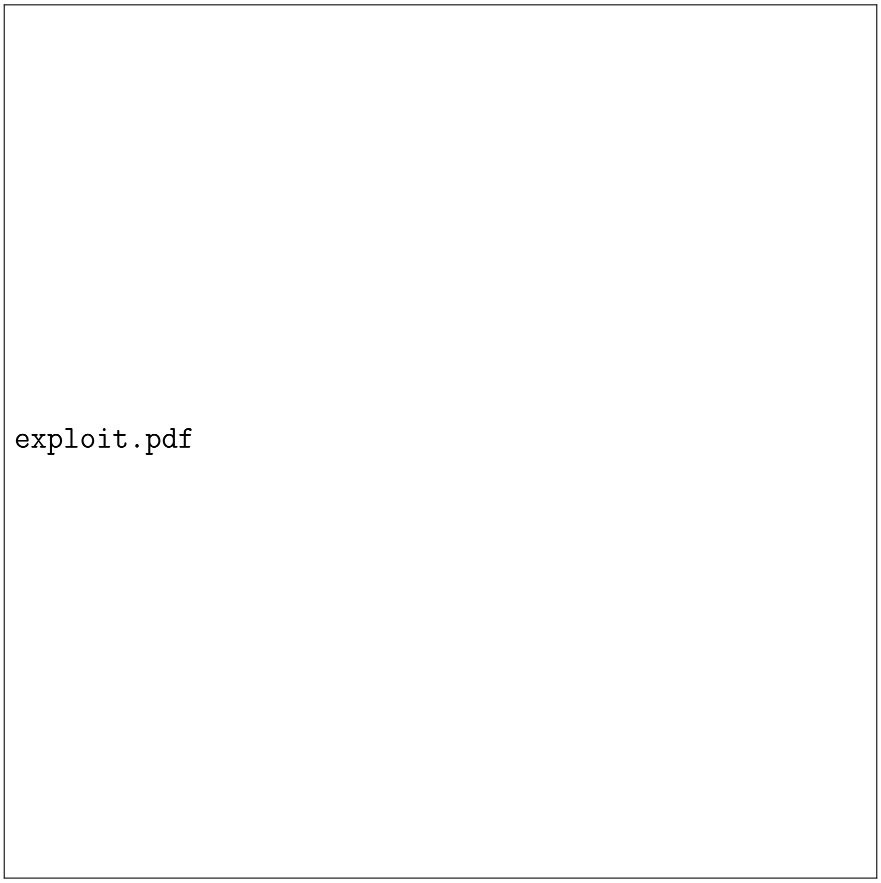


Figure 2: How a radical proposer can exploit the restrainer in a static environment.

(1) If a restrainer has ideal policy $r \geq s$, she will veto any proposal that moves policy toward the radical's ideal policy. Thus, the radical can do no better than propose the status quo.
(2) If a restrainer has ideal policy $r < s$, and her loss is symmetric around her ideal point, she will accept any proposal lying closer to her ideal point than s. Thus, the most reform she is prepared to accept is the policy $y < r$ satisfying $s - r = r - y$, i.e., the policy $y = 2r - s$.

If a radical proposer's ideal policy is sufficiently palatable to the restrainer relative to the status quo, i.e., if $2r - s \leq -e$, then the radical will propose her own ideal policy. Otherwise, she can do no better than $y = 2r - s$. Therefore, the (static) optimal proposal of the radical is:

$$y^*(-e, r, s) = \begin{cases} s & \text{if } s \leq r \\ s - 2(s - r) & \text{if } r < s < e + 2r \\ -e & \text{if } s \geq e + 2r. \end{cases} \tag{1}$$

A proposer's ability to move policy rises with the distance between the status quo s and the restrainer's ideal policy r. This is particularly relevant when a radical proposer and restrainer are partially aligned relative to the status quo, so that $s > r$, but not so much that the restrainer would allow the radical to implement her ideal policy, $s < e + 2r$. In this case, the radical fully exploits the restrainer's desire for reform, jumping policy past r by $s - r$.

Similarly, the optimal proposal of a reactionary with ideology $e > m$ is:

$$y^*(e, r, s) = \begin{cases} s & \text{if } s \geq r \\ s + 2(r - s) & \text{if } 2r - e < s < r \\ e & \text{if } s \leq 2r - e \end{cases} \tag{2}$$

Combining (1) and (2), we see that when the proposer cares only about the immediate consequences of her proposal, she moves the policy outcome as close as possible to her ideal policy.

The nature of this solution has implications for the dynamic setting. When a proposer and restrainer are only partially aligned relative to the status quo, a proposer's ability to move policy rises with the distance $|r - s|$ between the status quo and the restrainer's ideal policy. When the status quo arises from a previous proposal, this feature provides a proposer incentives to refrain from maximizing her static payoff in order to increase her future advantage.

Centrist Restrainer Always Holds Veto Power. Suppose now that today's proposer is uncertain about the identity of tomorrow's proposer, but that the restrainer is sure to remain a centrist. This could reflect a setting in which a legislative chamber is the proposer, the president is the restrainer, and only the legislative chamber faces an imminent midterm election.

At date 1, a centrist restrainer will accept any proposal that is closer to her ideal policy than the status quo. This is not a consequence of our assumption that a restrainer evaluates proposals according to her immediate payoff. On the contrary, if a restrainer is certain to retain veto power, her acceptance strategy is the same when she is dynamically sophisticated, and therefore internalizes the long-run consequences of her acceptance decisions.

Result 1. Suppose a centrist restrainer is certain to hold veto power at both dates. Then at date one a radical proposer proposes $y_1(-e) \leq 0$, while a reactionary proposes $y_1(e) \geq 0$.

The reason is that the induced distribution over date-2 policy outcomes is unaffected by the date-1 choice: for any date-2 status quo $s_2 = y_1 \in [0, s_1]$ or $s_2 = -y_1$, a radical will implement $-y_1$ and a reactionary will implement y_1.

This result does *not* mean that a proposer moves policy as close as possible to her ideal point. In fact, her proposal trades off between static and dynamic incentives. Catering to her immediate payoff also improves future outcomes *if* she is again realized as proposer. This is because she can do no better than lock in her gains at date 2 by maintaining the induced status quo, $s_2 (= y_1)$. However, as she moves policy closer to her ideal, the penalty from losing proposal power grows increasingly severe, rising at twice the rate, since an opposing proposer can reverse the policy in her own favored direction by a distance of $2|y_1|$. Thus,

Proposition 1. If a centrist restrainer always holds veto power at both dates, then an interior solution for a proposer with ideal policy $i \in \{-e, e\}$ satisfies:

$$y_1(i) = i - \delta i 2 \Pr(i \text{ loses proposal power}). \tag{3}$$

A radical's proposal $y_1(-e) \leq 0$ induces future *mis-alignment* between herself and a centrist, and raises the threat of a *hostile alignment* between a reactionary proposer and a centrist. The more likely is the possibility of a hostile alignment, the less reform a radical proposes in order to avoid antagonizing a centrist, who will be easier for a future reactionary to exploit. A reactionary proposer who proposes $y_1(e) \geq 0$ is guided by similar considerations.

When the restrainer is always a centrist, each proposer's concern for the long-run always induces policy moderation. Nonetheless, we will show that dynamic incentives need not induce policy moderation when the identity of the restrainer can also change between dates.

Restrainers are more extreme than proposers. Result 1 extends when (1) the identity of the restrainer may also change in between periods, but (2) $m > e$, so that the ideologies of non-centrist restrainers are more extreme than those of proposers. When $m > e$, a radical who faces a progressive at date 2 can achieve her ideal outcome $-e$ regardless of the location of the status quo $s_1 \geq -e$. The same is true for a reactionary-conservative pairing at date 2 when $s_1 \leq e$. Thus, a date-1 proposer faces no direct trade-offs from the prospects of future radical-progressive or reactionary-conservative pairings. Since the precise location of the date-1 policy only affects the date-2 outcome if the restrainer is a centrist, Result 1 extends: a radical at date 1 selects $y_1(-e) \leq 0$, while a reactionary proposer prefers $y_1(e) \geq 0$.[10]

Thus, the strategically interesting setting is where proposers are more extreme than restrainers, i.e., where $e > m$ (see Figure 1). In what follows, to ease presentation, we assume

[10] We prove this result in the Appendix.

an even greater degree of imperfect alignment between proposers and restrainers:

A1. $e - m > m + s_1$.

A1 implies that for any date-2 status quo resulting from date-1 interactions, each proposer wants to move policy closer to her ideal point than any restrainer would accept.[11]

4. Identities of Proposers and Restrainers May Change Over Time

We now study optimal proposals at date 1 when the identities of both the proposer and the restrainer may change over time, and proposers are more extreme than restrainers.

When the restrainer is always a centrist, we showed that a proposer always chose a policy that rendered the restrainer unwilling to accept further future policy shifts in the proposer's favored direction: the optimal date-1 proposal rendered the centrist restrainer exploitable only by the opposing proposer, at date 2. When the restrainer can change over time, by contrast, each proposer faces a non-trivial decision about which types of restrainers she wants to be partially aligned with her at date 2. When the restrainer is not dynamically sophisticated, she weakly prefers any policy $y_1 \in [-s_1, s_1]$ to s_1. This means that a proposer faces an initial decision about which side of a centrist's ideal policy to place her date-1 proposal. A proposer's continuation payoff from a policy y_1 that becomes the date-2 status quo s_2 is:

$$\begin{aligned} V_i(y_1) &= \alpha\Bigg[\sum_{r_2<y_1} \Pr(r_2) u_i(y_1 - 2(y_1 - r_2)) + \sum_{r_2 \geq y_1} \Pr(r_2) u_i(y_1)\Bigg] \\ &+ \beta\Bigg[\sum_{r_2>y_1} \Pr(r_2) u_i(y_1 + 2(r_2 - y_1)) + \sum_{r_2 \leq y_1} \Pr(r_2) u_i(y_1)\Bigg]. \end{aligned} \tag{4}$$

The date-2 proposer will be a radical with probability α. If the radical holds proposal power and $r_2 < y_1$, then the radical will exploit her *friendly alignment* to shift policy to $y_1 - 2(y_1 - r_2)$. If, instead, $r_2 \geq y_1$, the radical and restrainer are *mis-aligned*, and so she can do no better than maintain the status quo. With probability β, the proposer will be a reactionary. If $r_2 > y_1$, then a reactionary will exploit her own *friendly alignment* with a restrainer to move policy to $y_1 + 2(r_2 - y_1)$. If, instead, $r_2 \leq y_1$ then she and the restrainer

[11] When $0 < e - m < m + s_1$: if $y_1 \in [-s_1, 2m - e]$, a date-2 reactionary who faces a conservative restrainer can implement her ideal policy e and if $y_1 \in [e - 2m, s_1]$, and a date-2 radical who faces a progressive can implement $-e$. These additional cases complicate the analysis without providing insights.

will be *mis-aligned*, and maintain the status quo. Notice that friendly alignment from one proposer's perspective represents hostile alignment from the other's.

Substituting these possible date-2 policy outcomes into $V_i(y_i)$ and recalling the quadratic structure of preferences yields

$$\begin{aligned} V_i(y_1) &= -\alpha\Big[\sum_{r_2<y_1}\Pr(r_2)(y_1-2(y_1-r_2)-i)^2+\sum_{r_2\geq y_1}\Pr(r_2)(y_1-i)^2\Big] \\ &\quad -\beta\Big[\sum_{r_2>y_1}\Pr(r_2)(y_1+2(r_2-y_1)-i)^2+\sum_{r_2\leq y_1}\Pr(r_2)(y_1-i)^2\Big], \\ &= -\alpha\Big[\sum_{r_2<y_1}\Pr(r_2)(2r_2-i-y_1)^2+\sum_{r_2\geq y_1}\Pr(r_2)(i-y_1)^2\Big] \\ &\quad -\beta\Big[\sum_{r_2>y_1}\Pr(r_2)(2r_2-i-y_1)^2+\sum_{r_2\leq y_1}\Pr(r_2)(i-y_1)^2\Big]. \end{aligned} \tag{5}$$

Here, $2r_2-i$ is the date-2 status quo policy $s_2=y_1$ that would allow a proposer with ideology i to move policy all the way to i if she faced an aligned restrainer with ideal policy r_2. From Assumption **A1**, $e>2m+s_1$, so a proposer will never move date-1 policy this far.

