CESifo WORKING PAPERS

The Role of Electoral Incentives for Policy Innovation: Evidence from the U.S. Welfare Reform

Andreas Bernecker, Pierre C. Boyer, Christina Gathmann

CESifo

The Role of Electoral Incentives for Policy Innovation: Evidence from the U.S. Welfare Reform

Abstract

We investigate whether the decision to experiment with novel policies is influenced by electoral incentives. Our empirical setting is the U.S. welfare reform in 1996, which marked the most dramatic shift in social policy since the New Deal. We find that electoral incentives matter: governors with strong electoral support are less likely to experiment with policies than governors with little electoral support. Yet, governors who cannot be reelected experiment more than governors striving for reelection. The importance of electoral incentives is robust to controlling for governor ideology, voter preferences for redistribution, the influence of the legislature, or for learning among states. A comparison of the role of governor ideology and electoral incentives reveals that both contribute about equally to policy experimentation.

JEL-Codes: D720, D780, H750.

Keywords: policy innovation, electoral incentives, welfare reform, spillovers.

Andreas Bernecker
University of Mannheim
Mannheim / Germany
a09728@mail.uni-mannheim.de

Pierre C. Boyer
CREST, École Polytechnique
Palaiseau / France
pierre.boyer@polytechnique.edu

Christina Gathmann*
Department of Economics
University of Heidelberg
Heidelberg / Germany
christina.gathmann@awi.uni-heidelberg.de

*corresponding author

1 Introduction

> *"The country needs and, unless I mistake its temper, the country demands bold, persistent experimentation. It is common sense to take a method and try it: If it fails, admit it frankly and try another. But above all, try something"*
>
> Franklin D. Roosevelt

Just as technological innovations are crucial for an economy's long-run performance, policy innovations are important for a country's adaptability to changing environments. In both cases, decision-makers face a lot of uncertainty when adopting innovative policies or technologies. Policy-makers, for instance, often do not know the benefits of the reform for their local jurisdiction or when these potential benefits will manifest. If little is known about the effectiveness of certain policies or technologies, there are potentially large gains from experimentation and possibly large costs of failure.

Indeed many successful policy innovations started out as experiments – like the deregulation of the airline industry or the reform of the telecommunications sector in the U.S.. And yet, actual policy-making does not seem to follow the simple rules suggested by the quote of Franklin Roosevelt above. On the contrary, the history of policy-making is full of examples of reforms *never* implemented; or policies that persist despite perceived weaknesses. Hence, a central question in political economy is the following: under which circumstances do politicians actually engage in experimentation? And why do policy experiments seem to occur less frequently than desirable?

Theoretical models of political agency show that electoral incentives may be one reason why politicians are hesitant to engage in experiments when both voters and politicians are uncertain about the mapping from policies to realized outcomes (see, e.g., Rogoff, 1990; Harrington, 1993; Majumdar and Mukand, 2004). If experimentation affects a politician's reelection chances, a conflict of interest between voters and politicians may arise. A popular incumbent with strong electoral support among voters, for instance, has a lot to lose in case a policy experiment turns out to be a failure. As such, she might be hesitant to engage in policy experimentation even if it is beneficial for the electorate.

Despite a rich theoretical literature, there is very little systematic empirical evidence

when new ideas actually transform into policies – and when they do not. Our paper aims to fill this gap. We study policy experimentation at the state level before and after the 1996 welfare reform, the most important shift in U.S. social policy since the New Deal.[1] The 1996 reform marked a radical turn in the provision of income support for the poor, which emphasized workfare rather than welfare. The reform decentralized most of the authority to design welfare policies from the federal government to the U.S. states. Even prior to 1996, states could apply for welfare waivers to experiment with new welfare rules under the old 'Aid to Families with Dependent Children' (AFDC) program. After the adoption of the 'Temporary Aid to Needy Families' (TANF) program in 1996, states were largely free to choose their own welfare rules with respect to eligibility, requirements while on welfare or the sanctions if recipients failed to comply with any requirements (see Harvey et al., 2000). Yet, many of the new policy options had never been tried before and state policy-makers had little prior knowledge whether and which novel policies would benefit their local jurisdiction. The U.S. welfare reform is thus an ideal setting to explore why (and when) state policy-makers engaged in policy experimentation.[2]

Our empirical results provide strong support for the view that reelection concerns have an important effect on the decision to experiment with waivers during the 'Aid to Families with Dependent Children' (AFDC) program and welfare rules during the 'Temporary Aid to Needy Families' (TANF) program. Strong electoral support for a governor reduces experimentation, especially among governors who may be reelected. The influence of reelection concerns on experimentation is economically sizable: An increase in electoral support by an interquartile range reduces experimentation among governors who can be reelected by 17%. Lame ducks with strong support among the electorate however, are actually more likely to experiment than governors who can be reelected.

Experimentation with redistributive polices might also be influenced by ideological differences between policy-makers. Republicans, who have long pushed for the idea of

---

[1]The reform in 1996 was preceded by many years of heated political and public debate on how to reform a welfare system that was perceived to be in crisis. For detailed accounts that trace the political and public debate leading up to reform at the federal level in 1996 see Reintsma (2007); Weaver (2000); Haskins (2006).

[2]The decision to decentralize reflects the idea of laboratory federalism that in a world of imperfect information, decentralization is beneficial to foster experimentation in different jurisdictions and to learn about the best policy implemented (see Oates, 1999, for a survey).

workfare rather than welfare, certainly played an important role in the federal Welfare Reform of 1996. While the ideology of the governor matters, electoral incentives continue to play an important role even conditional on governor ideology. Comparing the role of governor ideology and electoral incentives, we find that both contribute about equally to policy experimentation.

In contrast to governor ideology, we find little influence of voter ideology, the preferences for redistribution or the demand for welfare among the electorate. Our results are further robust to controlling for characteristics of the state legislature as measured by party composition, legislative polarization, or electoral competition. Politicians who have to decide whether to implement a policy might look to their neighbors for inspiration or engage in welfare competition instead. Defining "neighbors" in terms of geography, population size or ideology, we find that cross-state spillovers have few effects on the amount of experimentation in a state.

This paper proceeds as follows. The next section discusses the related literature. Section 3 describes our empirical setting, the U.S. welfare reform in which governors had the option to experiment with novel, previously unused policies. Section 4 introduces a political agency model to guide our empirical analysis. Section 5 introduces the data and empirical strategy. Section 6 reports the main results and assesses alternative explanations. The last section concludes.

2   Related Literature

A vast literature has investigated the consequences of the U.S. welfare reform for caseloads, employment, earnings, poverty, marriage patterns and many other socio-economic outcomes (see Blank, 2002; Grogger and Karoly, 2005; Moffitt, 2002, for surveys). Yet, the question why certain states adopted new welfare policies and experimented with unknown policy measures – like the rules that determine access to resources and the potential penalties in the case of non-compliance, for instance, – has received little attention so far.[3] Our article provides the first empirical test that reelection

---

[3]In political science, Lieberman and Shaw (2000) relate the choice of welfare rules to local conditions while Soss et al. (2001) investigate the relationship between a state's racial composition or ideology and

concerns are important to understand the dynamics of welfare policy-making in the states.

Our empirical analysis further speaks to the political science literature on policy innovations like the adoption of state lotteries, new taxes or labor market regulation (e.g. Walker, 1969; Berry and Berry, 1990, 1992).[4] Most of this literature does not test a particular channel encouraging or inhibiting policy experimentation. Our paper focuses on the role of reelection concerns as one mechanism why policy experiments might or might not be implemented.

The analysis in this paper also contributes to a growing literature on the effects of electoral incentives (see, e.g., Besley and Case, 1995a, 2003; List and Sturm, 2006; Besley, 2007; Alt et al., 2011; Ferraz and Finan, 2011; Gagliarducci and Nannicini, 2013).[5] Besley and Case (1995a), for instance, show evidence that reelection concerns indeed keep politicians in check: Public spending and taxes are lower in a governor's first term of office but increase when the governor faces a binding term limit. More recently, Ferraz and Finan (2011) demonstrate that electoral accountability also reduces corruption and rent-seeking among local politicians in Brazil. Our paper focuses on a potential downside of electoral incentives which has so far not been investigated. We show that term limits may benefit voters because politicians who cannot be reelected worry less about their reputation and hence are more likely to experiment with risky, but potentially beneficial policies.[6]

Our theoretical setup builds on a large literature on reputation concerns following the seminal work of Holmström (1982, 1999). An early application to the political arena is the political agency model by Rogoff (1990). Closer to us are models where voters and politicians are both uncertain about the mapping from chosen policies to realized outcomes (Harrington, 1993; Majumdar and Mukand, 2004; Fu and Li, 2014; Willems,

---

selected policy rules under TANF.

[4]Karch (2007); Berry and Berry (2007) provide good surveys of this literature.

[5]Electoral accountability may benefit voters through two channels. Elections allow voters to select those incumbents whose prior track record suggests they are of high ability. In addition, electoral accountability constrains the opportunistic behavior of incumbents.

[6]In related theoretical work, Smart and Sturm (2013) show that term limits may benefit the voter if some politicians are "public-spirited", i.e. have payoffs that coincide with those of the electorate. As in our model, electoral accountability and the value of future office may discourage public-spirited politicians from taking actions that are in the interests of voters because it reduces their re-election prospects.

2013).[7] As in our setting, politicians trade off the potential welfare and electoral gains when undertaking (and continuing) a policy reform against the reputation loss that a failure would deliver. The main focus in our paper is to bring a simple version of these models to the data.

Finally, our analysis speaks to the literature on policy persistence which tackles the question why some policies persist even if they are known to be a failure. Potential mechanisms stressed in the previous literature are individual-specific uncertainty about the winners and losers of a reform (Fernandez and Rodrik, 1991; Jain and Mukand, 2003; Ciccone, 2004), conflict between different groups about who will bear the costs of reform (Alesina and Drazen, 1991; Drazen and Grilli, 1993), political constraints on economic reform plans (Dewatripont and Roland, 1992), electorate's demand for reform (Prato and Wolton, 2017), or vested interests benefiting from the status quo (Coate and Morris, 1999). We focus on the reputation costs of policy reversals as one explanation why policies might persist; most importantly, we provide empirical evidence under what conditions policies persist during the U.S. welfare reform.

3 The U.S. Welfare Reform and Experimentation

3.1 The Old AFDC Program

Since the New Deal, the program 'Aid to Families with Dependent Children' (AFDC) provided financial assistance to needy children lacking parental care or support (see e.g. Grogger and Karoly, 2005; Moffitt, 2008). The program was jointly administered by the federal and state governments but the rules of the program were determined at the federal level.[8] States could only chose the level of monthly benefits but had otherwise little room

---

[7]Recent papers by Callander (2011a,b) and Callander and Hummel (2014) model policy experimentation with a richer policy space: The "world" there is not restricted to a two-state world and policies can take more values than the binary policy choice (reform or not) adopted in most of the earlier literature. Callander and Harstad (2015) present a model where heterogeneous districts choose both whether to experiment and the type of policy experiments. Reelection concerns of the policy makers are absent in their analysis. Instead, we study a setup with two states of the world and a binary decision in order to analyze the consequences of reelection concerns for policy experimentation and reversals.