Given that a centrist restrainer evaluates a proposal solely according to its immediate payoff implications, the set of proposals that she will accept over the status quo is $[-s_1,s_1]$. Later, we consider a foresighted restrainer who understands both that (1) she may no longer be able to constrain a proposer if she loses veto power, and (2) the proposer may also change.

The optimal policy of an agent with ideology i solves:

$$\max_{y_1\in[-s_1,s_1]}(1-\delta)u_i(y_1)+\delta V_i(y_1). \tag{6}$$

We therefore obtain:

Lemma 1. If the optimal policy of a proposer with ideal point i is interior, then it satisfies:

$$\begin{aligned} y_1(i) &= (1-\delta)i+\delta\Big(\alpha\sum_{r_2>y_1(i)}\Pr(r_2)i+\beta\sum_{r_2<y_1(i)}\Pr(r_2)i\Big) \\ &+ \delta\Big(\alpha\sum_{r_2<y_1(i)}\Pr(r_2)(i+2(r_2-i))+\beta\sum_{r_2>y_1(i)}\Pr(r_2)(i+2(r_2-i))\Big), \end{aligned} \tag{7}$$

where $y_1(i)\in[-s_1,0)$, or $y_1(i)\in(0,s_1]$.

There are at most two solutions satisfying (7)—one on each side of the centrist restrainer's ideal policy—reflecting that whether the centrist restrainer is aligned with one proposer or the other changes as y_1 switches from one side of a centrist's ideal point to the other. Dynamic incentives are determined by two competing channels, an *alignment* channel,

$$\alpha \sum_{r_2 < y_1(i)} \Pr(r_2)(i + 2(r_2 - i)) + \beta \sum_{r_2 > y_1(i)} \Pr(r_2)(i + 2(r_2 - i)), \tag{8}$$

and a *mis-alignment* channel,

$$\alpha \sum_{r_2 > y_1(i)} \Pr(r_2)i + \beta \sum_{r_2 < y_1(i)} \Pr(r_2)i. \tag{9}$$

Alignment Channel: In equilibrium, the initial proposal $y_1(i)$ becomes the date-2 status quo s_2. With probability α, the date-2 proposer is a radical. If $r_2 < s_2$, the radical is aligned with a restrainer, allowing her to move policy to $y_2 = s_2 - 2(s_2 - r_2) = 2r_2 - s_2$. With complementary probability $\beta = 1 - \alpha$, the date-2 proposer is a reactionary. If $r_2 > s_2$, the reactionary is aligned with the restrainer and can move policy to $y_2 = s_2 + 2(r_2 - s_2) = 2r_2 - s_2$.

The absolute magnitude of $i + 2(r_2 - i) = 2r_2 - i$ captures the ideological conflict of interest between a proposer and a partially-aligned restrainer. For a radical, the first term in the alignment channel reflects future *friendly alignment*, and the second term reflects *hostile alignment*. For a reactionary, the first term reflects *hostile alignment* and the second reflects *friendly alignment*. Since $e > 2m$, both friendly and hostile alignment encourage a proposer to refrain from moving date-1 policy toward her ideal. However, a risk-averse proposer weighs *hostile alignment* more heavily than *friendly alignment*. So, unless she is likely to hold future proposal power, a proposer will hold back primarily to prevent future policy moves away from her ideal.

Mis-alignment Channel: At date 2, it may be that the proposer and the restrainer ideologies admit no mutually acceptable alternative to the induced status quo. This occurs if a proposer holds power but faces a restrainer whose ideal point lies on the opposite side of the status quo from her own ideal point. When this happens, s_2 will once again be implemented.

As the prospect of this policy inertia rises, a date-1 proposer prefers either to front-load reform (if she is a radical) or to hold the line against reform (if she is a reactionary). Future gridlock limits both the advantage of holding subsequent proposal power and the disadvantage of losing it. Mis-alignment thus constitutes a form of insurance for a proposer against the adverse consequences of initially accelerating her own agenda.

5. Forces Shaping Incentives to Step Back or Leap Forward

The immediate interest of a proposer is to move date-1 policy in the direction of her ideal policy. However, the future consequences of a proposal present conflicting imperatives. We first focus on 'local' comparative statics that change the location of an interior solution $y_1(i) \in (-s_1, 0)$ or $y_1(i) \in (0, s_1)$ within each interval. We then identify forces that lead to 'jumps' in y_1 from one side of the centrist restrainer's ideal point to the other.

Changes in Concerns for Current and Future Payoffs. A date-1 proposer has a short-run incentive to exploit a centrist as much as possible. However, *any* prospect of a date-2 proposer-restrainer alignment represents a dynamic force for restraint. Thus,

Proposition 2. If a proposer becomes more concerned about future payoffs (i.e., if δ rises), then she always holds back more from moving her initial proposal toward her ideal point.

Short-run incentives yield no trade-offs for a proposer, since she controls the agenda and is constrained only by the necessity of securing acceptance from a centrist restrainer. By contrast, a future prospect of (1) a *friendly* alignment or (2) a *hostile* alignment gives a date-1 proposer a dynamic incentive to refrain from unfettered exploitation of the centrist. Since dynamic incentives always urge more restraint than static incentives, raising a proposer's concern for future outcomes gives her a stronger incentive to hold back at date 1.

Proposition 2 implies that legislative activity (as measured by the submission of bills) should peak at the start of a legislative cycle, and steadily decline as the next cycle approaches. Our explanation is distinct from 'honeymoon' arguments that emphasize a legislature's deference to a president just after his or her election (**?**). Instead, we emphasize the relative imminence of subsequent opportunities to change policy in the future. Regardless of whether the circumstances in which these opportunities arise are expected to be friendly or hostile to today's proposer, their increased proximity always serves as a force for restraint in the short run. Our prediction that legislative activity should peak at the start of a cycle and diminish steadily thereafter is overwhelmingly reflected in the data. Over the period 1974-2013 in the United States Congress, in each two-year congressional session, on average, thirty-five per cent of all bills were introduced in the first four months, fifty per cent in the first seven months, and almost seventy per cent in the first year.[12]

Changes in Uncertainty about Future Power. We next characterize the possibly paradoxical effects of a probabilistic shift toward a more reform-minded restrainer: under plausible circumstances, *both* a radical *and* a reactionary proposer respond by accelerating reform.

Proposition 3. Consider a shift in the distribution over date-2 restrainers that redistributes probability mass from a conservative to a progressive. Regardless of whether a proposer is a radical or a reactionary, she responds by offering less reform if and only if the probability she holds future proposal power exceeds $\frac{1}{2} + \frac{m}{2e}$.

Thus, regardless of whether a proposer is a radical or a reactionary, making a future progressive restrainer more likely results in *more* current reform unless the proposer is very likely to hold future proposal power. When probability mass of $\epsilon > 0$ is taken from the future prospect of a conservative and redistributed to a progressive, the local change in the proposal is:

$$\delta\epsilon \left(i(\beta - \alpha) - \alpha(2m + i) - \beta(2m - i)\right). \tag{10}$$

The probability of a mis-aligned future proposer-restrainer pairing rises by $\epsilon(\beta - \alpha)$. This difference is positive when a reactionary proposer is more likely than a radical to hold future proposal power, since the principal role of a future progressive restrainer is to stand as a bulwark against subsequent counter-reform. A higher likelihood of *mis-alignment* encourages a proposer to move her proposal in the direction of her ideal policy i.

The other two terms in (10) come from the alignment channel, reflecting the change in the net likelihood of alignment, adjusted for the proposer's risk aversion. A radical's prospect of a future *friendly alignment* with a progressive rises by $\alpha\epsilon$, and the prospect of a *hostile alignment* between a reactionary proposer and a conservative falls by $\beta\epsilon$. The same change in primitives raises a reactionary's prospect of a future hostile alignment and reduces the prospect of a friendly alignment.

A risk-averse proposer cares most about policies that result from hostile alignment. Her 'risk-adjusted' change in the net likelihood of alignment is:

$$\delta\epsilon(-i(\alpha - \beta) - 2m), \tag{11}$$

[12] `https://www.govtrack.us/congress/bills/statistics`.

where (1) $\alpha - \beta$ is the net change in the probability of future alignment, (2) $-i$ reflects that a higher prospect of either friendly or hostile alignment—prior to the risk adjustment—encourages a proposer to hold back more from moving policy towards her ideal point and (3) $-2m$ is the risk adjustment, reflecting that a proposer cares more about outcomes further from her ideal point.

If the net probability of alignment rises, i.e., if $\alpha > \beta$, then a reactionary always concedes more initial reform: not only does the net likelihood of alignment rise, but the likelihood of *hostile* alignment rises at the expense of *friendly* alignment. The consequences for a radical are more subtle. Holding back more allows a radical to better exploit a future friendly alignment. However, unless she is very likely to retain proposal power, the first-order effect of a more 'reform-friendly' distribution of veto power is to *lower* her risk-adjusted alignment consideration via the reduced risk of a future hostile alignment. This leads her to bring reform forward. The impact of risk aversion is clearest when the distribution over future proposal power is balanced, i.e., $\alpha = \beta = \frac{1}{2}$: both proposers offer increased reform of $2\delta\epsilon m$.

Due to risk aversion, the requisite threshold on a radical's prospect of holding power also rises as her primitive alignment with a progressive and a reactionary's primitive alignment with a conservative (both captured by $\frac{m}{e}$) rise. This reflects that more aligned hostile pairings make bad policy outcomes even worse; while more aligned friendly pairings make good policy outcomes even better. This raises the wedge between the risk-averse radical's evaluation of these two considerations.

A symmetric logic implies that if the net probability of mis-alignment rises—i.e., if $\beta > \alpha$—a radical proposer responds to a higher prospect of a future progressive restrainer by bringing forward reform in anticipation of future grid-lock. A reactionary responds with less reform only if she is very likely to hold proposal power (i.e., only if β is very large), as only then will the mis-alignment channel dominate the change in her risk-adjusted alignment.

To place the proposition in context, consider a president who is facing a midterm election and is sure to remain in office, but is uncertain about the election's consequences for the ideology of the pivotal legislator in the lower chamber. The proposition implies that if a president anticipates a favorable shift in the preferences of the pivotal legislator, it is better to hold off more on executing her agenda. If instead, the president anticipates an unfavorable shift, then since the next legislative session yields less scope for reversing any initial concessions, she prefers to accelerate her agenda prior to the midterm election. Finally, if the president *also* faces election and there is sufficient uncertainty about whether she will retain office, the proposition reveals that *regardless* of her ideological preferences, she moves the initial policy toward the anticipated location of the new pivotal legislator's ideal policy.

Changes in Ideology. We next derive how changes in the ideological conflict between proposers and restrainers affect the date-1 trade-offs a proposer faces to: (1) increase her *friendly alignment* with future restrainers, (2) lower the *hostile alignment* of an opposing proposer with restrainers, and (3) accelerate her initial agenda in anticipation of future *mis-alignment*.

Greater polarization of restrainers (larger m) affects primitive conflicts of interest. It *lowers* conflict between a radical proposer and a progressive restrainer, and between a reactionary proposer and a conservative restrainer. It also affects how a proposer trades off friendly and hostile alignments, in proportion to their relative likelihood. Greater polarization also *raises* conflict between mis-aligned proposers and restrainers; but such conflicts only matter for aligned agents.

Proposition 4. Regardless of whether the date-1 proposer is a radical or a reactionary, more polarized restrainers (increased m) induce the proposer to offer more reform if and only if:

$$\beta \Pr(r_2 = m) < \alpha \Pr(r_2 = -m). \tag{12}$$

When a friendly alignment is more likely than a hostile alignment, raising m reduces the imperative to raise the value of future *friendly alignment* by holding back, since a proposer can achieve more with the now more-aligned restrainer for any date-2 status quo. The proposer responds by moving policy in the direction of her ideal point.

Conversely, when a hostile alignment is more likely, if a hostile restrainer moves closer to a hostile proposer, it raises the imperative to mitigate future *hostile alignment*. This is because a future opposing proposer can move policy closer to her own ideal point for any date-2 status quo. A proposer responds by moving policy less aggressively toward her ideal point, in order to endogenously lower the alignment between an opposing proposer and restrainer.

To place Proposition 4 in context, suppose there is a right-wing status quo, and an imminent election is expected to bring both the presidency and legislature under the control of the Left. This could arise from a 'coattail' effect, in which legislators who are politically aligned with the presidential candidate are likely to benefit from their candidate's popular support (**?**), and which **?** show leads to the election of more ideologically-extreme senators who support the president. When the legislature is expected to become more ideologically polarized, the initial incumbent proposer—regardless of whether she is a radical or a reactionary—offers more reform. A reactionary makes concessions to avert more drastic future policy shifts. The motives of a radical are quite different: she initiates more reform today since she can already achieve more in the future with a more ideologically polarized aligned restrainer regardless of her initial proposal.

In contrast to the effects of increased polarization of restrainers, greater polarization of proposers affects both static and dynamic trade-offs. It raises a proposer's immediate incentive to move policy toward her ideal, since more extreme ideological preferences raise the direct cost of holding back. Also, in contrast to the effects of increased polarization of restrainers, greater polarization of proposers affects both the alignment and mis-alignment channels.

Proposition 5. Suppose the polarization e of proposers rises. Then each proposer moves her date-1 proposal closer to her ideal policy if mis-aligned proposer-restrainer pairings are more likely than aligned pairings. If, instead, aligned pairings are more likely, then there exists a $\bar{\delta} < 1$ such that if and only if $\delta \geq \bar{\delta}$, each proposer moves her initial proposal further from her ideal policy.

A more extreme proposer suffers a higher date-1 cost from failing to move policy toward her ideal. She also suffers a higher cost of date-2 mis-alignment, since the status quo will be implemented. If the net likelihood of future mis-alignment exceeds that of alignment, static and dynamic considerations both lead a more extreme proposer to accelerate her agenda.

However, a more extreme proposer also has a greater intrinsic conflict of interest with all restrainers. This raises her incentive to hold back to raise future friendly alignment and reduce hostile alignment. By holding back, she lowers her conflict with aligned friendly restrainers; and she reduces an opposing proposer's ability to exploit aligned hostile restrainers.

If aligned proposer-restrainer pairings are more likely than mis-aligned pairings, static and dynamic incentives oppose each other. Then, if and only if a date-1 proposer cares enough about future outcomes—for example, due to an imminent election—will she respond by holding back. As the prospects of aligned pairings rise, the requisite size of δ falls since the initial proposer has greater certainty about the need to hold back from exploiting the centrist for the sake of her date-2 payoff.

To place Proposition 5 in context, suppose that the next election may change the identity of both the president and legislative majority. If control of the two branches is likely to fall to different political parties, a more ideological president will accelerate her agenda before the election. If, instead, the same party is likely to control both branches, the impact of more extreme proposer preferences depends on the imminence δ of the election. If an election is imminent and there will be an opportunity to revisit the issue in the next legislative session, the president holds off working on the issue. This may be due to (1) a fear of losing power to an opposing aligned proposer-restrainer pairing, or (2) an attempt to create even more favorable conditions for aggressive reform. Otherwise, despite the likely prospect of either favorable or unfavorable unified government, a more extreme proposer accelerates her agenda.

Discrete Changes in Proposals. Changes in tastes, uncertainty and concern for the future affect local comparative statics. They also affect the discrete trade-offs associated with whether a proposer wishes to (partially) align herself with a centrist restrainer, as long as there is uncertainty about whether the centrist will hold future veto power. If the centrist will never hold future veto power, the alignment of any future restrainers and proposers is constant for any initial proposal accepted by the initial centrist restrainer, and the local solutions characterized in Lemma 1 coincide. If the centrist always holds veto power, Result 1 implies that each proposer strictly prefers a policy that does not align herself with the centrist.

Suppose, therefore, that the probability tomorrow's proposer faces a centrist restrainer is strictly positive, but less than one. Jumps in optimal policies require that there exist *multiple* interior solutions $y_1^-(-e) \in [-s_1, 0)$ and $y_1^+(-e) \in (0, s_1]$. In turn, multiple interior solutions require that the date-1 proposer be more likely than not to retain proposal power. To see why, recognize that Lemma 1 implies that for a date-1 radical proposer,

$$y_1^+(-e) - y_1^-(-e) = 2e\delta(\alpha - \beta)\,\Pr(\text{centrist}), \tag{13}$$

which can only result in $y_1^+(-e) > 0 > y_1^-(-e)$ if $\alpha > \beta$.

We focus our analysis of 'jumps' in optimal policies by exploring how changes in the uncertainty associated with future proposal power affect an initial radical proposer's preference for aligning herself with a future centrist restrainer. Suppose, then, that $\alpha > \beta$. The radical must address the question: when is it worthwhile to refrain from exploiting the centrist at date 1, (1) in the hopes of retaining proposal power and facing a progressive restrainer, and/or (2) inoculating herself against a future reactionary-conservative pairing? Let $y_1^*(-e)$ denote a radical's globally optimal interior solution. Then, we have:

Proposition 6. Suppose $\alpha > \beta$, so a radical is more likely to hold future proposal power.

1. If $\frac{\Pr(\text{conservative restrainer})}{\Pr(\text{progressive restrainer})} \leq 1 - \frac{2m}{e}$ then there exists $\alpha^*(\delta)$ (*decreasing* in δ) such that

$y_1^*(-e) > 0$ if $\alpha > \alpha^*(\delta)$ and $y_1^*(-e) < 0$ otherwise.

2. If $1 - \frac{2m}{e} < \frac{\Pr(\text{conservative restrainer})}{\Pr(\text{progressive restrainer})} \leq 1$, then $y_1^*(-e) < 0$ for all δ.

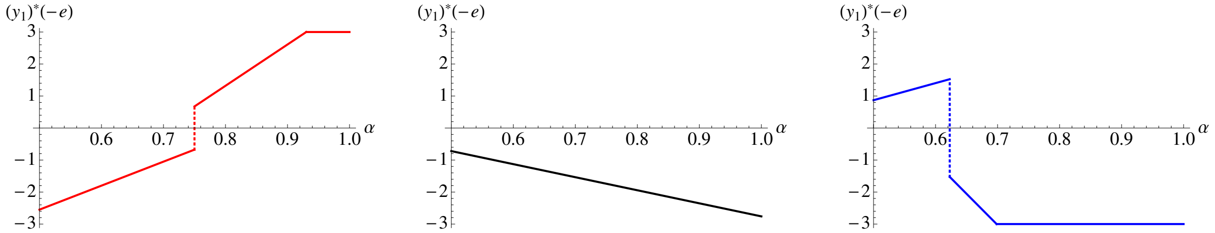
3. If a conservative restrainer is more likely than progressive restrainer, then there exists $\alpha^{**}(\delta)$ (*increasing* in δ) such that $y_1^*(-e) < 0$ if $\alpha > \alpha^{**}(\delta)$ and $y_1^*(-e) > 0$ otherwise.

(1) When $\frac{\Pr(\text{conservative restrainer})}{\Pr(\text{progressive restrainer})} \leq 1 - \frac{2m}{e}$ a conservative restrainer is much less likely than a progressive. Then, when proposal power is fairly balanced—i.e., when $\alpha > \beta$, but the difference is small—a date-1 radical proposer is largely concerned about the risk of a future reactionary proposer, who is likely to face a progressive restrainer, with whom she is mis-aligned. This implies that future policy is likely to remain 'stuck' at the induced status quo, $s_2(= y_1)$. Hence, an initial radical proposer wants to exploit the centrist restrainer immediately.

As the radical's prospects α for retaining proposal power rise, so does the relative value to her of holding back, since she is more likely to retain office where she is likely to benefit from a friendly alignment with a progressive. There exists a threshold $\alpha^*(\delta)$ at which the radical switches from exploiting the centrist to holding back in the hope of extracting more from a future progressive restrainer. This is the point at which it is better to 'step back in order to leap forward more vigorously'. The threshold $\alpha^*(\delta)$ *decreases* in δ since a greater concern for the future increases the willingness of a radical to hold back with even less favorable prospects of holding proposal power.

(2) When $1 - \frac{2m}{e} < \frac{\Pr(\text{conservative restrainer})}{\Pr(\text{progressive restrainer})} \leq 1$, a progressive restrainer is still more likely than a conservative, but their likelihoods are closer. When the initial distribution of proposal power is evenly balanced, today's radical proposer again favors accelerating early reform. However, now as her prospect α of retaining proposal power rises, the greater prospect of a conservative restrainer reduces the prospect of future alignment. This leads the radical to choose an initial proposal to the left of the centrist's ideal point: a radical *never* steps back to leap forward, because the possibility of drawing an aligned progressive is not high enough to sacrifice her ability to better exploit a centrist in one 'jump' rather than two. As a proposer becomes more concerned for date-1 outcomes (i.e., as δ falls), incentives to hold back fall even further.

(3) When $\frac{\Pr(\text{conservative restrainer})}{\Pr(\text{progressive restrainer})} > 1$, a conservative restrainer is more likely than a progressive. Then, if proposal power is initially fairly balanced, the radical is primarily concerned about a hostile reactionary-conservative pairing. The radical initially prefers to neutralize



(a) $\frac{\Pr(r_2=m)}{\Pr(r_2=-m)} \leq 1 - \frac{2m}{e}$ (b) $1 - \frac{2m}{e} < \frac{\Pr(r_2=m)}{\Pr(r_2=-m)} \leq 1$ (c) $\frac{\Pr(r_2=m)}{\Pr(r_2=-m)} > 1$

Figure 3: Illustration of how a radical's optimal date-1 proposal varies with her prospects of holding future proposal power. Parameters: $\delta = 1$, $e = 9$ and $m = 3$. In (a) $\frac{\Pr(r_2=m)}{\Pr(r_2=-m)} = 0$, in (b) $\frac{\Pr(r_2=m)}{\Pr(r_2=-m)} = \frac{7}{13}$, and in (c) $\frac{\Pr(r_2=m)}{\Pr(r_2=-m)} = 30$.

the reactionary's ability to affect a potent counter-reform in the future by opting for a policy that aligns the centrist with the radical. That is, relatively low prospects of holding future proposal power now lead the radical to favor *less* initial reform.

As the prospect that a radical retains proposal power rises, the value of forestalling a reactionary falls. Instead, the first-order effect is to raise the prospect of misalignment between the radical and a conservative restrainer. There exists an $\alpha^{**}(\delta)$ at which a radical switches to accelerating reform, in anticipation of future gridlock. At $\alpha^{**}(\delta)$, the need to hold back from exploiting the centrist for fear of a future reactionary-conservative pairing is trumped by a desire to accelerate reform in anticipation of the induced status quo again being implemented. The threshold $\alpha^{**}(\delta)$ *rises* in δ since a more patient proposer is more willing to hold back from exploiting the centrist to inoculate herself against a future reactionary-conservative pairing.