[8]A separate program for jobless two-parent families (AFDC-UP) was established in 1961, but single-parent families remained the primary beneficiaries of the AFDC program prior to 1996.

to shape welfare policy.[9]

Starting in the early 1960s, state governments obtained more autonomy to experiment with welfare rules. Since 1962, Section 1115 of the Social Security Act allowed the Secretary of Health and Human Services to waive federal AFDC rules and regulations. States could petition the U.S. Department of Health and Human Services (DHHS) to change welfare rules in temporary pilot projects. Waivers were primarily used to implement policies that would reduce welfare dependency and encourage work participation. The most popular experiments during the waiver period were stricter work requirements, sanctions for benefit units failing to meet work requirements and family caps (eliminating or reducing the benefit increase if a child is conceived during welfare receipt). To apply for a waiver, state governments submitted detailed requests which rules and program elements they planned to modify, and what rules and regulations they wanted to implement instead.[10] The proposed waiver provisions were then reviewed, and sometimes altered, by the federal offices with jurisdiction over the various aspects of each state's proposal. The Secretary of Health and Human Services made the final decision whether to approve a waiver, request changes or deny it. In most cases, the waiver was approved as is or after some changes.

To measure policy experimentation during the AFDC period, we code whether a state applied for a waiver which allows to implement rules and procedures that deviated from the national program. Information on waiver applications in each state is available from Koerper (1996), Crouse (1999), and Lieberman and Shaw (2000). We count all waiver applications irrespective of whether the state planned to implement it only in specific counties or the state as a whole. To reduce any bias through the federal review process, we include all waiver applications that have been approved and implemented, but also those that were denied by the Secretary of Health and Human Services or withdrawn by the state. If a state did not apply for a waiver in a given year, the variable is set to zero.

Welfare waivers were rare until the late 1980s when AFDC caseloads approached

---

[9]For a family of three with no other income, for instance, monthly benefits in 1995 varied from $119 in Mississippi to $720 in Alaska. States could also implement a few eligibility rules like "fit parent" or "suitable home" provisions in order to limit payments to families with unsatisfactory behavior.

[10]Most waivers implemented multiple changes to welfare rules simultaneously (see Crouse, 1999; Koerper, 1996, for detailed descriptions of the state waivers).

almost 4 million families or around 11 million recipients nationwide. Waivers became more popular under President Reagan when federal aid for AFDC declined and government funds tightened during the 1989-1992 recession. Fifteen waivers were approved in 14 states under the Reagan administration and another 15 applications from 12 states under the Bush administration (see Harvey et al., 2000). As reform efforts accelerated during the first term of the Clinton administration, the federal government approved 83 waivers (see Figure 1 for the evolution of waiver activity between the late 1970s and 1996). In total, all but five states applied for one or more waivers with an average of 0.2 waivers per year. Figure 2 shows that waiver applications varied a lot across states ranging from 0 in Alaska to as many as 10 in Hawaii.

3.2 The 1996 Welfare Reform and State TANF Programs

The passage of the Personal Responsibility and Work Opportunity Reconciliation Act (PRWORA) in August 1996 abolished the old AFDC program and replaced it with the 'Temporary Aid to Needy Families' (TANF) program. The 1996 reform reshaped the landscape of U.S. welfare policy along at least three dimensions: First, the reform decentralized the authority to design welfare programs to the state level. State governments were now free to choose their own welfare rules, however generous or restrictive, bound only by some federal requirements.[11] As states had to decide on many new policy rules – like eligibility rules, work requirements, sanctions or time limits – policy rules have changed even after the adoption of TANF at the state level and do so until today.

Second, states not only obtained the right to choose but also the full fiscal responsibility of their choices. After 1996, federal funding changed from a matching grant into a single block grant. States who spent more than the federal block grant had to shoulder all the additional costs of more generous provisions or higher caseloads.[12]

---

[11]In practice, waiver provisions remained in place until the waiver expired or were replaced by new statewide TANF programs. State TANF programs replacing existing AFDC programs were implemented between September 1996 (Massachusetts, Michigan, and Vermont) and as late as January 1998 (California).

[12]The block grant for each state was fixed at the federal spending level in the 1992-1995 fiscal years. States were further required to contribute to other program components like subsidized child care. Under the old AFDC program, in contrast, a matching grant by the federal government co-financed state

The federal government further introduced financial penalties, in the form of percentage reductions in the block grant, on states with excessive caseloads or states that failed to comply with the federally mandated work participation rates of recipients. As a consequence, states had strong incentives to keep their spending levels and welfare caseloads in check to avoid any additional burden on the state budget. These incentives are also reflected in the actual numbers: Caseloads, for example, fell nationwide by more than 56% between 1994 and 2000.[13]

The third novelty of the welfare reform was the explicit goal to reduce welfare dependency and boost work participation. By setting time limits to benefit receipt, for example, TANF pushed the idea that government support was a temporary measure rather than a permanent source of income for poor families.[14] Further, work requirements became much stricter in the TANF legislation compared to AFDC, and focused on active participation in the labor market rather than training and education. Recipients who failed to comply with the new work requirements also faced much harsher sanctions compared to the old AFDC program.

To identify policy experiments during the TANF period is challenging as states had many policy dimensions to experiment with. We focus on four key policy innovations that were at the center of the public and academic debate during the reform process (see, e.g., Crouse, 1999; Fang and Keane, 2004; Grogger and Karoly, 2005): (1) Whether the state adopted a family cap, which limits the additional benefits if a child was conceived and born while the parent is on welfare. (2) The rules imposed for time limits on benefit receipt. (3) The type of work requirements adults had to fulfill in order to remain eligible for benefit receipt. (4) The sanctions imposed if a benefit unit did not fulfill the work requirements. We collect information on these welfare rules in each state and year from the Welfare

---

and local contributions such that the costs of rising caseloads were shared between federal and state governments.

[13]The literature suggests that the booming U.S. economy of the 1990s is responsible for about a third of the caseload reduction; the remainder is likely due to states' incentives to tighten access to welfare benefits (see, e.g., Council of Economic Advisors, 1999; Fang and Keane, 2004).

[14]The federal government set a lifetime time limit for benefit receipt of 60 months. States that decided to offer longer time limits (or no lifetime time limit) have to finance recipients beyond the federal time limit out of state funds. States could also decide to implement shorter lifetime time limits on benefit receipts. We discuss the specific welfare rules that are part of the empirical analysis in more detail in Section 5.

Rules Database of the Urban Institute which provides the most comprehensive and up-to-date information on state policy rules for the post-1996 period (Urban Institute, 2015). The data appendix and Table A1 provide more details about each policy rule analyzed.

It is important to note that for all four sets of rules, an innovation implied a more restrictive policy relative to the AFDC program. To illustrate how we measure an experiment or reversal, take the example of family caps which did not exist under the old AFDC program. Between 1996 and 2010, twenty-two states (among them, California, Florida and Illinois) adopted a family cap, while five states (among them, Illinois) later abolished it. We thus code a policy experiment as one if a state first adopts a family cap in any year after 1996 and zero if no family cap was adopted; we code a policy reversal if a state (that had adopted a family cap until 1996 or later) abolishes it in any year between 1997 and 2010. The coding for other welfare rules follows a very similar logic: A tightening of a policy rule in a specific year is coded as an experiment, abolishing that restriction in a later year is coded as a policy reversal.[15] Figure 3 shows that many policy experiments occur shortly after the PRWORA reform. The figure also demonstrates that policy reversals (the dashed line) are rare throughout the whole period. Despite a common focus on reducing welfare dependency and boost work participation, Figure 4 reveals that experimentation varied a lot across states during the TANF period – just like during the AFDC period.

3.3 Role of Governors

The era of U.S. welfare reform made state governments the central actors in shaping welfare policy. In this process, state governors played a key role in initiating welfare

---

[15]We can measure the adoption and abolishing of specific policy rules only when they are covered by the Welfare Rules Database of the Urban Institute. Changes in policy rules during the waiver period, for instance, are not captured systematically in the database because the waiver applications available are not specific enough to be merged to the Welfare Rules Database in a consistent manner. The coverage of rules is most complete and reliable after 1996. Hence, if a state both adopted and abolished a family cap prior to 1996, we will not capture this policy change in our measure of policy reversal. Anecdotal evidence however suggests that policy reversals were rare before 1996 as many waivers only expired after the federal TANF reform. Similarly, we might have a bias in our post-1996 policy experimentation measure as well. Suppose a state experiments with a policy rule, say a family cap, using a waiver, but did not have a family cap in 1996 (as coded in the Welfare Rules Database). If the waiver expired before 1996 and the state again adopted a family cap after 1996, we would code it as an experiment (even though the state has tried this policy rule during the waiver period). As a consequence, we will have some measurement error in our experimentation measure during the TANF period.

waivers, for instance. Wisconsin's governor Tommy Thompson, a leading figure in the welfare reform process, provides a good example. He made welfare reform a top priority and immediately created a task force to reform the existing AFDC system (Mead, 2004; Kaplan, 2000). The Thompson administration applied for three waivers in 1988, 1992 and 1993.[16] In 1993, Governor Thompson used his line item veto to implement the new Wisconsin Works, better known as W-2. With its emphasis on labor market participation, stricter work requirements and harsh sanctions for noncompliance, the W-2 program became a blueprint for the federal welfare reform in 1996. Governors in Delaware, Michigan, Minnesota, Nebraska and Ohio played similar defining roles for the speed and direction of welfare reform (see Weissert, 2000; Winston, 2002, for case studies).[17]

State governors, like all policy-makers, thereby faced a lot of uncertainty which welfare policies to choose. Even if a governor had a general sense of the direction of reform, she still had to decide which and how to implement many new rules. A governor who wanted to reduce welfare caseloads, for instance, could restrict eligibility, exclude certain groups, impose stricter employment requirements, define more generous earnings disregards or impose harsh sanctions upon non-compliance (or a combination thereof). And even after 1996, when governors could build on some evidence from waiver states, it was unclear whether a similar policy bundle would achieve the same results if implemented in a state with a different population structure, demand for welfare or economic situation.

4 Theoretical Framework

To guide our empirical analysis on the decision to experiment with (and possibly reverse) policies during the U.S. welfare reform, we develop a political agency model inspired by Majumdar and Mukand (2004). Our main goal is to highlight the conditions under which a politician engages in experimentation and their interactions with electoral incentives. To simplify the exposition, all proofs are relegated to the Appendix.

---

[16]In 1988, for example, Wisconsin was the first state to make transfer receipt of a household conditional on the school attendance of its teen children.

[17]State governors also played a prominent role in pushing and keeping federal welfare reform on the political agenda, mostly through lobbying of the National Governors' Association (NGA) (see Weaver, 2000). Fourteen governors testified in welfare hearings of the 104th Congress compared to only three state representatives (Winston, 2002).

4.1 Setup

Policies. Consider an elected politician who has an opportunity to enact a reform. The incumbent needs to decide whether to continue with the safe status quo policy a_S or experiment with the new policy a_N. Both policies affect overall welfare which is publicly observable. Different policies are appropriate for different environments; hence, the success of the new policy a_N is contingent on the underlying state of the world. The state of the world may be one of two types denoted by S and N, with a_S and a_N being the appropriate policies for the two environments respectively. If the underlying state of the world is S, then enacting the new policy a_N causes a net loss in welfare. The state N (resp. S) occurs with probability p (resp. $1 - p$).

The status quo policy a_S is assumed to deliver a fixed welfare, normalized to 0, independently of the state of the world. No benefits from the new policy occur in state S. In state N, the new policy initiative delivers a benefit Δ. However, this benefit occurs either in the short term or in the long term. The benefit realizes in the short term (resp. in the long run) with probability q (resp. $1 - q$).