The asymmetry in the thresholds for $\frac{\Pr(\text{conservative restrainer})}{\Pr(\text{progressive restrainer})}$ reflects risk aversion, since a radical proposer's date-2 payoff is most strongly affected by hostile alignment or misalignment. The ratio $\frac{2m}{e}$ reflects the intrinsic alignment between the progressive and radical. As m rises, the urgency of holding back at date 1 to raise her future alignment with a progressive falls. This raises the bar for a radical to forego early exploitation of the centrist.

Figure 3 illustrates a radical's globally optimal proposal. Note the non-monotonicity in the third panel: here, the date-2 restrainer is most likely to be a centrist, but a conservative restrainer is far more likely than a progressive. When an initial radical proposer is unlikely to hold future proposal power so that $\alpha \leq \alpha^{**}$, she foregoes her ability to exploit the centrist by keeping policy to the right of the centrist's ideal point at date 1. She does so because of her imperative to reduce the future alignment of a hostile aligned reactionary-conservative proposer-restrainer pairing. As a consequence, a radical proposer will be aligned with a centrist restrainer at date 2. As her prospect α of holding future proposal power rises, a radical initially holds back *even more*, but not out of fear of a conservative restrainer. Instead, she holds back to raise her alignment with the centrist. The radical's decision not to exploit the centrist initially leaves open the possibility of exploiting her in the future, and a centrist restrainer is relatively likely to arise at that date.

At α^{**}, the fear of a hostile aligned reactionary-conservative pairing is trumped by the prospect of a mis-aligned radical-conservative pairing. If the radical holds future proposal power, she is most likely either to face a centrist with whom she can achieve no more than she could today, or a conservative with whom she can achieve no reform, at all. She does best to accelerate reform. Conditional on rendering the centrist unwilling to accept any further reform at date 2, however, the radical is almost certainly going to be mis-aligned with tomorrow's restrainer. So, further increases in proposal power lead the radical to *accelerate* reform as much as possible, as if she had based her initial proposal solely on static considerations.

Reversals. We close by highlighting conditions under which a paradoxical 'reversal' occurs: a reactionary proposer moves policy further from her ideal and closer to the radical's ideal than would the radical, herself. This happens if a future radical-progressive pairing is likely:

Proposition 7. If, at date two, the radical proposer is likely to hold power ($\alpha > \beta$) and the restrainer is likely to be a progressive ($\Pr(r_2 = -m) > \frac{1}{2}$) then there exists a $\delta^* < 1$ such that if $\delta \geq \delta^*$, a reactionary proposer proposes more reform at date one than a radical.

If agents who favor reform are likely to enjoy future proposal and veto power, a radical proposer 'steps back' in order to 'leap forward more vigorously' in the future. For the same reason, a reactionary proposer offers incremental reform in order to forestall a wave of even more potent future reform. Risk aversion plays no role in this result. In Appendix D, we extend Proposition 7 to a setting with linear disutility in the distance between agents' ideal policies and policy outcomes. That risk aversion plays no role with linear policy loss follows from the fact that all possible policy outcomes at both dates lie on the same side of a proposer's ideal point. Rather, the key force is a net present value calculation, which trades off a date-1 proposer's prospective future policy gains from holding back (which receive a 'double weighting' with linear loss), relative to the immediate policy loss from failing to exploit the centrist. As the prospect of a future friendly or hostile alignment rises, the magnitude of δ needed to sustain a reversal falls, since there is less uncertainty about the future benefits of initially holding back.

This result can illuminate contemporary and historical examples in which politicians advocate or oppose policies that do not cater to their contemporaneous interests. We earlier elaborated on an attempt in 1969 to reform the House of Lords by the British Labour government that was vanquished, in part, by opposition from within the Labour party. Strikingly, an earlier Conservative government implemented the *Life Peerages Act of 1958*. This Act allowed individuals who did not hold hereditary peerages to be appointed to the House of Lords,[13] and it allowed female peers to sit in the House of Lords. It was bitterly opposed by the Labour party, embodied in Hugh Gaitskell's accusation during the bill's debate:

> "[t]he Bill is not really a reform Bill, as we see it.... It leaves the present powers of the House of Lords unchanged and it gives, conveniently, an apparently slightly more respectable appearance to the House of Lords. We are opposed to a cloak of respectability put upon a person when the reality is quite unchanged."[14]

Subsequent retrospection by Conservatives supports the spirit of Gaitskell's objection. In a policy briefing to fellow parliamentarians in 1998, Conservative Member of Parliament Andrew Tyrie argued: "It was Conservative reforms of the late 1950s and early 1960s which... modernised the Lords enough to protect it from those who wanted it abolished"(**?**, ii). In this policy context, the forces identified in Proposition 7 appear to be quite relevant.

Discussion. In Appendix A, we allow for arbitrary numbers of restrainer and proposer types, and a general recognition rule. For example, we could allow for a centrist proposer who has the same ideal point as the centrist restrainer. Then, when a centrist holds proposal power, and she is aligned with the date-2 restrainer, she can implement her ideal policy regardless of the date-2 status quo. Hence, conditional on realizating a centrist proposer and an aligned restrainer, there are no marginal dynamic trade-offs for *any* date-1 proposer associated with local changes in policy: they all result in the same date-2 outcome. However, the strategic forces we identify conditional on not realizing a centrist proposer remain. More generally, all qualitative characterizations of locally-optimal proposals extend, but with more restrainer types, there are more policies associated with each relevant interval.

[13] Law Lords were previously the only class of non-hereditary peers.

[14] HC Deb 12 February 1958 vol 582, c 423

6. Dynamically-Sophisticated Restrainer

We have focused on the strategic considerations of a proposer given a premise that the restrainer evaluates proposals solely according to their period payoffs. When the restrainer is a pivotal legislator, it may be more natural to afford her the same dynamic sophistication as the proposer so that the restrainer evaluates proposals based on both current and future payoffs.

Since date-2 trade-offs are the same for a restrainer in both the dynamic and myopic cases, we focus on date-1 trade-offs. In general, there may be policies that a date-1 centrist restrainer (a) accepts when she is dynamically sophisticated, but would reject were she myopic, and (b) rejects when she is dynamically sophisticated, but would accept were she myopic.

Suppose, for example, that a date-2 restrainer is certain to be either a progressive or a centrist. Today's centrist restrainer internalizes the dynamic benefit from policies that restrict the scope for future movement away from her ideal point. Since either she or the progressive will hold future veto power, she is concerned about the prospect of a radical-progressive pairing. A dynamically-sophisticated centrist even more strongly prefers a policy $y_1 \in [-s_1, s_1)$ to the status quo than does a myopic centrist. Like a myopic centrist, she enjoys a higher period payoff from such a policy vis-à-vis the status quo. In addition, her dynamic benefit rises because the new status quo reduces the ability of the radical to exploit the progressive.

This observation means that there are now policies $y_1 < -s_1$ that are further from a centrist's ideal than the status quo—policies that are closer to a radical's ideal point—that a date-1 dynamically-sophisticated centrist restrainer will accept over the status quo, even though such policies yield a lower period payoff: a sophisticated centrist will accept policies $y_1 < -s_1$ when the probability of a future radical-progressive pairing is high enough.

Indeed, a dynamically-sophisticated centrist restrainer may even accept policies $y_1 \in (-e, -m)$ that lie to the *left* of a progressive restrainer's ideal. This is because $s_2 \in (-e, -m)$ *ensures* that policy cannot move further toward a radical's ideal point. When a radical-progressive pairing is likely and δ is high, a centrist restrainer may want to inoculate herself against the future ability of a radical proposer to exploit a progressive restrainer.

The set of proposals that a sophisticated centrist restrainer accepts need not be connected: she may accept proposals that are both to the left of the progressive restrainer's ideal point, and to the right, but do not include an interval around the progressive's ideal. This can happen when a future reactionary-progressive pairing is likely: policies to the left of the progressive (but not the right) align a reactionary and progressive; and a far-sighted centrist can then

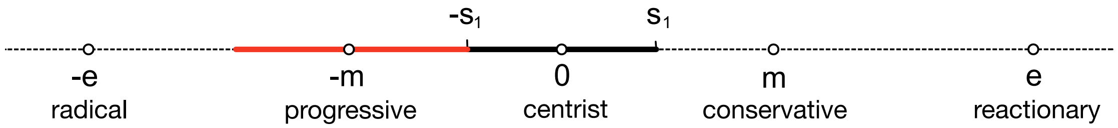


Figure 4: Illustration of the dynamically-sophisticated centrist restrainer's acceptance set at date 1 when the probability of a radical proposer at date 1 is high. The red line represents additional policies that she accepts because she partly internalizes the value of averting a future policy outcome that is closer to the radical's ideal policy.

gain when a future reactionary 'exploits' this by moving policy closer to the centrist's ideal.

Proposition 8. If a date-2 restrainer is always a progressive or centrist, then a dynamically-sophisticated centrist restrainer prefers

$$y_1 \in [\max\{-m, -s_1 - 4\delta\alpha \Pr(r_2 = -m)m\}, s_1)$$

to s_1. If $\delta\alpha \Pr(r_2 = -m) \geq \frac{1}{4}\left(1 - \frac{s_1}{m}\right)$, she also prefers some policies $y_1 \in [-2m - s_1, -m]$ to s_1.

A radical proposer can exploit a dynamically-sophisticated centrist restrainer's fear of a future radical-progressive pairing. Since the restrainer is prepared to accept policies closer to the radical's ideal point at the outset, the opportunity cost to a radical proposer of holding back at date 1 rises. This induces a radical to exploit the dynamically-sophisticated centrist to a greater extent than she would exploit a myopic centrist. In fact, a centrist restrainer may be worse off for her dynamic sophistication, since it renders her vulnerable to exploitation. In the Appendix, we provide conditions under which the 'reversals' documented in Proposition 7 arise with a sophisticated centrist restrainer, despite the incentives of a radical to exploit a centrist's fear of a future radical-progressive pairing.

Analogous results obtain when the date-2 restrainer is always a centrist or a conservative. If a future reactionary-conservative pairing is likely, a dynamically-sophisticated centrist restrainer may accept proposals that move date-1 policy *past* the status quo toward the reactionary's ideal point, since they forestall more extreme future outcomes.[15] In this case, both the radical proposer and the centrist restrainer benefit from the centrist's sophistication.

[15] Note that in this case, the centrist would *not* accept proposals close to $-s_1$—her acceptance set expands in one direction, but shrinks in the other.

Proposition 9. Suppose the date-2 restrainer is always a centrist or a conservative. Then, a date-1 radical proposer proposes a policy $y_1 \in [s_1, e)$ if δ is sufficiently large and either:

1. s_1 is close enough to a centrist's ideal, $s_1 \leq 2\beta \Pr(r_2 = m)m$, or;
2. s_1 is far enough from a centrist's ideal, $s_1 > 2\beta \Pr(r_2 = m)m$; a reactionary-conservative pairing is relatively likely; and proposers are polarized, i.e., e is sufficiently large.

Proposition 9 can explain the well-documented phenomenon that left-wing governments are as likely as right-wing governments to privatize state-owned industries, or to engage in deficit-cutting and other pro-market reforms (**?**, **?**). A prominent explanation offered by **?** is that politicians have private information about the necessity of these policies. In such a setting, left-wing parties can more credibly appeal to the necessity of such policies than can a right-wing party because left-wing parties are intrinsically more hostile to these policies, regardless of fundamentals. Though we also lean on the primitive hostility of a radical to the status quo as a source of 'reversals', the only uncertainty in our model concerns who holds power in the future. Our explanation is closest to Schroeder's defense of 'Agenda 2010': "Either we modernize ourselves, and by that I mean as a social market economy, or others will modernize us, and by that I mean unchecked market forces which will simply brush aside the social element".[16]

7. Conclusion

Knowing when to 'step back' —whether primarily to leap forward or instead to keep back—is a strategic imperative for political agents seeking not only to make short-run gains, but also to achieve long-term policy goals. We show that the prospect of losing *or* retaining political power yield two distinct rationales for agents to refrain from moving policy fully toward their ideal points. We characterize when radical reform advocates prefer less short-run reform than would opponents of reform, and illuminate our results with examples in which politicians advocate or oppose policies that do not cater to their contemporaneous interests.

Although our interpretations and illustrations of the political context have been legislative, the dynamic trade-offs we uncover have more general significance. In her study of

[16] Gerhard Schroeder, 'Agenda 2010—The Key to Germany's Economic Success', *Social Europe*, 23 April 2012, `http://goo.gl/yCuxgd`

social movements, **?** argues that incremental victories can have unintended consequences for a movement's ability to mobilize resources in the future. She finds "movements seek to make incremental gains in advancing their larger policy agenda; [but] this success carries a risk of long-term movement decline, as it can... enervate programmatic activity as continued gains potentially diminish the urgency of the issue or the demonstrable need for greater activism" (**?**, 406). She concludes: "[s]uccess can be a bit of a poisoned chalice to groups if their demonstrated ability to achieve good outcomes leads to subsequent attrition in support levels" (**?**, 408).

Similar issues arise in legal contexts. **?** assesses the trade-offs faced by the NAACP in pursuing legal attacks on racial segregation in U.S. schools. The NAACP sought "to secure decisions, rulings and public opinion on the broad principle instead of being devoted to merely miscellaneous cases".[17] After *Brown v. Board of Education of Topeka* (1954), civil rights lawyers who prosecuted local cases faced a tension between "serving two masters": their local clients, and the NAACP, which sought "to develop a broad scale attack on Jim Crow institutions" (**?**, 1288). In particular, "civil rights lawyers would not settle for anything less than a desegregated system", even when local plaintiffs might have settled litigation in return for promises of better segregated schools. As a counsel to the NAACP in Mississippi, Bell advised a community whose segregated school had been closed by local authorities. He warned that they would not receive support if they just attempted to re-open the school, but that they would receive support if they pursued a full-scale desegregation suit, which was eventually filed in 1963—one of the first in the state (**?**, 476-477).

[17] **?** quoting from 1934 NAACP Report 22.

8. Appendix A: More Proposers and Restrainers, General Recognition Rule.

In this Appendix, we generalize our benchmark setup to allow for a set of N agents with ideal policies $x_1 < x_2 < ... < x_N$. To avoid a plethora of sub-cases, we assume that $x_i - x_{i-1} = \epsilon > 0$ for all $i \in \{2, ..., N\}$, i.e., the ideal policies are evenly located across the policy space.

At date 2, an agent with ideal policy x_i is recognized to serve as the proposer with some probability $p(x_i)$. Similarly, an agent with ideology x_j—including, possibly, the proposer—is recognized as the date-2 restrainer, with probability $r(x_j)$. The model is otherwise unchanged. We characterize optimal (interior) date-1 proposals, and relate them to our benchmark analysis. We first write down the continuation payoff of an agent with ideal policy x_i, for any date-1 outcome y_1, generalizing (4), in the main text:

$$\begin{aligned} V_i(y_1) &= \sum_{x_j<y_1} p(x_j)\Bigg[\sum_{x_k\leq\frac{1}{2}(x_j+y_1)} r(x_k)u_i(x_j) + \sum_{x_k\in(\frac{1}{2}(x_j+y_1),y_1)} r(x_k)u_i(y_1-2(y_1-x_k)) + \sum_{x_k\geq y_1} r(x_k)u_i(y_1)\Bigg] \\ &+ \sum_{x_j>y_1} p(x_j)\Bigg[\sum_{x_k\geq\frac{1}{2}(x_j+y_1)} r(x_k)u_i(x_j) + \sum_{x_k\in(y_1,\frac{1}{2}(x_j+y_1))} r(x_k)u_i(y_1+2(x_k-y_1)) + \sum_{x_k\leq y_1} r(x_k)u_i(y_1)\Bigg] \\ &+ \sum_{x_j=y_1} p(x_j)u_i(x_j). \end{aligned} \tag{14}$$

To understand this expression, first consider a date-two interaction in which the proposer has ideal policy x_j, and the date-1 policy outcome is $y_1 > x_j$. If the date-2 restrainer has an ideal point $x_k \leq \frac{1}{2}(x_j + y_1)$, then the receiver weakly prefers the proposer's ideal policy, x_j, to the status quo. If the date-2 restrainer has an ideal point $x_k \in (\frac{1}{2}(x_j+y_1), y_1)$, then the proposer and restrainer are imperfectly aligned: they both prefer a policy $y_2 \in [y_1 - 2(y_1 - x_k), y_1]$ to the policy y_1. Since the proposer has bargaining power, she will propose the policy closest to her ideal policy, from this interval. Finally, if $x_k \geq y_1 > x_j$, the receiver and proposer are mis-aligned: there are no policies that both strictly prefer to the status quo policy. In that event, the policy outcome will be y_1. The second line analyzes the analogous setting in which the date-1 policy outcome is $y_1 < x_j$. Finally, if the date-2 proposer's ideal policy coincides with the status quo policy, y_1, the date-2 proposer can implement this policy by proposing it.

It follows that if the date-1 proposer has ideal policy x, and its optimal proposal is on the interval (x_i, x_{i+1}), for some $i \in \{1, ..., N-1\}$ then the proposal is given by

$$y(x, x_i) = \frac{(1-\delta)x + \delta Mx + \delta A}{1 - \delta F} \tag{15}$$

where:

$$\begin{aligned} A = & \sum_{x_j < y(x,x_i)} p(x_j) \sum_{x_k \in (\frac{1}{2}(y(x,x_i)+x_j), y(x,x_i))} r(x_k)(x + 2(x_k - x)) \\ & + \sum_{x_j > y(x,x_i)} p(x_j) \sum_{x_k \in (y(x,x_i), \frac{1}{2}(y(x,x_i)+x_j))} r(x_k)(x + 2(x_k - x)), \end{aligned} \tag{16}$$

$$M = \sum_{x_j < y(x,x_i)} p(x_j) \sum_{x_k \geq y(x,x_i)} r(x_k) + \sum_{x_j > y(x,x_i)} p(x_j) \sum_{x_k \leq y(x,x_i)} r(x_k), \tag{17}$$

and

$$F = \sum_{x_j < y(x,x_i)} p(x_j) \sum_{x_k \leq \frac{1}{2}(y(x,x_i)+x_j)} r(x_k) + \sum_{x_j > y(x,x_i)} p(x_j) \sum_{x_k \geq \frac{1}{2}(y(x,x_i)+x_j)} r(x_k). \tag{18}$$

Our expression for $y(x, x_i)$ generalizes the interior solution characterized in (7), in the main text. The term M constitutes the *mis-alignment* channel. It sums over all proposer-restrainer pairings for which no mutually preferred policy to the status quo, $y(x, x_i)$, exists. As in the benchmark setting, this is a force for the initial proposer to move the date-1 policy closer to her ideal policy, x. The term A constitutes the *alignment* channel. The term F is the total probability that the proposer and restrainer both prefer the proposer's ideal policy to the status quo. Whenever this proposer-restrainer pairing occurs, the precise location of $y(x, x_i)$ on the interval (x_i, x_{i+1}) makes no difference to the outcome: the proposer proposes her ideal point. In the benchmark setting, we ruled out any such pairings by assuming that proposers are sufficiently more extreme than the polarized restrainers, i.e., $e - m > m + s_1$: our interior solution in equation (7) reflects that $F = 0$, in that case.

9. Appendix B: Proofs

We adopt parameterization $\Pr(r_2 = -m) = p$, $\Pr(r_2 = 0) = q$, and $\Pr(r_2 = m) = 1 - p - q$.

Proof of Result 1. We first show that if the centrist restrainer is certain to hold veto power then a radical proposes $y_1 \leq 0$ and a reactionary proposes $y_1 \geq 0$. A proposer with ideology i derives payoff $(1 - \delta)u_i(y_1) + \delta(\alpha u_i(-y_1) + \beta u_i(y_1))$ from $y_1 \in (0, e)$ and payoff $(1 - \delta)u_i(-y_1) + \delta(\alpha u_i(-y_1) + \beta u_i(y_1))$ from proposal $-y_1 < 0$, The payoff difference is $(1 - \delta)(u_i(y_1) - u_i(-y_1))$ is strictly negative if $i = -e$ and strictly positive if $i = e$.

Suppose, next, that the date-2 restrainer may be a progressive, centrist or conservative. We show that if $e < m$, then Result 1 again applies. The payoff from $y_1 > 0$ is

$$\begin{aligned} & (1-\delta)u_i(y_1) + \delta\alpha(pu_i(-e) + qu_i(-y_1) + (1-p-q)u_i(y_1)) \\ + \quad & \delta\beta(pu_i(y_1) + qu_i(y_1) + (1-p-q)u_i(e)), \end{aligned} \tag{19}$$

and that from proposal $-y_1 < 0$ is:

$$\begin{aligned} & (1-\delta)u_i(-y_1) + \delta\alpha(pu_i(-e) + qu_i(-y_1) + (1-p-q)u_i(-y_1)) \\ & +\delta\beta(pu_i(-y_1) + qu_i(y_1) + (1-p-q)u_i(e)). \end{aligned} \tag{20}$$

Taking the difference of these two expressions yields the result. □

Proof of Lemma 1. Solve the first-order condition to $\max_{y_1 \in [0,s_1]}(1-\delta)u_i(y_1) + \delta V(y_1)$ for:

$$\begin{aligned} y_1(i) = (1-\delta)i + \delta\Big(\alpha \sum_{r_2 > y_1(i)} \Pr(r_2) + \beta \sum_{r_2 \le y_1(i)} \Pr(r_2)\Big)i \\ + \delta\Big(\alpha \sum_{r_2 \le y_1(i)} \Pr(r_2)(2r_2 - i) + \beta \sum_{r_2 > y_1(i)} \Pr(r_2)(2r_2 - i)\Big), \end{aligned} \tag{21}$$

which characterizes an interior solution in $[0, s_1]$. Likewise, an interior solution to $\max_{y_1 \in [-s_1,0]}(1-\delta)u_i(y_1) + \delta V(y_1)$, is characterized by:

$$\begin{aligned} y_1(i) = (1-\delta)i + \delta\Big(\alpha \sum_{r_2 \ge y_1(i)} \Pr(r_2) + \beta \sum_{r_2 < y_1(i)} \Pr(r_2)\Big)i \\ + \delta\Big(\alpha \sum_{r_2 < y_1(i)} \Pr(r_2)(2r_2 - i) + \beta \sum_{r_2 \ge y_1(i)} \Pr(r_2)(2r_2 - i)\Big). \ \square \end{aligned} \tag{22}$$

Proof of Proposition 2. We have:

$$\frac{1}{2}\frac{\partial y(i)}{\partial \delta} = \alpha \sum_{r_2 < y_1(i)} \Pr(r_2)(r_2 - i) + \beta \sum_{r_2 > y_1(i)} \Pr(r_2)(r_2 - i), \tag{23}$$

and since $|i| \ge |r_2|$ for all $r_2 \in \{-m, 0, m\}$, $\operatorname{sgn}\left(\frac{\partial y(i)}{\partial \delta}\right) = -\operatorname{sgn}(i)$. □

Proof of Proposition 3. This is proven in the text. □

Proof of Proposition 4. The result is immediate from $\frac{1}{2\delta}\frac{\partial y_1(i)}{\partial m} = -\alpha p + \beta(1-p-q)$. $\square$

Proof of Proposition 5. $\text{sgn}\Big(\frac{\partial y_1(i)}{\partial |i|}\Big) = \text{sgn}(i)\Big(1-2\delta\big(\alpha \sum_{r_2<y_1(i)} \Pr(r_2)+\beta \sum_{r_2>y_1(i)} \Pr(r_2)\big)\Big)$. If $\alpha \sum_{r_2<y_1(i)} \Pr(r_2) + \beta \sum_{r_2>y_1(i)} \Pr(r_2) \leq \frac{1}{2}$, then $\text{sgn}\left(\frac{\partial y_1(i)}{\partial i}\right) = \text{sgn}(i)$. If the reverse strict inequality holds, then:

$$\text{sgn}\left(\frac{\partial y_1(i)}{\partial |i|}\right) = \begin{cases} \text{sgn}(i) & \text{if } \delta \leq (2(\alpha \sum_{r_2<y_1(i)} \Pr(r_2) + \beta \sum_{r_2>y_1(i)} \Pr(r_2)))^{-1} \\ -\text{sgn}(i) & \text{if } \delta > (2(\alpha \sum_{r_2<y_1(i)} \Pr(r_2) + \beta \sum_{r_2>y_1(i)} \Pr(r_2)))^{-1}. \ \square \end{cases} \tag{24}$$

For a proposer with ideology i, let $y_1^-(i) < 0$ be a proposer's interior solution aligning a reactionary and centrist, and $y_1^+(i) > 0$ be an interior solution aligning a radical and centrist.