The total cost of enacting and continuing the new policy initiative is c which is incurred independently of the success of the reform. This cost may represent the additional training of the bureaucracy to effectively administer the new policy or other costs of implementation. The total cost c consists of short- and long-run costs so that $\frac{c}{2}$ is incurred when the reform is implemented and again if the reform is maintained in the long term. If the reform is reverted back to a_S instead, then the long-term cost is not incurred. However, the short-term benefit cannot be kept in the long term if the reform is reversed.

We make the following assumptions:

$$p\Delta - c > 0, \tag{A1}$$

$$pq\Delta - \frac{c}{2}(1 + pq) > 0, \tag{A2}$$

and

$$\beta\Delta - \frac{c}{2} < 0, \tag{A3}$$

where $\beta = \frac{p(1-q)}{1-pq} < p$ is the probability that the state of the world is N conditional on a short-term failure of the reform.

Assumption $(A1)$ implies that a reform that is continued has positive expected welfare gains ex-ante. In our setting, this assumption can be justified on three grounds: first, there was a widespread perception among policy-makers in the late 1980s and early 1990s that the old AFDC program was too permissive and needed reform (see e.g. Haskins, 2006; Reintsma, 2007). Second, decentralizing decision rights should improve average welfare if local policies provide a better fit to local needs (compared to a common policy for all). Finally, even if the experiment tightened rules and thus reduced recipients' welfare, average welfare in a state might still increase if financial resources freed up from TANF are used to benefit the average inhabitant. Assumption $(A2)$ implies that undertaking the reform and reverting back to the status quo in the face of short-term failure has positive expected benefits ex-ante. Finally, assumption $(A3)$ implies that, if the benefit does not realize in the short term, the expected future benefit from continuing with the reform is negative.[18]

Timing. The game has three periods, $T = \{0, 1, 2\}$. At the beginning of the game $(T = 0)$, the incumbent politician has the opportunity to change the existing policy a_s. She receives a private signal about the appropriateness of the policies, i.e. the state of the world, and then faces the choice of either maintaining the status quo policy a_s or enacting the new policy a_N. If she decides to maintain the status quo, then the ex-post welfare is 0. On the other hand, if the new policy is enacted, a cost $\frac{c}{2}$ is initially incurred. Then, the politician learns about the realization of short-term benefit from the reform at the beginning of $T = 1$. If the new policy turns out to be successful, the politician knows that the state is N, and therefore a_N is the appropriate policy. This implies that the benefit Δ is secured by continuing the reform and incurring the long-term cost $\frac{c}{2}$ at the beginning of period $T = 2$. However, in the case where the benefit does not realize in the short term, it is unclear whether the failure is because of bad luck or because the policy is not appropriate. The politician now has to decide whether to continue with the policy

---

[18]We implicitly assume that the parameters of the model take values such that assumptions $(A1)$, $(A2)$, and $(A3)$ hold simultaneously. This is the case for a large range of parameter values.

initiative a_N or to revert back to the initial status quo a_S. If she reverts back to a_S, ex-post welfare is $-\frac{c}{2}$. If she continues with a_N, the society again incurs a cost $\frac{c}{2}$ and gets the long-run benefit only if the state of the world is N. The ex-post welfare is observed before the next election that takes place at the end of period $T = 2$, corresponding to the end of the incumbent's term in office.

Politicians. The economy has many politicians that differ in their capability to assess the appropriateness of policies for the economy, and can be either of high quality or of low quality. The incumbent politician knows her type. For simplicity, we assume that a high-quality incumbent knows the state of the world. The high-quality politicians serves as a benchmark to compare the decisions taken by different politicians depending on their electoral incentives and welfare considerations. On the other hand, a low-quality politician does not know anything ex ante, i.e. she only knows that the state of the world N is realized with probability p, and that in this state the benefit realizes in period 1 (resp. period 2) with probability q (resp. $1 - q$).[19]

The incumbent politician cares about both welfare of her constituency as well as her own future electoral prospects. The incumbent's objective function is

$$\gamma(\text{Welfare}) + (1 - \gamma)(\text{Probability of Reelection}), \tag{1}$$

where $\gamma \in [0, 1]$ is the relative weight that she puts on welfare.

Finally, at the beginning of the game ($T = 0$) the incumbent has earned a reputation among the electorate represented by $\lambda \in [0, 1]$. We assume that this reputation does not depend on future decisions taken by the incumbent and remains constant until the next election.[20]

Voters and Reelection Rule. The electorate consists of a representative voter that chooses either to reelect or throw the incumbent out of office at the end of the term $T = 2$. The voter cares only about ex-post welfare. We assume that the voter does not

---

[19]We discuss below a version of the model where politicians have ideological bias in favor or against the reform.

[20]See Majumdar and Mukand (2004) for an analysis where the reputation of the incumbent changes with the decision to undertake reforms.

know that there are high- or low-quality politicians so that he does not think that the incumbent and the (large number of) potential challengers differ in competence. However, the voter may take into account the reputation of the politician when deciding to reelect the incumbent. Assuming the representative voter does not know that politicians differ in abilities simplifies the analysis considerably. Under this assumption, there is no signaling game in the policy choice of the incumbent: Both the challenger and the incumbent are perceived to be identical by the voter except in the case where the incumbent's reputation is positive.[21]

We assume that the representative voter follows a simple reelection rule: Reelect the incumbent with probability 1 if ex-post welfare is positive; never reelect the incumbent if ex-post welfare is negative; and reelect the incumbent with probability λ if ex-post welfare is 0. This reelection rule is in the spirit of agency models where voters punish the incumbent for bad performance (see Persson and Tabellini, 2000).

4.2   Results

4.2.1   Policy Choices of a High-quality Politician

We first analyze the decision of a high-quality incumbent which serves as a benchmark for our analysis.

Proposition 1. *For any relative welfare weight γ and any reputation λ, a high-quality incumbent always undertakes a reform if and only if the underlying state of the world calls for a reform (state N). In addition, a reform undertaken is never reverted back to the status quo.*

A high-quality politician knows the state of the world. If the state of the world is S, she will never experiment because there are neither welfare nor electoral gains from a reform. Proposition 1 shows that a high-quality incumbent has her objective fully aligned with the interest of the representative voter: There is no conflict of interest between maximizing voter welfare and her reelection chances.

---

[21]The signaling game where the representative voter makes inferences on quality of the incumbent based on the sequence of policy choices and the realized outcomes is analyzed in Majumdar and Mukand (2004). The predictions they derive are similar to the ones we obtain in the simplified model presented here.

4.2.2 Policy Choices of a Low-quality Politician

We now show that the decision-making of a low-quality incumbent may exhibit a conflict between maximizing welfare and reelection chances. In order to disentangle the different forces, we first analyze two polar cases: The one of a purely welfare maximizing politician ($\gamma = 1$) and the one of a purely office-motivated politician ($\gamma = 0$).

Welfare-maximizing Politician. The following proposition presents the case where the policy choices of the incumbent are driven only by welfare considerations.

Proposition 2. *A low-quality incumbent who only cares about welfare ($\gamma = 1$) always implements the reform, and always reverts back to the status quo if the reform does not deliver the short-term benefit.*

The decision-making policy described in Proposition 2 is the optimal one from a welfare perspective under the veil of ignorance. For a politician who does not know the state of the world, the reform is welfare maximizing from an ex-ante perspective. However, it is welfare-maximizing to revert back to a_S if the reform fails in the short run as the expected future benefit from the reform is then negative.

Office-motivated Politician. The second case presents the policy choice of a purely office-motivated politician.

Proposition 3. *A low-quality incumbent who is purely office-motivated ($\gamma = 0$) implements the reform if and only if*

$$p \geq \lambda, \tag{2}$$

and never reverts back to the status quo when the reform is undertaken.

Proposition 3 shows that it is the reelection motive of the low-quality politician that is responsible for her choices. A low-quality politician might not maximize voter welfare for two reasons. First, equation (2) implies that the politician's initial reputation can conflict with the decision to experiment with beneficial reforms. An uninformed politician

15

with a high enough initial reputation will not experiment even though it would maximize ex-ante welfare (see Proposition 2). As a consequence, reputation concerns result in too little policy experimentation from a welfare perspective. Second, in the face of short-term failure a purely office-motivated politician does not take into account the expected welfare loss of continuing with the reform and "gambles" for reelection with the welfare of the electorate. This happens because a policy reversal back to the status quo implies that the politician would lose the election for sure, whereas if the politician persists with the reform, she can win the election if the long-term benefit realizes (which happens with probability $\beta > 0$).

Intermediate Politician. We now turn to the general case where the low-quality politician is concerned with both welfare and her reelection chances ($\gamma \in [0, 1]$). We first focus on the decision to revert an experiment.

Proposition 4. *A low-quality incumbent reverts back to the status quo in face of a short-term failure of the reform if and only if $\gamma \geq \gamma^*$, where*

$$\gamma^* := \frac{1}{1 + \frac{c}{2\beta} - \Delta}. \tag{3}$$

A low-quality politician who observes the short-term failure of a reform does not know whether the reform will be successful in the long term. This effect pushes her to revert back to the status quo to increase expected welfare (see Proposition 2). However, reelection concerns pull her towards sticking with the reform as in Proposition 3. Proposition 4 disentangles these two conflicting forces: if the relative welfare weights is above the threshold γ^* defined in Equation (3), the low-quality incumbent reverses the reform as welfare considerations outweigh reelection considerations.

Corollary 1. *A low-quality incumbent is more likely to revert in the face of short-term failure of the reform when γ increases and Δ decreases.*

While high-quality politicians never revert an experiment (Proposition 1), Corollary 1 shows that low-quality politicians make their reversal decision dependent on the welfare gains and costs of the experiment.

We now analyze whether an uninformed politician who cares about both welfare and reelection ($\gamma \in [0,1]$) undertakes the reform.

Proposition 5. *(I.) When $\gamma \geq \gamma^*$, then a low-quality incumbent undertakes a reform if and only if*

$$\gamma \left(pq\Delta - \frac{c}{2}(1 + pq) \right) + (1 - \gamma)pq \geq (1 - \gamma)\lambda. \tag{4}$$

(II.) When $\gamma < \gamma^*$, then a low-quality incumbent undertakes a reform if and only if

$$\gamma \left(p\Delta - c \right) + (1 - \gamma)p \geq (1 - \gamma)\lambda. \tag{5}$$

Proposition 5 follows from a comparison of the politician's payoff if she undertakes the reform (for the two cases where she will or will not revert back to the status quo in the face of short-term failure) with the payoff from sticking with the status quo and running the next election on her initial reputation instead. The proposition states that a low-quality incumbent compares the possible welfare and reelection gains that experimentation will deliver (taking into account her optimal reversal decision) to her current chances of reelection determined by her initial reputation.

Corollary 2. *A low-quality incumbent is more likely to experiment when Δ increases, γ increases, and λ decreases.*

Corollary 2 states that low-quality politicians experiment more when the potential welfare gains are high, when they care more about welfare (and less about reelection) and when the initial reputation is low (because then the incumbent has little to lose in terms of her reelection chances).