Lemma 2. A proposer with ideal point i is indifferent between a proposal $y_1^+(i) > 0$ aligning the centrist restrainer with the radical and a proposal $y_1^-(i) < 0$ aligning the centrist with the reactionary if and only if the centrist restrainer is indifferent between these proposals.

Proof. Letting $p \equiv \Pr(r_2 = -m)$ and $q = \Pr(r_2 = 0)$, define the payoff difference function:

$$Z(i, \alpha, m, p, q) \equiv (1-\delta)(u_i(y_1^+(i)) - u_i(y_1^-(i)) + \delta\left(V_i(y_1^+(i)) - V_i(y_1^-(i))\right). \tag{25}$$

$Z(i, \alpha, m, p, q)$ can be written $(y_1^+(i) - y_1^-(i))(y_1^+(i) + y_1^-(i))$, which has roots at $y_1^+(i) = y_1^-(i)$ and $y_1^+(i) = -y_1^-(i)$. In both cases, the centrist restrainer is indifferent between these proposals. However, we have $y_1^+(i) = y_1^-(i)$ only if $y_1^+(i) = y_1^-(i) = 0$, which implies $\alpha = \beta = \frac{1}{2}$. We have $y_1^+(i) - y_1^-(i) = 2(\beta - \alpha)\delta i \Pr(r_2 = 0)$, so $y_1^+(-e) = -y_1^-(-e) > 0$ only if $\alpha > \beta$. $\square$

Proof of Proposition 6. $Z(i, \alpha, m, p, q) = 0$ at $\alpha = \frac{1}{2}$ and at most one other value of $\alpha \in \left(\frac{1}{2}, 1\right]$, which solves $y_1^+(i) = -y_1^-(i)$. For $\alpha > \frac{1}{2}$, we have:

$$\begin{aligned} \varphi(\alpha, \delta, i, p, q) \equiv y_1^+(i) + y_1^-(i) &= (1-\delta)i + \delta\big(\alpha \sum_{r_2>0} \Pr(r_2) + \beta \sum_{r_2\leq 0} \Pr(r_2)\big)i \\ &+ \delta\big(\alpha \sum_{r_2\leq 0} \Pr(r_2)(2r_2 - i) + \beta \sum_{r_2>0} \Pr(r_2)(2r_2 - i)\big) \\ &+ (1-\delta)i + \delta\big(\alpha \sum_{r_2\geq 0} \Pr(r_2) + \beta \sum_{r_2<0} \Pr(r_2)\big)i \\ &+ \delta\big(\alpha \sum_{r_2<0} \Pr(r_2)(2r_2 - i) + \beta \sum_{r_2\geq 0} \Pr(r_2)(2r_2 - i)\big), \end{aligned} \tag{26}$$

where $p = \Pr(r_2 = -m)$ and $q = \Pr(r_2 = 0)$. Substitution yields:

$$\varphi(\alpha, \delta, i, p, q) = i(\delta(4\alpha - 8\alpha p + 4p - 4\alpha q + 2q - 4) + 2) - 4\delta m(\alpha + p - \alpha q + q - 1), \quad (27)$$

which is linear in δ and in α. For $\alpha \in \left(\frac{1}{2}, 1\right)$, $\varphi(\alpha, \delta, -e, p, q)$ strictly increases in δ. Finally, $\varphi(\alpha, \delta, -e, p, q)$ strictly increases in α only if: $\frac{1-p-q}{p} < \frac{e-m}{e+m}$.
1. Suppose $\frac{1-p-q}{p} < 1 - \frac{2m}{e}$. Since $1 - \frac{2m}{e} < \frac{e-m}{e+m}$, $\varphi(\alpha, \delta, -e, p, q)$ strictly increases in $\alpha \in \left(\frac{1}{2}, 1\right)$, and strictly increases in δ. Define:

$$\alpha^*(\delta) = \frac{e(\delta(2p + q - 2) + 1) + 2\delta m(p + q - 1)}{2\delta(e(2p + q - 1) + m(q - 1))} \quad (28)$$

Thus, $y_1^*(-e) > 0$ if and only if $\alpha > \alpha^*(\delta)$. The cut-off $\alpha^*(\delta)$ strictly *decreases* in δ, since:

$$\frac{\partial \alpha^*(\delta)}{\partial \delta} = -\frac{e}{2\delta^2(e(2p + q - 1) + m(q - 1))}, \quad (29)$$

and so $\frac{\partial \alpha^*(\delta)}{\partial \delta} < 0$ by $\frac{1-p-q}{p} < \frac{e-m}{e+m}$. We have $\alpha^*(\delta) < 1$ if and only if $\delta > \frac{e}{2ep+eq-2mp} \equiv \delta_1$, where $\delta_1 < 1$ by $\frac{1-p-q}{p} < 1 - \frac{2m}{e}$.
2. Consider $\frac{1-p-q}{p} \in \left[1 - \frac{2m}{e}, 1\right]$. Then $\varphi(\frac{1}{2}, 1, -e, p, q) \leq 0$ and $\varphi(1, 1, -e, p, q) \leq 0$. Since $\varphi(\alpha, \delta, -e, p, q)$ is linear in α and strictly increases in δ, $y_1^*(-e) < 0$ for all $\alpha \in \left(\frac{1}{2}, 1\right]$.
3. Consider $\frac{1-p-q}{p} > 1$. Since $\frac{e-m}{e+m} < 1$, $\varphi(\alpha, \delta, -e, p, q)$ falls in $\alpha \in \left(\frac{1}{2}, 1\right)$, and rises in δ. Thus, $y_1^*(-e) > 0$ if and only if $\alpha < \alpha^*(\delta)$. But, $\alpha^*(\delta)$ *increases* in δ, since $\frac{1-p-q}{p} > 1 > \frac{e-m}{e+m}$. Thus, $\alpha^*(\delta) > \frac{1}{2}$ if and only if $\delta > \frac{e}{e+m(1-2p-q)} \equiv \delta_2$, where $\delta_2 < 1$ by $\frac{1-p-q}{p} > 1$. □

Proof of Proposition 7. Let the global solution for agent i in $[-s_1, s_1]$ be $y_1^*(i)$. Suppose $y_1^*(-e) \geq 0$. If $y_1^*(e) \leq 0$, the claim is trivial. If $y_1^*(e) > 0$, then $y_1^*(e) = \min\{y_1^+(e), s_1\}$, and $y_1^*(-e) \geq 0$ implies $y_1^*(-e) = \min\{\max\{0, y_1^+(-e)\}, s_1\}$. Then, $y_1^*(e) \leq y_1^*(-e)$ if $\delta \geq (1 + (2\alpha - 1)(2(p + q) - 1))^{-1} \equiv \delta_1(\alpha, p, q)$.

Suppose $y_1^*(-e) \leq 0$. Then it suffices to show (1) $y_1^-(e) \leq y_1^-(-e)$ and (2) $y_1^*(e) = \min\{\max\{y_1^-(e), -s_1\}, 0\}$. (1) holds if $\delta \geq (1 + (2\alpha - 1)(2p - 1))^{-1} \equiv \delta_2(\alpha, p)$. To see (2), $\alpha > \frac{1}{2}$ implies $y_1^+(e) < y_1^-(e)$ and $y_1^+(-e) > y_1^-(-e)$. Suppose $y_1^-(-e) \geq 0$. Then, $y_1^+(-e) > y_1^-(-e) \geq 0$ implies $y_1^*(-e) = \min\{y_1^+(-e), s_1\} > 0$, contradicting $y_1^*(-e) \leq 0$. So, $y_1^*(-e) \leq 0$ implies $y_1^-(-e) < 0$ and $y_1^*(-e) = \max\{y_1^-(-e), -s_1\}$. Since $\alpha > \frac{1}{2}$ and $\delta \geq \delta_2$ implies $y_1^+(e) < y_1^-(e) \leq y_1^-(-e) < 0$, so $y_1^*(e) = \max\{y_1^-(e), -s_1\} \leq \max\{y_1^-(-e), -s_1\} = y_1^*(-e)$.

Proof of Proposition 8. We characterize the set of policies weakly preferred by a centrist restrainer over the status quo. Define $p \equiv \Pr(-m)$ and $1-p = \Pr(0)$. Define:

$$\psi(\alpha, p, m, \delta, s_1) \equiv 4\delta m^2 p \left(2\alpha + \alpha^2\delta p - 2\alpha\delta p + \delta p - 1\right) + 4\alpha\delta m p s_1 + s_1^2. \tag{30}$$

We show that if the restrainer at date 1 is always a progressive or centrist, a dynamically-sophisticated centrist restrainer prefers to the status quo s_1 any y_1 satisfying:

$$y_1 \in \left[\max\{-m, -s_1 - 4\alpha\delta p m\}, s_1\right]. \tag{31}$$

If, in addition, $\psi(\alpha, p, m, \delta, s_1) \geq 0$ and $-(1-\alpha)2\delta p m - \sqrt{\psi(\alpha, p, m, \delta, s_1)} < -m$, then a sophisticated centrist restrainer also prefers to s_1 any policy y_1 satisfying:

$$y_1 \in \left[-(1-\alpha)2\delta p m - \sqrt{\psi(\alpha, p, m, \delta, s_1)}, \min\{-m, -(1-\alpha)2\delta p m + \sqrt{\psi(\alpha, p, m, \delta, s_1)}\}\right]. \tag{32}$$

The payoff of a centrist restrainer from y_1 is $(1-\delta)u_0(y_1) + \delta V_0(y_1)$, where $u_0(y_1)$ is the date-1 payoff and $V_0(y_1)$ is the continuation payoff. So, the centrist's payoff from $y_1 = s_1$ is:

$$(1-\delta)u_0(s_1) + \delta\alpha\left(p u_0(-2m - s_1) + (1-p)u_0(-s_1)\right) + \delta\beta u_0(s_1). \tag{33}$$

Define $\Delta(y_1) \equiv (1-\delta)(u_0(y_1) - u_0(s_1)) + \delta(V_0(y_1) - V_0(s_1))$, which is the difference in a centrist's payoff from y_1 and her payoff from the status quo, s_1.
(i) The payoff to a centrist restrainer from $y_1 < -e$ is:

$$(1-\delta)u_0(y_1) + \delta\alpha u_0(-e) + \delta\beta\left(p u_0(\min\{-2m - y_1, e\}) + (1-p)u_0(e)\right). \tag{34}$$