Ideology and Experimentation. So far there is no role for ideology in the model as politicians agree on the decision to undertake the reform given the state of the world. We now extend the model to allow for some ideological differences between politicians. In particular, we assume that Republicans assign a higher probability that the reform is the right thing to do, i.e. that the state of the world calls for a reform. This assumption captures the common perception of the welfare reform, and certainly the

federal PRWORA reform, that it was fueled by a conservative agenda to curb welfare spending. To capture that, we assume that Republicans who do not observe the state of the world believe that the state of the world N (where an experiment is welfare-enhancing) occurs with probability $p_R > p$.[22] In contrast, Democrats who do not observe the state of the world continue to have the belief that the state of the world is N with probability p.[23] Since the representative voter cares only about ex-post welfare, the reelection rule remains unaffected by the bias of the politicians. In this new environment, Republicans are more likely to experiment with the reform (which increases expected welfare from Proposition 2) and to persist in the face of short-term failure (which decreases expected welfare from Propositions 3 and 5) relative to a Democratic governor.[24]

4.3   From Theory to Data

In the empirical analysis below, we assume that all governors do not know the state of the world (i.e. are low-quality in the setting of the model). We later relax that assumption allowing some governors to be better able to predict whether an experiment was needed. Using proxies for a governor's ability – like her educational qualifications or prior political experience – we find that that this heterogeneity matters but has little impact on our main results.

Our model then highlights several channels how electoral incentives and ideology may influence policy experimentation. The first channel is related to the initial electoral support: politicians with higher initial electoral support among the electorate are less likely to experiment. This is particulary the case if politicians are purely office-motivated. Experiments are risky and make incumbents who seek reelection hesitant to risk their electoral support in order to pursue risky experiments that might fail.

The second channel is that the influence of electoral incentives should be weaker for politicians with a lower weight on reelection. Politicians who are more concerned

---

[22]Democrats and Republicans could also differ in the assessment of Δ even if they agree on p. In this case, the relevant comparative statics are those in Corollaries 1 and 2.

[23]The analysis yields the same prediction if Democrats underestimate the probability of the state being N relative to Republicans instead.

[24]To illustrate, take the example where $\gamma = 0$. Then, from Proposition 3, a politician experiments whenever $p \geq \lambda$. Since λ is not correlated with the party of the governor in the model, the inequality is more likely to be satisfied for a Republican. As a consequence, she experiments more.

about welfare are more likely to experiment than politicians who care mostly about their own reelection. A third channel is that the potential gains from a reform encourages experimentation. Finally, if Republicans are more likely to believe that an experiment is the right thing to do in the current state of the world, they should experiment more with the reform than Democrats.

The model also speaks to the likelihood of observing a policy reversal. The initial electoral support of the politician should have no effect on the likelihood of a reversal. The reason is that a politician who reverts back to the status quo is voted out of office for sure. A politician who does not revert is either successful in the long run (in which case she is reelected for sure) or is not successful (in which case she is voted out of office). In all of these cases, the initial electoral support has no impact on reelection chances and hence, does not matter for the decision to revert an experiment. A politician is less likely to revert if she cares more about reelection and the potential welfare gains from an experiment are high. Finally, Republicans are more likely to persist in the face of short-term failure which decreases expected welfare relative to Democrats.

5 Data Sources and Estimation Approach

5.1 Data Sources

For the empirical analysis, we require measures for the electoral support of a governor (an empirical counterpart to λ), the importance of reelection concerns (a proxy for $1-\gamma$) and a measure of the potential gains from experimentation (an empirical counterpart to Δ). All of these concepts are difficult to measure as they are ultimately unobservable to the researcher.

We use the vote margin in the past gubernatorial race to proxy the governor's electoral support. Given the incumbency effect, governors who were elected with a larger majority are more likely to be reelected conditional on the state of the world and the type of challenger. To rule out that the past vote margin just captures electoral competition in gubernatorial races, we control for the competitiveness of state elections, for example, through the Holbrook-Van Dunk index (Holbrook and Van Dunk, 1993). To measure

reelection concerns among governors, we use information on whether a state governor faces a binding term limit in her current term in office, i.e. whether she is a lame duck. Lame ducks might still have some electoral incentives, for example, because they plan to run for another office at a higher level. All we need for our analysis is that lame ducks care somewhat less about their electoral prospects than governors who can and want to be remain in office.

Further, we require a proxy for the potential gains of a policy experiment which is difficult to measure. An important motivation for state governors to apply for a waiver prior to 1996 was to reduce caseloads and hence, welfare spending in their state. The incentive to curb spending became even stronger during the TANF period because states now have to cover all welfare expenditures exceeding the federal block grant (see Section 3.2). We thus use a state's AFDC/TANF spending to proxy the potential gains from policy experimentation. The idea here is that the block grant encourages politicians with growing welfare expenditures to experiment with rules that effectively reduce the state's welfare burden and free up resources for other public goods and services. Below, we show that the potential savings through lower welfare spending is an important motor for experimentation whose influence is very robust to including the demand for welfare, the ideology or redistributive preferences of its electorate.

We use the governor's party and a measure of governor ideology (based on Berry et al. (1998)) to control for the ideology of state governors. The latter measure takes into account that governors belonging to the same party might have different ideologies. The ideology measure varies between 0 and 100 where larger values represent a more liberal attitude. We further collected and coded biographical information (experience in prior political offices and educational background) for each governor.

To control for other potential influences on the decision to experiment and reverse a policy, we collect information on the political and socio-economic characteristics in each state. Our basic set of state-level controls include the state unemployment rate, population and the share of the population above 65 and between 5 and 17 as well as income per capita. In additional regressions, we check whether additional state characteristics like the racial composition, the share of immigrants, the share of unmarried

births and income inequality (measured as household income at the 90th over the 10th percentile) are important to explain policy experimentation.

The decision to experiment might also be influenced by voter preferences or the demand for welfare in the state population. To account for voter preferences, we use Berry's measure of citizen ideology as well as the voter support for the Democratic candidate in the last Presidential election as proxy for state ideology. In addition, we employ individual-level information on preferences for redistribution and preferences over welfare spending from the General Social Survey to proxy for the state-level preferences for redistribution. To account for the legislative influence on policy innovation, we control for party composition and political polarization. Polarization is calculated as the difference between the DW-NOMINATE scores for each party where the score for each party is measured by the state's representatives in the House or Senate (Poole and Rosenthal, 1984; 1997). We further code whether the government in the state is politically divided between Democrats and Republicans. States might also imitate or learn from other states. To analyze these cross-state effects, we define geographic neighbors (states sharing a boundary); states similar in population size (based on their population in 1978) and ideological neighbors in the same Census region (with a similar voting record in the 1978 presidential election). Our measure of cross-state spillovers is then the average number of experiments undertaken in neighboring states where neighbors are defined by a share boundary, population size or ideology.

Table 1 contains summary statistics of the variables for the waiver and the TANF period. The online appendix contains additional details on the source and coding of our dependent and independent variables.

5.2   Estimation Approach

To assess the role of electoral incentives for policy experimentation, we use a difference-in-differences approach. To implement our main empirical tests, we estimate variants of the following model:

$$Exp_{st} = \alpha_1 ElecSup_{st} + \pi_1 LD_{st} * ElecSup_{st} + \phi_1 LD_{st} + \mu_1 Gain_{st} + \gamma_1 * X_{st} + t_t + \theta_s * \tau_t + \varepsilon_{1st} \quad (6)$$

where Exp_{st} denotes our measure of policy experimentation which we pool over the AFDC period (1978-1996) and the TANF period (1996-2008). $ElecSup_{st}$ measures the governor's electoral support, LD_{st} is an indicator whether the governor faces a binding term limit and $Gain_{st}$ measures the potential gains from experimentation. θ_s denotes state fixed effects, t_t year fixed effects and $\theta_s * \tau_t$ are state-specific linear trends. We further control for the following state characteristics X_{st}: state unemployment rate, personal income per capita, state population and age structure of the population.

We expect policy experimentation to decrease with the electoral support for the governor, i.e. $\alpha_1 < 0$. A governor with high initial electoral support will be reluctant to engage in risky policy experiments, while a governor with little electoral support is more likely to experiment - in the hope to boost her reelection chances. In addition, we expect the electoral incentives to be weaker for governors who cannot be reelected and hence worry less about the risk of policy experimentation, so $\pi_1 > 0$. In addition, we expect that the likelihood of an experiment increases with the potential gains from experimentation, i.e. $\mu_1 > 0$.

A potential concern with the specification in equation (6) is reverse causality with respect to electoral support and spending. Empirically, we deal with this concern in two ways. Our first approach is to use the vote margin in the past election and lagged spending on the welfare program as regressors. Lags help to cope with contemporaneous shocks that might influence both the decision to experiment and program spending, for instance. Lags would however, not help with persistent shocks or systematic omitted variables that affect both sides of the equation. In the absence of plausible instruments, we augment (6) with a lagged dependent variable which controls for persistent factors driving policy experimentation, and estimate this model using an Arellano-Bond panel estimator (Arellano and Bond, 1991).[25]

---

[25]An alternative method to address endogeneity would be to employ an approach similar to Lee et al. (2004) who used a sample of close elections (say in 1992) which generates quasi-random variation in the party of the governor elected into office. The basic idea is then to compare the behavior of a governor who was just elected into office in 1992 and re-elected in 1996 to the behavior of a candidate who won the seat in 1996 from an incumbent of the opposite party who in turn barely won the election in 1992. We would expect that the first governor has more electoral support than the second governor because of the incumbency advantage. Unfortunately, such a sample (using a vote margin below 2 percentage points to define close elections) has less than 100 observations which renders the use of a RDD to identify the role of electoral support on policy experimentation impossible.

Furthermore, we would argue that any endogeneity in our key independent variables, electoral support and welfare spending, should bias our estimates toward zero rather than inflate the true effects. The model suggests that higher AFDC or TANF spending should raise the likelihood of experimentation. In the case of reverse causality, in turn, we would expect that policy experimentation today, by tightening sanctions, work requirements and eligibility criteria, for instance, reduces future welfare spending because of declining caseloads. Similarly, if a governor campaigns on policy experimentation, the decision to experiment with welfare rules would boost her future vote margin. Our model in contrast predicts that the past vote margin reduces experimentation because of the potential electoral costs of failure. In both cases, the coefficients on welfare spending and electoral support are likely to be conservative estimates of the true effect.

Finally, we study the determinants of policy reversals using the following model:

$$Reversal_{st} = \mu_2 Gains_{st} + \phi_2 LD_{st} + \gamma_2 * X_{st} + t_t + \theta_s + \theta_s * \tau_t + \varepsilon_{2st}, \tag{7}$$

where all variables are defined as before. The higher the gains from experimentation and the more concerned the governor is about reelection, the less likely a politician will revert a policy; so, $\mu_2 < 0$ and $\phi_2 > 0$.

6 Empirical Results

6.1 Electoral Concerns and Policy Experimentation

We start out with assessing the role of electoral incentives for policy experimentation according to (6). The dependent variable in Table 2 is the number of policy experiments in a state and year during the 1978-2008 period.[26] All specifications control for state and year fixed effects as well as state characteristics and a governor's age. Even columns also add state-specific trends in order to control for upward or downward trends in experimentation over time. Columns (1) and (2) of Table 2 document that governors

---

[26]We find similar though statistically slightly weaker effects, if we use a dummy variable for experimentation (rather than a count variable) or if we estimate the baseline separately for the waiver period (1978-1996) and the TANF period (1996-2008) instead.

with high electoral support indeed experiment less than governors with less electoral support. Our model provides one explanation for this relationship: Politicians, especially if they are purely office-motivated, are hesitant to reform because they weigh the risky electoral gains from an experiment against their safe initial electoral support.

We next investigate whether the influence of electoral support on experimentation is muted for governors who cannot be reelected. Columns (3) and (4) add whether a governor is a lame duck and its interaction with past electoral support to the specification. The estimates show that electoral support still has a negative effect on experimentation among governors who can be reelected; governors who face a term limit, in turn, are actually more likely to experiment. The effect is economically sizable. An increase in the past vote margin by an interquartile range reduces policy experiments among governors who can stand for reelection by 17%.[27] For lame ducks in contrast, the sum of the coefficient on the vote margin and the interaction effect of vote margin and lame duck turns positive. Hence, lame ducks are more likely to engage in risky experimentation compared to governors who can be reelected.

The last two columns of Table 2 show that adding the potential gains from experimentation does not affect the coefficients on electoral incentives. As expected, higher potential gains (as measured by past AFDC/TANF spending) are associated with more experimentation. An increase in welfare spending by an interquartile range increases experimentation by 64.9%.[28]

As some of our independent variables, like the past vote margin or past welfare spending, might be influenced by past experimentation, we also estimate an augmented model with lagged dependent variable to absorb persistence in policy experimentation. The results for the Arellano-Bond estimator shown in Appendix Table A2 are surprisingly similar to the baseline estimates in Table 2. We still find that electoral support reduces experimentation while the opposite is true for governors who cannot be reelected. Furthermore, higher past welfare spending encourages experimentation even conditional

---

[27]The interquartile range (75th-25th percentile) of the past vote margin in our time period is 10.8 percentage points, while the mean number of policy experiments per year is 0.283. Hence, (-0.0076*10.8)/0.283=0.172.

[28]The interquartile range of real welfare expenditures is 505.76 (US$ million). Hence, we get 0.0005*505.76/0.283=0.649.

on the lagged dependent variable.[29]

We next explore in Table 3 the robustness of our empirical estimates to alternative specifications of the key independent variables. One concern with our empirical model in 6 is that electoral support in the past election just proxies for the competitiveness of state elections. Weaker competition, i.e. large margins of victory, rather than the electoral incentives of the governor would then be the factor reducing experimentation. To assess this argument, we add two variables to measure electoral competition: whether the governor has changed in the past two electoral cycles (in column (1) of Table 3); the Holbrook and Van Dunk index of competitiveness (in column (2) of Table 3); and both measures simultaneously (in column (3) of Table 3). The estimates in columns (1)-(3) show no statistically significant influence of either measure of competition on policy experimentation. Even more importantly, the coefficients on the past vote margin remain unchanged compared to the baseline and statistically significant. Hence, the effect of the past vote margin cannot be explained by the fact that swing states or states with tougher electoral competition experiment less. Instead, policy experiments in a state are less likely because the current governor has more to lose (in terms of reelection chances) if her electoral support is high and the experiment turns out to be a failure.

A related concern is that the past vote margin might affect policy experimentation in a nonlinear fashion; governors who just barely won an election might experiment more while governors with a comfortable margin of victory might experiment less, for instance. To test for such nonlinearities, we add in column (4) of Table 3 a quadratic term of the vote margin to the specification. The results show that it is only the linear effect that matters. The quadratic terms, in turn, are small and not statistically significant. One might also worry that the vote margin in the past election is endogenous because it measures the outcome of party platforms, candidate positions and voter turnout. To mitigate the concern of endogeneity, we replace the past vote margin with the average vote margin of a governor over this whole career as governor (in column (5) of Table 3). The coefficient on the average vote margin of a governor becomes larger and remains statistically significant;

---

[29]The test statistic for the error structure of the residuals further suggests that we do not have persistent autocorrelation in the error term. Finally, the Sargan test statistic cannot reject the null hypothesis that the overidentifying restrictions from the instruments (lags from $t-2$ to $t-4$ of the dependent and predetermined independent variables) are valid.

the interaction effect remains positive but loses statistical significance.

Furthermore, we assess whether our results are robust to the definition of lame duck governors. We first drop states with a one term limit to rule out that the positive effect of lame ducks on experimentation is accounted for by governors elected in these states. Column (6) of Table 3 shows that our results are unaffected and hence, not driven by states with a one term limit. One might also worry that states with no term limit for governors differ from states with a term limit along some dimension that make them less likely to experiment. To address this concern, column (7) restricts the sample to states with term limits. The identification of the interaction effect now comes from governors in their first and second term in states with a term limit. Again, the interaction effect remains qualitatively similar. Hence, the fact that lame ducks experiment more cannot be accounted for by unobservable differences between states with and without term limits for governors.

We next use alternative measures of the potential gains from experimentation. Even conditional on the included state characteristics, the potential gains from experimentation might also depend on the demand for welfare. One explicit goal of the 1996 welfare reform was to decentralize decision-making to the state level. A potentially important benefit of decentralization is that policies are, under certain conditions, closer to the preferences of the electorate. Many states implemented strict rules excluding recent immigrants from access to welfare, for instance. To assess the influence of the demand side, column (7) of Table 3 includes the size of the immigrant population, share of Blacks, the share of unmarried births and a measure for income inequality to the baseline specification. Interestingly, none of these variables has an effect on policy experimentation. Hence, states with higher welfare spending are more likely to experiment even conditional on the size of the immigrant population, the share of Blacks or single moms in a state.

The last specification (in column (8) of Table 3) uses total welfare spending of a state which includes spending on other complementary programs like Food Stamps. If states engage in policy experimentation in order to reduce caseloads, the gains from experimentation would be higher the more a state spends on all welfare programs (as many beneficiaries of AFDC or TANF are also eligible for Food Stamps, for example).

At the same time, we would expect that endogeneity is less of a concern for total welfare spending than for AFDC/TANF spending. The estimates in the last column show that higher total welfare spending encourages policy experimentation just like the baseline estimates in Table 2.

So far, we have treated all governors as if they had little information about the state of the world. Yet, some governors might be better in recognizing and discerning valuable information from noisy cues. To explore such heterogeneity, we use two proxies for a governor's ability to predict the state of the world: her prior political experience and her educational background. Experience in prior political offices should be a good indicator of the governor's ability either if elections select the most able candidate; or if governors improve their political decision-making under uncertainty while in office (see, e.g., Alt et al., 2011, for evidence supporting the latter view). Alternatively, we proxy ability as whether a governor has a Masters or higher degree (see Galasso and Nannicini, 2011; Gagliarducci and Nannicini, 2013; Kotakorpi and Poutvaara, 2011, for similar approaches).[30] The first specification in Table A3 shows whether governors with political experience below the median years (in column (1)) or less than a Masters degree (in column (3)) are more or less likely to experiment. A governor with little experience or education who does not worry about her reelection is somewhat more likely to experiment.[31] The second specification interacts the indicator for low experience or education with being a lame duck. There is some evidence that governors with little prior experience are more likely to experiment when they are a lame duck. While the interaction effect between lame duck and low educational attainment is positive, it is not statistically significant. Hence, there is some evidence that governors might differ in their ability to assess the state of the world. More importantly however, our baseline estimates (also shown in Table A3) do not change when we account for heterogeneity among governors.

---

[30]While voters may in principle observe her tenure as a governor, they are unlikely to know exactly how much political experience a governor has accumulated prior to being elected. Voters might be more aware of the educational credentials of a governor however.

[31]A low-quality politician may be more or less likely to experiment than a high-quality politician. The high-quality politician experiments when the state of the world is N which occurs with probability p (Proposition 1). The low-quality politician is more likely to experiment if $\gamma = 1$ (Proposition 2) but less likely to experiment if $\gamma = 0$ and $\lambda \geq p$ (Proposition 3).

6.2   Effect of Governor Ideology

A common perception of the welfare reform, and certainly the federal PRWORA reform, is that it was fueled by a conservative agenda to curb welfare spending. Yet, can ideology among governors also explain experimentation with welfare rules at the state level? Most importantly, are ideological differences among governors responsible for the role electoral incentives play for policy experimentation as shown in Tables 2 and 3? To assess these questions, we first study the relationship between governor ideology and experimentation by including an indicator whether the governor is a Republican (in column (1) of Table 4) or Berry's measure of party ideology to account for ideological differences within parties across states (in column (2) of Table 4). The measure of governor ideology varies from 0 to 100 where larger values indicate more liberal views. As expected, Republican governors are more likely to experiment than Democratic governors. Also, more liberal governors are significantly less likely to experiment. Both measures support the idea that governor ideology matters for the decision to experiment.

We next add our measures of electoral incentives to the specification: we start with the past vote margin in addition to the two alternative measures of governor ideology (in columns (3) and (4) of Table 4). The results show that the past vote margin influences policy experimentation even conditional on governor ideology. We next add our other key independent variables to the specification: the interaction effect with being a lame duck and past welfare spending (in columns (5) and (6) of Table 4). We find that our estimates controlling for governor ideology are very similar to the baseline in Table 2.

The last two specifications in Table 4 re-estimate the baseline after splitting the sample into whether the governor is a Republican (column (5)) or a Democrat (column (6)). The results show that higher electoral support and higher potential gains have a stronger effect on experimentation for Republicans than for Democrats: The negative effect for governors who can be reelected is almost twice as large as for Democratic governors, for instance. Yet, the impact of electoral support on experimentation is again zero if a governor is a lame duck – irrespective of whether the governor belongs to the Republican or Democratic party.

Overall then, our estimates show that electoral incentives remain an important

factor even after accounting flexibly for the ideological stance of the governor. To assess the relative importance of the two channels, we use the estimates in columns (5) of Table 4. Accordingly, governor ideology accounts for about 30.9% (0.0876/0.283) of policy experiments, while electoral incentives (or the absence thereof) accounts for 24.4% (-0.0064*10.8/0.283) and 8.4% (-0.0064+0.0086)*10.8/0.283) respectively. Hence, both electoral incentives and ideology are about equally important in explaining policy experimentation.

6.3 Alternative Explanations

Overall, our findings provide strong evidence that electoral incentives influence the decision to experiment with welfare policy. Our analysis has so far abstracted from other potential influences on policy innovations like voter preferences, the legislature or spillovers between states. We now investigate whether these factors play a role for the decision to engage in policy experimentation.

6.3.1 Influence of Heterogeneity in Voter Preferences

Our analysis has so far abstracted from voter heterogeneity. The model assumed for simplicity that there is no disagreement about the potential gains from experimentation as there is a single representative voter. In the empirical analysis, we have proxied the potential gains as past welfare spending. However, voters might differ in their assessment of the optimal amount of welfare spending: some might find it too high, others too low.

We first check whether voter preferences, measured by Berry's measure of voter ideology (in column (1) of Table 5) or the Democratic vote share in the last Presidential election (in column (2) of Table 5) affect experimentation. Both columns show no systematic relationship between state-level shifts in voter ideology and policy experimentation. We next add our main independent variables for electoral incentives in columns (3) and (4) of Table 5. We find that voter ideology still plays no role for experimentation. Yet, we again find the same patterns than in the baseline: past electoral support reduces experimentation, lame duck governors are more likely to experiment and past spending encourages experimentation just as in Table 2.