Since $e > 2m + s_1$ and $y_1 < -e$, we have $-2m - y_1 > s_1$. Then, since $u_0(y_1) < u_0(s_1)$ and $V_1(y_1) < V_1(s_1)$ for any $y_1 < -e$, we have shown $\Delta(y_1) < 0$.
(ii) The payoff to a centrist restrainer from $y_1 \in [-e, -m]$ is:

$$(1-\delta)u_0(y_1) + \delta\alpha u_0(y_1) + \delta\beta\left(p u_0(\min\{-2m - y_1, e\}) + (1-p)u_0(-y_1)\right). \tag{35}$$

By inspection, we have $\Delta(y_1) < 0$ if $y_1 < -2m - s_1$. Consider, instead, $y_1 \geq -2m - s_1$. Then:

$$\Delta(y_1) = 4(2\alpha - 1)\delta m^2 p + 4\alpha\delta m p s_1 - 4(1-\alpha)\delta m p y_1 + s_1^2 - y_1^2, \tag{36}$$

which is strictly concave in y_1, and has roots $-2(1-\alpha)\delta pm \pm \sqrt{\psi(\alpha,p,m,\delta,s_1)}$. When $\psi(\alpha,p,m,\delta,s_1) > 0$ and $-2(1-\alpha)\delta pm - \sqrt{\psi(\alpha,p,m,\delta,s_1)} < -m$, $\Delta(y_1) \geq 0$ only if:

$$y_1 \in [-2(1-\alpha)\delta pm - \sqrt{\psi(\alpha,p,m,\delta,s_1)}, \min\{-m, -(1-\alpha)2\delta pm + \sqrt{\psi(\alpha,p,m,\delta,s_1)}\}]. \quad (37)$$

The second claim in the Lemma follows because $\Delta(-m) = (m+s_1)(m(4\alpha\delta p - 1) + s_1)$ is strictly positive if $\delta\alpha p > \frac{1}{4}\left(1 - \frac{s_1}{m}\right)$.

(iii) The payoff to a centrist restrainer from $y_1 \in [-m, 0]$ is:

$$(1-\delta)u_0(y_1) + \delta\alpha\left(pu_0(-2m - y_1) + (1-p)u_0(y_1)\right) + \delta\beta\left(pu_0(y_1) + (1-p)u_0(-y_1)\right). \quad (38)$$

Thus, $\Delta(y_1) = (s_1 - y_1)(4\alpha\delta mp + s_1 + y_1) \geq 0$ if and only if $y_1 \geq \max\{-s_1 - 4\alpha\delta mp, -m\}$.

(iv) The payoff to a centrist restrainer from $y_1 \in [0, s_1]$ is:

$$(1-\delta)u_0(y_1) + \delta\alpha\left(pu_0(-2m - y_1) + (1-p)u_0(-y_1)\right) + \delta\beta u_0(y_1). \quad (39)$$

We obtain $\Delta(y_1) = (s_1 - y_1)(4\alpha\delta mp + s_1 + y_1)$, which implies $\Delta(y_1) \geq 0$ since $y_1 \in [0, s_1]$.

(v) The payoff to a centrist restrainer from $y_1 \in [s_1, e - 2m]$ is:

$$(1-\delta)u_0(y_1) + \delta\alpha\left(pu_0(-2m - y_1) + (1-p)u_0(-y_1)\right) + \delta\beta u_0(y_1). \quad (40)$$

Thus, we obtain $\Delta(y_1) = (s_1 - y_1)(4\alpha\delta mp + s_1 + y_1) < 0$ for $y_1 \in [s_1, e - 2m]$.

(vi) Consider $y_1 \in (e - 2m, e]$. The payoff to a centrist restrainer from $y_1 \in (e - 2m, e]$ is:

$$(1-\delta)u_0(y_1) + \delta\alpha\left(pu_0(-e) + (1-p)u_0(-y_1)\right) + \delta\beta u_0(y_1). \quad (41)$$

By inspection, $u_0(y_1) < u_0(s_1)$ and $V_1(y_1) < V_1(s_1)$, so $\Delta(y_1) < 0$ for $y_1 \in (e - 2m, e]$.

(vii) The argument for $y_1 > e$ is similar to (i). □

Proof of Proposition 9. Define $q \equiv \Pr(r_2 = 0)$, $1 - q \equiv \Pr(r_2 = m)$, and

$$\phi(\alpha, q, m, \delta, s_1) = 4\delta m^2(1-q)\left(1 - 2\alpha + \alpha^2\delta(1-q)\right) - 4(1-\alpha)\delta m(1-q)s_1 + s_1^2. \quad (42)$$

Lemma 3. If the restrainer at date 1 is certain to be either a centrist or conservative ($p = 0$), then the dynamically-sophisticated centrist restrainer at date 1 prefers any policy

$$y_1 \in \begin{cases} [-s_1 + \delta 4\beta(1-q)m, s_1] & \text{if } s_1 \geq 2\delta\beta(1-q)m \\ [s_1, \min\{-s_1 + \delta 4\beta(1-q)m, m\}] & \text{if } s_1 \leq 2\delta\beta(1-q)m \end{cases} \tag{43}$$

over the status quo. If $\phi(\alpha, q, m, \delta, s_1) \geq 0$ and $\alpha 2\delta(1-q)m + \sqrt{\phi(\alpha, q, m, \delta, s_1)} > m$ then the proposer weakly prefers to s_1 any proposal

$$y_1 \in [\max\left\{m, \alpha 2\delta(1-q)m - \sqrt{\phi(\alpha, q, m, \delta, s_1)}\right\}, \alpha 2\delta(1-q)m + \sqrt{\phi(\alpha, q, m, \delta, s_1)}]. \tag{44}$$

Proof: A centrist's payoff from y_1 is $(1-\delta)u_0(y_1) + \delta V_0(y_1)$, where $u_0(y_1)$ is the date-1 payoff and $V_0(y_1)$ is the continuation payoff. So, the payoff to a centrist restrainer from $y_1 = s_1$ is:

$$(1-\delta)u_0(s_1) + \delta\alpha(qu_0(-s_1) + (1-q)u_0(s_1)) + \delta\beta\left(qu_0(s_1) + (1-q)u_0(2m - s_1)\right) \tag{45}$$

It is easy to show that a restrainer never prefers $y_1 < -s_1$ to s_1. So, we focus on the following cases: $y_1 \in [-s_1, 0]$, $y_1 \in (0, s)$, $y_1 \in [s, m]$, $y_1 \in (m, e]$, $y_1 > e$. For any such y_1, define:

$$\Delta(y_1) \equiv (1-\delta)(u_0(y_1) - u_0(s_1)) + \delta(V_0(y_1) - V_0(s_1)), \tag{46}$$

which is the difference in a centrist's payoff from policy y_1 rather than the status quo s_1.
(i) The payoff to a centrist restrainer from $y_1 \in [-s_1, 0]$ is:

$$(1-\delta)u_0(y_1) + \delta\alpha u_0(y_1) + \delta\beta\left(qu_0(-y_1) + (1-q)u_0(2m - y_1)\right). \tag{47}$$

We thus obtain $\Delta(y_1) = (s_1 - y_1)(s_1 + y_1 - 4\beta\delta m(1-q))$ which implies $\Delta(y_1) \geq 0$ if and only if $y_1 \geq -s_1 + \delta 4\beta(1-q)m$. This is consistent with $y_1 \leq 0$ if and only if $s_1 \geq \delta 4\beta(1-q)m$.
(ii) The payoff to a centrist restrainer from $y_1 \in [0, s_1]$ is:

$$(1-\delta)u_0(y_1) + \delta\alpha(qu_0(-y_1) + (1-q)u_0(y_1)) + \delta\beta(qu_0(y_1) + (1-q)u_0(2m - y_1)). \tag{48}$$

We thus obtain $\Delta(y_1) = (s_1 - y_1)(s_1 + y_1 - 4\beta\delta m(1-q))$ which implies $\Delta(y_1) \geq 0$ if and only if $y_1 \geq -s_1 + \delta 4\beta(1-q)m$. This is consistent with $y_1 \leq s_1$ only if $s_1 \geq \delta 2\beta(1-q)m$.

(iii) The payoff to a centrist restrainer from a policy $y_1 \in [s_1, m]$ is:

$$(1-\delta)u_0(y_1) + \delta\alpha(qu_0(-y_1) + (1-q)u_0(y_1)) + \delta\beta(qu_0(y_1) + (1-q)u_0(2m - y_1)). \tag{49}$$

We therefore obtain $\Delta(y_1) = (s_1 - y_1)(s_1 + y_1 - 4\beta\delta m(1-q))$ which implies $\Delta(y_1) \geq 0$ if and only if $y_1 \leq -s_1 + \delta 4\beta(1-q)m$. This is consistent with $y_1 \geq s_1$ if and only if $s_1 \leq \delta 2\beta(1-q)m$.
(iv) The payoff to a centrist restrainer from policy $y_1 \in [m, e]$ is:

$$(1-\delta)u_0(y_1) + \delta\alpha\left(qu_0(-y_1) + (1-q)u_0(2m - y_1)\right) + \beta u_0(y_1). \tag{50}$$

Thus, $\Delta(y_1) = 4(1-2\alpha)\delta m^2(1-q) - 4(1-\alpha)\delta m(1-q)s_1 + 4\alpha\delta m(1-q)y_1 + s_1^2 - y_1^2$, which is negativefor $y_1 > 2m + s_1$. Then, $\Delta(y_1) \geq 0$ if and only if $\phi(\alpha, q, m, \delta, s_1) \geq 0$ and

$$y_1 \in [\max\left\{m, \alpha 2\delta(1-q)m - \sqrt{\phi(\alpha, q, m, \delta, s_1)}\right\}, \alpha 2\delta(1-q)m + \sqrt{\phi(\alpha, q, m, \delta, s_1)}]. \tag{51}$$

(v) It is easy to show that a centrist restrainer strictly prefers s_1 to any policy $y_1 > e$. □

We now prove the proposition, starting with point (i). By Lemma 3, $s_1 \leq 2\beta m(1-q)$ implies that the restrainer weakly prefers l y_1 to s_1 only if $y_1 \geq s_1$. We next prove point (ii). If $s_1 > 2\delta\beta(1-q)m$ and $\delta \geq \frac{s_1}{m4(1-q)\beta} \equiv \delta_1(\beta, q, m, s_1)$, then a policy $y_1 \in [-e, s_1]$ is preferred by a centrist restrainer to the status quo only if $y_1 \in [0, s_1]$. Since $4(1-q)\beta > 1$ for $q < \frac{1}{2}$ and $\beta > \frac{1}{2}$, we have $\delta_1(\beta, q, m, s_1) < 1$. This step implies that for $\delta > \delta_1$, we have:

$$y_1^*(-e) \geq \min\{\max\{0, -e(1-\delta) + \delta e(\alpha - \beta)(2q-1) + 2\delta\beta(1-q)m\}, s_1\} \tag{52}$$

Thus, $y_1^*(-e) \geq s_1$ if $e\left(\delta(\alpha-\beta)(2q-1) + \delta - 1\right) \geq s_1 - \delta\beta 2(1-q)m > 0$ (by supposition). The LHS is positive if $\delta \geq (1 + (\alpha-\beta)(2q-1))^{-1} \equiv \delta_2$. Since $\alpha < \beta$ and $q < \frac{1}{2}$, $\delta_2 < 1$. So, for $\delta > \max\{\delta_1, \delta_2\}$, there exists $\underline{e}(\delta)$ such that $e \geq \underline{e}(\delta)$ implies $y^*(-e) \geq s_1$. □

FOR ONLINE PUBLICATION

10. Appendix C: Reversals With a Dynamically Sophisticated Centrist Restrainer

Proposition 7 showed that if the prospect of a radical-progressive pairing is quite likely, the radical initially proposes less reform than a reactionary. This result presumed a myopic centrist restrainer. We provide sufficient conditions for reversals with a sophisticated centrist restrainer, considering a setting in which the date-2 restrainer is always a progressive or a centrist.