Yet, our measures of voter ideology might be too crude to pick up preferences for redistribution. As an alternative strategy, we use data on the preferences for redistribution from the General Social Survey. The measure in column (5) of Table 5 is an indicator equal to one if respondents in a state are more likely to support redistribution than respondents in the median state in that year. The measure in column (6) of Table 5 is an indicator if respondents in a state are more likely than respondents in the median state to say that their government spends too little on welfare. In both cases, higher values imply that respondents prefer more redistribution. Interestingly, we find that stronger preferences for redistribution among the electorate encourage policy experimentation. This effect is weaker however, in states with high expenditures on welfare in the past (see columns (6) and (8) of Table 5). Again, our baseline estimates are unaffected by the inclusion of voter preferences or heterogeneity about the benefits of welfare transfers.

6.3.2 Influence of the Legislature

Our main empirical analysis restricts attention to the governor as the relevant decision-maker. Yet, the legislature is certainly also an important player in state policy-making. Veto players in the legislative process, for example, might delay or reduce policy experimentation while more competition in the legislature or between governor and legislature possibly increases it. Previous evidence suggests that party composition and tightness of legislative elections seem to have an effect on welfare spending under AFDC (e.g. Barrilleaux et al., 2002).

In Table 6, we add a number of controls for the legislature to our baseline specification in Table 2: The party composition in the lower and upper house (in column (1)); the degree of polarization in the parliament which is often viewed as an indicator for legislative competition (in column(2)); whether the government is divided (in column (3)); and all controls simultaneously (in column (4)). The only legislative variable that influences policy experimentation is whether the state has a divided government (see Bernecker, 2016, for a more detailed empirical investigation). Most importantly, the coefficients on the past vote margin, its interaction with being a lame duck and past welfare spending are hardly affected by controlling for the structure of the legislature.

30

6.3.3 Spillover Effects across States

Our analysis has also abstracted from spillover effects across states. If potential welfare recipients are geographically mobile, states may engage in welfare competition. In that case, experimentation in one state is influenced by policy choices in neighboring states (see, e.g., Figlio et al., 1999; Saavedra, 2000; Wheaton, 2000). Spillovers can also arise if policy choices in neighboring states provide information for voters as in a model of yardstick competition (Besley and Case, 1995b). In fact, a central insight of the literature on laboratory federalism is that decentralization promotes learning and spillover of knowledge among decision-makers within a federal system (see, e.g., Oates, 1999; Strumpf, 2002).

We create three different measures to capture learning from neighboring states. Our first measure calculates how much geographic neighbors experiment in prior years. Alternatively, we define neighbors as states with similar population size and create their mean propensity to experiment in previous years. Finally, we also consider spillovers from states that are ideologically similar (as measured by the Democratic vote share in the 1978 Presidential election) and located within the same census region. Table 7 shows that there is little systematic statistical relationship between policy experiments in a state and experimentation in neighboring states - independently of how we define neighbors. While all coefficients are positive suggesting some imitation across states, the effects never reach statistical significance.

Including measures of cross-state spillovers leaves our baseline estimates unchanged suggesting that electoral incentives of governors play a role for experimentation even beyond learning from other states. How does the absence of cross-state spillovers square with the rich case study evidence that certain states and governors (like Tommy Thompson in Wisconsin) played an decisive role in the welfare reform process? The key observation here is that cross-state spillovers matter if we do not control for aggregate waiver activity (through year effects). One possible interpretation is that state governors learn from and possibly imitate neighboring states when there is a lot of experimentation going on. Yet, once we control for the overall activity pattern of experimentation, spillovers through policy diffusion play much less of a role.

6.4 Evidence on Policy Reversals

After having established that electoral incentives are an important driver of the decision to experiment with risky policies, we now turn to the reversal of policy rules. Our theoretical framework suggests that policy reversals should be a rare event because a governor who reverses an experiment is not reelected for sure. Table A4 in the appendix provides some descriptive evidence on policy reversals during the TANF period.[32] For each of the eleven policy rules included in our measure, the table lists the number of experiments and the number of reversals (in columns (1) and (2) of Table A4). The numbers show that policy reversals are indeed a rare event: while 22 states experimented with a family cap, for instance, only 5 states reversed the family cap later on. Across all rules we analyze, about 20% of experiments get reversed.

Column (3) of Table A4 further shows that very few policy reversals happen within the same electoral cycle as the experiment. There are at least two potential explanations for such a pattern: first, governors do not like to reverse an experiment because it demonstrates to voters that the governor made a mistake (as stressed in our theoretical framework). Reversals might then be observed only when governors from the opposite side of the ideological spectrum come to power in a later cycle. Second, reversals within the same electoral cycle might be rare because it takes time to observe the success or failure of an experiment. Column (4) of Table A4 shows that the mean number of years between an experiment and a reversal is with 5.5 years longer than a typical electoral cycle. The last two columns of Table A4 indicate that a reversal is implemented in many cases by a governor from the same party as the initial experimenter. The descriptive evidence then suggests that reversals of experiments are not only rare, but that it also takes time to learn about and evaluate the outcome of an experiment. We find little support for the argument that policy experiments are mainly reversed by political opponents after they get voted into office.

To investigate policy reversals a bit more systematically, we relate whether a policy reversal occurs (columns (1)-(4) of Table 8) and the number of policy reversals (in columns (5)-(8) of Table 8) to our baseline variables using equation (7). An important

---

[32]As discussed in Section 2, we cannot measure policy reversals during the waiver period before 1996.

caveat is that a policy experiment should only be reversed when the experiment fails. Unfortunately, we do not observe whether the experiment failed to deliver short term benefits or not. Our reversal measure therefore contains substantial measurement error which reduces the precision of our estimates. We thus view the results in Table 8 as suggestive in nature. In line with our model, we find that higher potential gains from experimentation make a policy reversal less likely. In our theoretical framework, governors prefer to gamble: They stick to a policy experiment and hope that the experiment turns out to be successful in the second period.

In the second specification (in columns (2) and (6)), we add whether a governor is a lame duck. We would expect that lame ducks are more likely to revert a policy experiments because a reversal is less costly for voter welfare if the experiment is not successful in the short term. The coefficient is positive suggesting that lame ducks are indeed more likely to revert a policy than governors who can be reelected. Yet, the coefficient is not statistically significant. The third specification in columns (3) and (7) also includes whether the governor is a Republican: the coefficient is negative suggesting that Republicans are more hesitant to revert an experiment in line with our model. Again, the large standard errors suggest there is substantial noise in our reversal measure. The final specification in columns (4) and (8) include the past vote margin of the governor which, according to our model, should not play a role for the decision to reverse a policy experiment. And indeed, the correlation between initial reputation and a policy reversal is very weak (relative to the effect on policy experimentation in Table 3, for example) and never close to statistically significant. The fact that electoral support has little relationship with the reversal decision suggests that voters do not seem to update their opinion about a politician based on the observed policy choices. The absence of updating could be an indication that voters may not be able to fully observe or evaluate the politicians' actions and the policy's consequences within a relatively short period of time.[33]

---

[33]The result that initial electoral support does not matter for the decision to revert a policy differs from the prediction in the full signaling game of Majumdar and Mukand (2004). There, voters update their belief about the politician's quality after observing policy choices and their outcomes. In that case, the initial reputation would also have an effect on reversal decisions.

7 Conclusion

This article analyzes whether policy experimentation is influenced by electoral incentives. Our empirical setting is the U.S. welfare reform in 1996 which marked the most important shift in social policy since the New Deal. Both during the waiver period and the post-1996 period, state governments faced new policy options many of which had never been implemented before. To guide our empirical analysis, we develop a political agency model which shows that reelection concerns reduce incentives to engage in policy experimentation.

Our results confirm that reelection concerns reduce the probability of experimentation during the AFDC and TANF period. We also show that governors who face binding time limits behave systematically different from governors who face reelections and hence worry more about their electoral support. We further find that ideological differences among policy-makers influence experimentation; yet, our main result that electoral incentives matter for policy experimentation holds even after accounting for ideological influences. Finally, we find that none of the alternative factors we consider – role of voter preferences, the influence of the legislature, and learning from other states – invalidates our conclusion that electoral incentives matter for the decision to implement novel policies.

One implication emerging from our analysis is that mitigating reelection concerns may be beneficial when both voters and politicians are uncertain about the best policy. Strong electoral concerns make politicians reluctant to experiment with risky policies; they also reduce the likelihood that politicians reverse failed experiments. As electoral concerns are more muted when politicians face a binding term limit, our analysis provides one possible rationale for why term limits exist.

References

Alesina, A. and A. Drazen (1991). Why are stabilizations delayed? *American Economic Review 81*(5), 1170–1188.

Alt, J., E. B. de Mesquita, and S. Rose (2011). Disentangling accountability and competence in elections: Evidence from U.S. term limits. *Journal of Politics 73*(1), 171–186.

Arellano, M. and S. Bond (1991). Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *Review of Economic Studies 58*, 277–297.

Barrilleaux, C., T. Holbrook, and L. Langer (2002). Electoral competition, legislative balance, and American state welfare policy. *American Political Science Review 46*(2), 415–427.

Bernecker, A. (2016). Divided we reform? Evidence from US welfare policies. *Journal of Public Economics 142*, 24–38.

Berry, F. S. and W. D. Berry (1990). State lottery adoptions as policy innovations: An event history analysis. *American Political Science Review 84*(2), 395–415.

Berry, F. S. and W. D. Berry (1992). Tax innovation in the states: Capitalizing on political opportunity. *American Journal of Political Science 36*(3), 715–742.

Berry, F. S. and W. D. Berry (2007). Innovation and diffusion models in policy research. In P. A. Sabatier (Ed.), *Theories of the Policy Process*, second edition, pp. 223–260. Boulder, CO: Westview.

Berry, W. D., E. J. Ringquist, R. C. Fording, and R. L. Hanson (1998). Measuring citizen and government ideology in the American States, 1960-93. *American Journal of Political Science 42(1)*, 327–348.

Besley, T. (2007). Principled agents?: The political economy of good government. *Oxford University Press*.

Besley, T. and A. Case (1995a). Does electoral accountability affect economic policy choices? Evidence from gubernatorial term limits. *Quarterly Journal of Economics 110*(3), 769–798.

Besley, T. and A. Case (1995b). Incumbent behavior: Vote-seeking, tax-setting, and yardstick competition. *American Economic Review 85*(1), 25–45.

Besley, T. and A. Case (2003). Political institutions and policy choices: Evidence from the United States. *Journal of Economic Literature 41*(1), 7–73.

Blank, R. M. (2002). Evaluating welfare reform in the United States. *Journal of Economic Literature 40*, 1105–66.

Callander, S. (2011a). Searching and learning by trial and error. *American Economic Review 101*(6), 2277–2308.

Callander, S. (2011b). Searching for good policies. *American Political Science Review 105*(04), 643–662.

Callander, S. and B. Harstad (2015). Experimentation in federal systems. *Quarterly Journal of Economics 130*(2), 951–1002.

Callander, S. and P. Hummel (2014). Preemptive policy experimentation. *Econometrica 82*(4), 1509–1528.

Center for Economic and Policy Research (2012). *Current Population Survey ORG Uniform Extracts*, Volume Version 1.7. Washington, D.C.

Ciccone, A. (2004). Resistance to reform: Status quo bias in the presence of individual-specific uncertainty: Comment. *American Economic Review 94*(3), 785–795.

Coate, S. and S. Morris (1999). Policy persistence. *American Economic Review 89*(5), 1327–1336.

Council of Economic Advisors (1999). The effects of welfare policy and economic expansion on welfare caseloads: an update. Technical report, Executive Office of the President of the United States.

Crouse, G. (1999). Implementation of major changes to welfare policies, 1992-1998. http://www.aspe.hhs.gov/hsp/Waiver-Policies99/policy_CEA.htm (accessed on February 7, 2014).

Dewatripont, M. and G. Roland (1992). Economic reform and dynamic political constraints. *Review of Economic Studies 4*, 703–730.

Drazen, A. and V. Grilli (1993). The benefit of crises for economic reforms. *American Economic Review 83*(3), 598–607.

Fang, H. and M. P. Keane (2004). Assessing the impact of Welfare Reform on single mothers. *Brookings Papers on Economic Activity 1*, 1–95.

Fernandez, R. and D. Rodrik (1991). Resistance to reform: Status quo bias in the presence of individual- specific uncertainty. *American Economic Review 81*(5), 1146–1155.

Ferraz, C. and F. Finan (2011). Electoral accountability and corruption in local governments: evidence from audit reports. *American Economic Review 101*(4), 1274–1311.

Figlio, D., V. Kolpin, and W. Reid (1999). Do states play welfare games? *Journal of Urban Economics 46*, 237–254.

Fu, Q. and M. Li (2014). Reputation-concerned policy makers and institutional status quo bias. *Journal of Public Economics 110*, 15 – 25.

Gagliarducci, S. and T. Nannicini (2013). Do better paid politicians perform better? Disentangling incentives from selection. *Journal of the European Economic Association 11(2)*, 369–398.

Galasso, V. and T. Nannicini (2011). Competing on good politicians. *American Political Science Review 105*(01), 79–99.

Grogger, J. and L. A. Karoly (2005). *Welfare Reform: Effects of a Decade of Change*. Harvard University Press.

Harrington, J. (1993). Economic policy, economic performance, and elections. *American Economic Review 83*(1), 27–42.

Harvey, C., M. J. Camasso, and R. Jagannathan (2000). Evaluating welfare reform waivers under Section 1115. *Journal of Economic Perspectives 14*, 165–188.

Haskins, R. (2006). *Work over Welfare: The Inside Story of the 1996 Welfare Reform Law*. Brookings Institution Press.

Holbrook, T. and E. Van Dunk (1993). Electoral competition in the American States. *American Political Science Review 87*(4), 955–962.

Holmström, B. (1982). Managerial incentive problems - A dynamic perspective. In *Essays in Economics and Management in Honor of Lars Wahlbeck*. Swedish School of Economics.

Holmström, B. (1999). Managerial incentive problems: A dynamic perspective. *Review of Economic Studies 66*(1), 169–182.

Jain, S. and S. W. Mukand (2003). Redistributive promises and the adoption of economic reform. *American Economic Review 93*(1), 256–264.

Kaplan, T. (2000). Wisconsin works. In S. F. Liebschutz (Ed.), *Managing Welfare Reform in Five States: The Challenge of Devolution*. Rockefeller Institute.

Karch, A. (2007). *Democratic Laboratories: Policy Diffusion among the American States.* University of Michigan Press.

Klarner, C. (2003). Measurement of partisan balance of state government. *State Politics and Policy Quarterly 3*, 310–319.

Koerper, K. (1996). Weekly tracking update - welfare reform: Section 1115 Waiver Activity. Office of Planning, Research and Evaluation.

Kotakorpi, K. and P. Poutvaara (2011). Pay for politicians and candidate selection: An empirical analysis. *Journal of Public Economics 95*(7–8), 877 – 885.

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

Leip, D. (2012). Dave Leip's Atlas of U.S. Presidential Elections. http://www.uselectionatlas.org. (accessed on February 7, 2014).

Lieberman, R. C. and G. M. Shaw (2000). Looking inward, looking outward: The politics of state welfare innovation under devolution. *Political Research Quarterly 53*(2), 215–240.

List, J. A. and D. M. Sturm (2006). How elections matter: Theory and evidence from environmental policy. *Quarterly Journal of Economics 121*(4), 1249–1281.

Majumdar, S. and S. W. Mukand (2004). Policy gambles. *American Economic Review 94*(4), 1207–1222.

Mead, L. M. (2004). *Government Matters. Welfare Reform in Wisconsin.* Princeton University Press.

Moffitt, R. A. (2002). Welfare programs and labor supply. In A. J. Auerbach and M. Feldstein (Eds.), *Handbook of Public Economics*, Chapter 34, pp. 2393–2430. Elsevier.

Moffitt, R. A. (2008). Welfare reform: The US experience. *Swedish Economic Policy Review 14 (2)*, 11–54.

Oates (1999). An essay on fiscal federalism. *Journal of Economic Literature 37*(3), 1120–1149.

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

Poole, K. T. and H. Rosenthal (1984). The polarization of american politics. *Journal of Politics 46*, 1061–79.

Poole, K. T. and H. Rosenthal (1997). *Congress: A Political-Economic History of Roll Call Voting.* New York: Oxford University Press.

Prato, C. and S. Wolton (2017). Rational ignorance, populism, and reform. *European Journal of Political Economy.*

Reintsma, M. (2007). *The Political Economy of Welfare Reform in the United States.* Edward Elgar.

Rogoff, K. (1990). Equilibrium political budget cycles. *American Economic Review 80*(1), 21–36.

Saavedra, L. (2000). A model of welfare competition with empirical evidence from AFDC. *Journal of Urban Economics 47*, 248–279.

Smart, M. and D. M. Sturm (2013). Term limits and electoral accountability. *Journal of Pulic Economics 107*, 93–102.

Soss, J., S. F. Schram, T. P. Vartanian, and E. O'Brien (2001). Setting the terms of relief: Explaining state policy choices in the devolution revolution. *American Journal of Political Science 45*(2), 378–395.

Strumpf, K. S. (2002). Does government decentralization increase policy innovation? *Journal of Public Economic Theory 4*(2), 207 – 241.

United States Census Bureau (2011). Statistical Abstract of the United States: 2012 (131st edition).

Urban Institute (2015). Welfare Rules Database. Washington, D.C. (accessed on May 7, 2015).

U.S. Department of Health and Human Services, Administration for Children and Families (1998-2009). TANF Annual Reports to Congress, I-VIII. Technical report, GAO.

U.S. Department of Homeland Security (2011). *Yearbook of Immigration Statistics.* Washington, D.C.: Office of Immigration Statistics.

Walker, J. L. (1969). The diffusion of innovations among the American States. *American Political Science Review 63*(3), 880–899.

Weaver, K. (2000). *Ending Welfare as we Know it.* Brookings Institution.

Weissert, C. S. (2000). *Learning from Leaders: Welfare Reform Politics and Policy in five Midwestern States.* Rockefeller Institute Press.

Wheaton, W. C. (2000). Decentralized welfare: Will there be underprovision? *Journal of Urban Economics 48*, 536–555.

Willems, T. (2013). Political accountability and policy experimentation: Why to elect left-handed politicians? *Mimeo. Princeton University.* mimeo.

Winston, P. (2002). *Welfare Policymaking in the States: The Devil in Devolution.* Georgetown University Press.

Online Appendix

A   Theory: Proofs

Proof of Proposition 1.   The proof of Proposition 1 follows from the optimization problem of the incumbent politician. A high-quality incumbent politician knows the state of the world. At $T = 0$, the incumbent chooses to reform or not (with the option to revert back to a_S if the reform is undertaken) in order to maximize equation (1).

In state S, no gain can be obtained from the reform ex-post: The ex-post welfare of the reform in state S is either $-c$ or $-\frac{c}{2}$ if reverted back to a_S. Therefore, the reform should not be undertaken from a welfare perspective. Since the ex-post welfare of a reform in state S is negative, the probability of reelection (from the reelection rule) in state S is (weakly) positive if and only if no reform is undertaken. The optimal decision of the high-quality incumbent is then to stick with the status quo policy a_S independently of γ and λ.

In state N, undertaking the reform maximizes ex-post welfare since there are positive welfare gains ex-post with probability 1, i.e. the ex-post welfare with a reform in state N is equal to $\Delta - c$ which is strictly positive from $(A1)$, and is never reversed independently of the realization of the benefit in the short term. In addition, the probability of reelection is also maximized by implementing the reform since a positive ex-post welfare ensures reelection with probability 1, which is (weakly) greater than λ for any $\lambda \in [0, 1]$.

Proof of Proposition 2.   The proof of Proposition 2 follows from the optimization problem of the incumbent politician given by equation (1) for $\gamma = 1$. For a low-quality politician who does not know the state of the world, the reform is welfare maximizing from an ex-ante perspective from $(A1)$ and $(A2)$. Therefore, she always implements the reform in $T = 0$. However, it is welfare maximizing to revert back to a_S if the reform fails in the short run since the expected future benefit from the reform is negative from $(A3)$.

Proof of Proposition 3.   The proof of Proposition 3 follows from the optimization problem of the incumbent politician given by equation (1) for $\gamma = 0$.

A purely office-motivated politician never reverts back to the status quo after implementing a reform. If the reform delivers benefits in the short term it is always continued. If the reform does not deliver short-term benefits in $T = 1$, a policy reversal back to the status quo implies that the politician would lose the election with probability 1 (because welfare ex-post would be equal to $-\frac{c}{2}$ and hence, the probability of reelection would be zero from the reelection rule). Whereas if the politician persists with the reform, she can win the election for sure if the long-term benefit realizes which happens with probability $\beta > 0$.

For a low-quality politician who does not reverse a reform, the reform maximizes the probability of winning from an ex-ante perspective if it delivers higher chances of reelection than her current reputation. This statement is equivalent to equation (2) where the left-hand side represents the ex-ante winning probability if the reform is undertaken and continued which is equivalent to the probability that the state of the world is N. The right-hand side is the ex-ante winning probability if the reform is not undertaken which is equal to the initial reputation of the incumbent.

Proof of Proposition 4. The proof of Proposition 4 directly follows from comparing the politician's payoffs in case of reverting the reform and of continuing the reform after a short-term failure. Reverting the reform is at least as good as continuing it if and only if

$$\gamma\left(\frac{-c}{2}\right) \geq \gamma\left(\beta\Delta - c\right) + (1 - \gamma)\beta,$$

or

$$-\beta \geq \gamma\left(\beta\Delta - \frac{c}{2} - \beta\right).$$

Since the term in parenthesis is negative from $(A3)$, the last inequality can be rewritten to obtain the threshold γ^* in equation (3).

Proof of Corollary 1. The proof of Corollary 1 directly follows from differentiating equation (3) with respect to γ and Δ, respectively.

Proof of Proposition 5. The proof of Proposition 5 directly follows from the comparison of the politician's payoff if she undertakes the reform (for the two cases where she will or will not revert back to the status quo in the face of short-term failure) with the payoff from sticking with the status quo and running the next election on her initial reputation instead.

Proof of Corollary 2. The proof of Corollary 2 follows directly from differentiating equations (4) and (5) with respect to Δ, γ and λ respectively.

B   Data

B.1   Measuring Policy Experimentation during the TANF period

To measure policy experimentation and reversals for the post-1996 period, we rely on the Welfare Rules Database by the Urban Institute. The database provides detailed information about states' TANF policies obtained from caseworker manuals and regulations which are typically more detailed and up-to-date than the official plans state submit periodically to the federal government. While the database contains hundreds of rules on eligibility, benefit calculation, and many other aspects of welfare reform, we focus on a set of rules in policy areas that were at the center of the public and political debate surrounding welfare reform (see Table A1 for a list of the rules and how we code policy experiments and policy reversals).