Suppose a radical-progressive pairing is sufficiently likely that a dynamically-sophisticated centrist restrainer would accept some proposals $y_1 \in (-e, -m)$. On this interval, a radical wants to propose policy *as close* as possible to her ideal point. This is because there is no prospect of future reform: her initial proposal renders her mis-aligned with all future restrainers. In contrast, a date-1 reactionary who is sufficiently fearful of a radical-progressive pairing prefers to propose $-m$, since it is the closest policy to her ideal that ensures date-2 policy will move no further away from her ideal. Thus, a high prospect of a radical-progressive pairing may no longer imply that a radical adopts less initial reform than a reactionary.

A radical may still choose less initial reform than a reactionary. However, it requires that the radical be likely to hold proposal power, but not so likely as to trigger the above effects.

Proposition 10. Suppose the date-2 restrainer is a centrist or a progressive, where the progressive is more likely. If $\alpha \in \left(\frac{1}{2}, \frac{1}{2} + \frac{m}{2e}\right)$, and δ is sufficiently large, a reactionary proposes more initial reform than a radical.

Proof. First, we provide conditions on α and δ such that a radical proposes $y_1^*(-e) \in [\max\{-m, -s_1 - 4\alpha\delta pm\}, 0)$. We then show these conditions are sufficient for $y_1^*(e) \leq y_1^*(-e)$.
Step 1: *If $\alpha \in \left(\frac{1}{2}, \frac{1}{2} + \frac{m}{2e}\right)$, then for δ sufficiently close to 1, a radical strictly prefers an interior solution $y_1(-e) \in [\max\{-m, -s_1 + 4\delta\alpha pm\}, 0]$, to interior solutions $y_1(-e) \in (0, s_1]$ and*

$$y_1(-e) \in \left[-(1-\alpha)2\delta pm - \sqrt{\psi(\alpha, p, m, \delta, s_1)}, \min\{-m, -(1-\alpha)2\delta pm + \sqrt{\psi(\alpha, p, m, \delta, s_1)}\}\right].$$

Note that we do not claim the existence of these interior solutions. If $\alpha > \frac{1}{2}$, then $\psi(\alpha, p, m, \delta, s_1) > 0$ and by Proposition 8, $\alpha > \frac{1}{2}$ and $p > \frac{1}{2}$ and δ sufficiently close to 1 imply that a centrist restrainer would strictly prefer some policies on the interval $[-2m - s_1, -m)$ to the status quo, s_1. Suppose a date-1 radical proposer chooses an interior solution:

$$y_1(-e) \in \left[-(1-\alpha)2\delta pm - \sqrt{\psi}, \min\{-m, -(1-\alpha)2\delta pm + \sqrt{\psi}\}\right]. \tag{53}$$

The difference of a radical proposer's value from proposing an interior solution on this interval, and her value from proposing an interior solution $y_1(-e) \in [0, s_1]$ is:

$$4(2\alpha - 1)\delta \left(e^2(1-\delta) - 2(1-\delta)emp + m^2p(1-\delta p)\right), \tag{54}$$

which is strictly positive for $\delta = 1$. So, interior solution $y_1(-e) \in [0, s_1]$ is strictly dominated for a radical by interior solution $y_1(-e) \in [-2m - s_1, -m]$ when δ is sufficiently close to 1.

We now compare a radical proposer's payoff from an interior solution $y_1(-e) \in [\max\{-m, -s_1 - 4\delta\alpha mp\}, 0]$ to that from an interior solution on the interval in (53). The former is greater if:

$$4(2\alpha - 1)\delta p(m-e)(e(\delta(2\alpha - 2\alpha p + p - 2) + 1) + m(\delta p - 1)) \geq 0. \tag{55}$$

The LHS is strictly concave in α, with roots $\alpha = \frac{1}{2}$ and $\alpha = \frac{\delta e(2-p)+m(1-\delta p)-e}{2\delta e(1-p)} \equiv \bar{\alpha}(\delta, e, p, m) < 1$. If $\delta > \frac{e-m}{e-mp}$, then $\bar{\alpha}(\delta, e, p, m) > \frac{1}{2}$, and a radical strictly prefers interior solution $y_1(-e) \in [\max\{-m, -s_1 - 4\delta\alpha mp\}, 0]$ to interior solution $y_1(-e)$ on interval (53) if $\alpha < \bar{\alpha}(\delta, e, p, m)$. The threshold $\bar{\alpha}(\delta, e, p, m)$ strictly increases in δ, and satisfies $\bar{\alpha}(1, e, p, m) = \frac{1}{2} + \frac{m}{2e}$.

Step 2: *When $\alpha \in \left(\frac{1}{2}, \frac{1}{2} + \frac{m}{2e}\right)$, $p > \frac{1}{2}$ and δ is sufficiently close to one, an interior solution, $y_1(-e) \in [\max\{-m, -s_1 - 4\delta\alpha pm\}, 0]$ exists, and is a radical's globally optimum, $y_1^*(-e)$.* We show that for δ sufficiently close to one, the following conditions are satisfied:

$$\max\{-m, -s_1 - 4\delta\alpha pm\} \leq (1-\delta)(-e) + \delta e(\alpha - \beta)(2p-1) - 2\alpha\delta mp \leq 0. \tag{56}$$

where the middle expression is a radical's interior solution on $[\max\{-m, -s_1 - 4\delta\alpha pm\}, 0]$. This solution is strictly increasing in δ. For δ sufficiently large, the first inequality holds if $\max\{-m, -s_1 - 4\delta\alpha mp\} \neq -m$. Suppose, instead, $\max\{-m, -s_1 - 4\delta\alpha mp\} = -m$. Then, for the first inequality to be satisfied, for δ sufficiently large, we need $e(\alpha - \beta)(2p-1) - 2\alpha mp \geq -m$, which holds since $\alpha > \frac{1}{2}$ and $p > \frac{1}{2}$. So, we need only verify $e(\alpha - \beta)(2p-1) - 2\alpha mp < 0$

for (56) to hold for δ sufficiently close to one. Suppose, instead, that the inequality fails, i.e.,

$$2\alpha(e(2p-1)-mp) \geq e(2p-1). \tag{57}$$

Then $e(2p-1) \geq mp$. Inequality (57) is equivalent to $\alpha \geq \frac{e(2p-1)}{2(e(2p-1)-mp)}$. However, $\frac{e(2p-1)}{2(e(2p-1)-mp)} - \bar{\alpha}(1,e,m,p) = \frac{m(e(1-p)+mp)}{2e(e(2p-1)-mp)} > 0$. So, for δ sufficiently close to 1, $\alpha < \bar{\alpha}(1,e,m,p)$ rules out $\alpha \geq \frac{e(2p-1)}{2(e(2p-1)-mp)}$. Thus, an interior solution $y_1(-e) \in [\max\{-m, -s_1 - 4\delta\alpha mp\}, 0]$ exists. By the previous step, $\alpha < \bar{\alpha}(1,e,m,p)$ and δ sufficiently close to one imply that this interior solution is also the radical's global optimum.

Step 3: *When $\alpha \in \left(\frac{1}{2}, \frac{1}{2}+\frac{m}{2e}\right)$, $p > \frac{1}{2}$ and δ is sufficiently close to one, a reactionary proposer at date 1 makes a proposal satisfying $y_1^*(e) \leq y_1^*(-e)$.*

Suppose, first, $y_1^*(e) \in (0, s_1]$. Then $y_1^*(e) = \min\{e(1-\delta) - \delta e(\alpha-\beta) - 2\delta\alpha pm, s_1\}$, and for δ sufficiently close to one, $\alpha > \frac{1}{2}$ yields $y_1^*(e) < 0$, a contradiction. So, $y_1^*(e) \leq 0$. If $y_1^*(e) \leq \max\{-m, -s_1 - 4\delta\alpha pm\}$, the Proposition is correct, by Step 2. Suppose, instead, $y^*(e) \in (\max\{-m, -s_1 - 4\delta\alpha pm\}, 0]$. Then $y^*(e) = \max\{e(1-\delta) - \delta e(\alpha-\beta)(2p-1) - 2\delta\alpha pm, 0\}$, which yields $y_1^*(e) < 0$ for δ sufficiently close to one. Recalling the formula for $y_1^*(-e)$ from the previous step, we thus have $y_1^*(e) < y_1^*(-e)$ for δ sufficiently close to one. □

11. Appendix D: Reversals Without Risk Aversion

Proposition 7 provides conditions under which a date-1 reactionary proposes a policy that is closer to the radical's ideal policy than does the radical herself (a "reversal"). We now show that Proposition 7 does not depend on risk aversion, by proving the same result when agents incur linear policy losses. That risk aversion does not play a role follows from the fact that all possible policy outcomes at both dates lie on the same side of a proposer's ideal policy.

The sole change to the model is that we replace the quadratic disutility specification with linear loss: $u_i(y_t) = -|y_t - i|$. For simplicity, we assume that the restrainer at date 1 is certain to be a progressive (with probability p) or a centrist (with probability $1-p$). The characterization of date-2 policy outcomes as a function of the inherited status quo, $s_2 (= y_1)$ is unchanged since it follows from the symmetry of the date-2 restrainer's policy losses around her ideal point. We therefore focus on each proposer's date-1 proposal.

Proposition 11. Suppose that agents have linear policy losses. If, at date two, the radical proposer is relatively likely to hold power ($\alpha > \beta$) and the restrainer is likely to be a progressive ($p > \frac{1}{2}$) then there exists a $\delta^* \in \left(\frac{1}{2}, 1\right)$ such that: if $\delta \geq \delta^*$, a reactionary proposer successfully proposes more reform at date one than a radical proposer.

Proof. A radical proposer prefers s_1 to 0 if $\delta \geq \frac{1}{2\alpha} \equiv \delta_1(\alpha)$, and she prefers the policy s_1 to $-s_1$ if $\delta \geq \frac{1}{1+p(2\alpha-1)} \equiv \delta_2(\alpha, p)$, where $1 > \delta_2(\alpha, p) > \delta_1(\alpha)$ for $\alpha > \frac{1}{2}$. Finally, a radical proposer prefers the policy 0 to the policy $-s_1$ if $\delta \geq \frac{1}{2(1-p-\alpha+2p\alpha)} \equiv \delta_3(\alpha, p)$. So, $y_1^*(-e) = 0$ only if $\delta \geq \delta_3(\alpha, p)$, and $\delta \leq \delta_1(\alpha)$, which cannot hold since $\delta_1(\alpha) < \delta_3(\alpha, p)$. We conclude:

$$y_1^*(-e) = \begin{cases} s_1 & \text{if } \delta \geq \delta_2(\alpha, p) \\ -s_1 & \text{if } \delta < \delta_2(\alpha, p). \end{cases} \tag{58}$$

Similarly, we obtain the optimal date-1 proposal of a reactionary:

$$y_1^*(e) = \begin{cases} s_1 & \text{if } \delta \leq \delta_1(\alpha) \\ 0 & \text{if } \delta \in (\delta_1(\alpha), \delta_3(\alpha, p)] \\ -s_1 & \text{if } \delta > \delta_3(\alpha, p). \end{cases} \tag{59}$$

We conclude that when $\delta > \delta_2(\alpha, p)$, $y_1^*(e) < y_1^*(-e)$. □