Family caps. Under AFDC, benefit levels increased with family size. Under a family cap, additional benefits that an assistance unit would receive for a child born into the benefit unit while on welfare would be capped. Some states provide a percentage of the increase to the unit, while others provide no additional funds to the unit for the additional child. In several states, a family is never able to regain benefits for a capped child, even after the case has been closed for a period of time. In others, a family cap can be removed (and hence, the child can be included in the benefit computation should the family apply for assistance again) if the assistance unit remains off welfare for some time. We code a binary indicator equal to one if a state has a family cap and zero if not.

Work requirements. Under AFDC, states could require recipients to participate in the Job Opportunities and Basic Skills Training (JOBS) program, which provided education, training, and work experience activities. However, many individuals were exempt from these requirements (because of age, illness or having a small child). Under TANF, states require adults heading an assistance unit to perform some type of work-related activity. Work programs vary widely from state to state in terms of who must work, how much work is required, and what activities are considered work. The first rule defines the minimum number of hours a recipient must participate in work-related activities. The hours requirements vary from a mere effort to find a job up to full-time employment. The second rule defines whether the work requirement applies after several months of benefit receipt or by the time of application or approval. The third rule indicates whether there is a time limit of benefit receipt if a parent fails to work at least 20 hours per week in a regular job after a certain number of assistance months. The fourth rule counts the number of work exemptions due to, for example, pregnancy, disability or caring for a young child or elderly person.

Sanctions. Under AFDC, sanctions were sometimes imposed if the adult heading an assistance unit did not comply with the JOBS program; in that case, the adult portion of the benefit was not paid out to the assistance unit. Under TANF, states now require household heads to perform some type of work-related activity sooner or later. If a benefit unit does not comply with these requirements, states can impose drastic sanctions. The first policy rule characterizes the initial sanction if a benefit unit fails to comply with the work requirements for the first time. The initial sanction varies from reduction of 25% or less to a suspension of the full family benefit. The other three rules characterize the severity of the worst sanction that can be imposed. The second rule defines the severity of the worst sanction varying from less than 25% of the benefit to a suspension of the full family benefit and even case closure. The third rule defines the duration of the worst sanction ranging from until the unit complies with the requirement to a permanent suspension of the family benefit. The final rule defines whether a unit has to reapply (or not) after the worst sanction has been imposed.

Termination and work-requirement time limits. Under AFDC, families were entitled to receive benefits as long as they met the eligibility requirements. Under TANF, many states imposed both intermittent and lifetime time limits. The first rule characterizes the number of months an assistance unit can receive benefits without interruption (many states impose 24 months). The second rule defines how benefits are reduced when the assistance unit reaches the intermittent time limit. The loss of benefits might be just for the adult members or for the entire assistance unit. The third rule defines the duration of the lifetime limit ranging from no time limit to only 24 months. If states wish to extend benefits beyond the federal time lifetime limit of 60 months, they have to use their own state funds to finance it. The final rule defines whether the state allows for any extensions to the lifetime limit or not.

B.2 Politics and ideology measures

Governor characteristics. We collect biographies of U.S. governors from the website of the National Governor's Association and data provided by David J. Andersen from the Eagleton Institute of Politics at Rutgers University. We code information on the age

and educational degree of the governor. To measure governor heterogeneity whether an experiment is called for, we calculate her political experience prior to becoming governor. Specifically, the variable measures the number of years between the first electoral office (such as member of the State Senate or Attorney General) and the inauguration as governor. We then construct a dummy variable equal to one if a governor has prior political experience below the median and zero otherwise. Information on gubernatorial election results (especially vote margins) and term limits are taken from List and Sturm (2006); and Leip (2012).

Ideology. The vote share for the Democratic candidate in the last presidential election is taken from the Statistical Abstract (United States Census Bureau (2011)) and updated using Leip (2012). Voter and governor ideology is based on ideology ratings of the state's congressional delegation, the American for Democratic Action (ADA) rating and the AFL/CIO's Committee on Political Education (COPE) rating (Berry et al. (1998)). Berry et al. (1998) assign an ideology rating to the citizens of each congressional district by using an average of the score of the congressional member and his or her election opponent weighted by the number of votes the score received. Zero denotes the most conservative and 100 the most liberal. They then generate a state-wide measure by averaging over all congressional districts. The measure of governor ideology is constructed by assigning to the governor the ratings of the members of Congress from their party. Updates of these ideology data are available at http://rcfording.wordpress.com/state-ideology-data/.

State Legislature. Data for the composition of the state legislature, state competitiveness, the party of the governor and indicators for a divided government are obtained from Klarner (2003) and updated from http://www.indstate.edu/polsci/klarnerpolitics.htm. Polarization in the state legislature is calculated as |democratic seat share–50%| for the state senate and house respectively. The divided government indicator is equal to one if the governor belongs to a different party than the majority of legislators in either the state senate or the state house. To measure state competitiveness, we use the Holbrook-Van Dunk index calculated from district-level returns to state legislative elections (see Holbrook and Van Dunk (1993)).

B.3 State Demographics and Other Controls

Demographics. Population size, the number of Blacks and the age structure are taken from the Statistical Abstract (United States Census Bureau (2011)). The size of the immigrant population refers to the number of legal immigrants admitted by state of intended residence and is taken from Fang and Keane (2004) for 1970 to 2002 and updated using the Yearbook of Immigration Statistics (U.S. Department of Homeland Security (2011)). Personal income per capita is taken from the website of the Bureau of Economic Analysis. The unemployment rate for 1960 to 1998 is from the website of the Bureau of Labor Statistics. Our measure of income inequality is the ratio of the 90th over the 10th percentile of total household income calculated from the March Current Population Survey (Center for Economic and Policy Research (2012)). All income variables are deflated by the urban consumer price index with base year 2002. Unmarried birth refers to the % of births to unmarried women per 1,000 unmarried women aged 15-44 years by state of residence. For the years 1992 to 2003, the data are available from Table 8.3 in the TANF Annual Reports to Congress (U.S. Department of Health and Human Services,

Administration for Children and Families (2009)). For earlier and later years, we obtain the data from the Center for Disease Control and Prevention (CDC), the National Vital Statistics System and Vital Statistics.

State AFDC and TANF spending. To measure the potential gains from experimentation, we use state-level AFDC and TANF spending. These data come from Paul Ehmann at the U.S. Census Bureau. All fiscal variables are then converted into real measures using the urban consumer price index (with years 1982-1984=100) from the Bureau of Labor Statistics.

Spillover effects across States. To analyze spillover effects across states, we define a "neighbor" along two different dimensions. The first variable (geographic neighbors) codes the extent of experimentation for all states sharing a common border with the current state. The second variable (similar population sizes) calculates experimentation in states with a population size similar to that of the current state. For the calculation, we use the following ten bands based on population size in 1978: (CA NY TX PA IL), (OH MI FL NJ MA), (NC IN GA VA MO), (WI TN MD LA MN), (WA AL KY CT SC), (IA OK CO AZ OR), (MS KS AR WV NE), (UT NM ME RI HI), (ID NH MT NV SD) and (ND DE VT WY AK). The third measure (ideological neighbors) takes the average number of experiments of the four states within the same census region who voted similarly in the presidential election of 1978.

Table 1: Summary Statistics

	Waiver Period		Tanf Period	
	Mean	Std. Dev.	Mean	Std. Dev.
# Policy Experiments	0.212	0.591	0.559	0.936
# Policy Reversals			0.099	0.332
Dummy for Policy Reversal (TANF Period)			0.089	0.284
Governor's Past Vote Margin	8.53	7.34	10.05	8.77
Governor Lame Duck	0.236	0.425	0.282	0.450
Governor Age	52.77	7.87	55.17	7.61
Governor Party	-0.013	0.281	0.076	0.319
Governor Ideology (Berry et al. 1998)	-0.013	0.281	0.076	0.319
Citizens Ideology (Berry et al. 1998)	46.95	15.08	50.99	15.42
Democratic Presidential Vote Share	43.44	8.99	47.44	9.24
Divided Government	0.546	0.498	0.566	0.496
Democratic Seat Share Upper House	0.600	0.186	0.504	0.153
Democratic Seat Share Lower House	0.589	0.180	0.511	0.152
Polarization Nominate Upper House	0.573	0.168	0.820	0.184
Polarization Nominate Lower House	0.585	0.211	0.598	0.194
Competitiveness Index (Holbrook and Van Dunk)	40.70	12.52	38.89	10.96
AFDC/TANF Spending (in US$ millions)	547.98	1024.12	517.69	1036.32
Unemployment Rate	6.50	2.07	4.76	1.11
Per Capita Income (/1000)	15586.42	5295.12	31254.28	7287.64
Population (/1000)	5.01	5.27	5.91	6.35
% Black Population	9.80	9.30	11.12	10.71
% Population 65+	12.13	1.89	12.75	1.60
%Population 5-17	19.64	2.02	18.26	1.40
% Immigrant Population	1.68	2.13	2.18	1.74
% Unmarried Births	24.29	8.95	34.97	7.38
Income Inequality (P90/P10)	7.95	1.77	8.80	1.94
Geographic Neighbors' Experiments (t-2)	0.207	0.341	0.598	0.699
Pop. Size Neighbors' Experiments (t-2)	0.206	0.356	0.605	0.690
Ideological Neighbors' Experiment (t-2)	0.230	0.395	0.598	0.699

Notes: For details on the welfare experimentation and reversal measures, see the data appendix. Governor Lame duck is equal to one if the governor cannot run for reelection. The past electoral support measures the winner's distance in votes to the runner-up in the last gubernatorial election. The ideology measure for the government is calculated from Berry et al. 1998 and ranges from zero for most conservative to 100 for most liberal, see the data appendix for details. The citizen ideology measure is also taken from Berry et al. 1998. Democratic Presidential vote share refers to the share in the last presidential election. Divided Government is equal to one if the party of the governor is different from the party of the majority of legislators in either the state's lower or upper house. The polarization variables are calculated as difference in the Nominate scores of the two parties. The competitiveness index is the vote share of winning candidates in state legislative elections over the current and past 3 years. Geographic neighbors' experiments refers to the average number of experiments for adjacent states (geographically neighboring states). Population size neighbors' experiments refers to the average number of experiments for states of similar population size (where all states are divided into ten bands of similar population size). Ideological neighbors refers to the average number of experiments in states who voted similarly in the 1978 Presidential election.

Table 2: Policy Experimentation during the AFDC and TANF Period

	Main Effect Electoral Support		Differential Effect for Lame Ducks		Control for Potential Savings	
	(1)	(2)	(3)	(4)	(5)	(6)
Electoral Support Last Gubernatorial Election	-0.0044**	-0.0045*	-0.0069**	-0.0076**	-0.0061**	-0.0067*
	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Lame Duck * Electoral Support			0.0066*	0.0083*	0.0076*	0.0087*
			(0.004)	(0.004)	(0.004)	(0.005)
Lame Duck			-0.0299	-0.0512	-0.0229	-0.0651
			(0.053)	(0.054)	(0.062)	(0.062)
Past AFDC/TANF Spending					0.0004***	0.0005***
					(0.000)	(0.000)
Unemployment rate	0.0005	-0.0116	0.0019	-0.0101	-0.0102	-0.0219
	(0.014)	(0.018)	(0.014)	(0.019)	(0.013)	(0.023)
State income per capita	-0.0000***	-0.0000	-0.0000***	-0.0000	-0.0000*	-0.0000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Population	-0.0369	0.1368	-0.0376	0.1385	-0.0887**	-0.1619
	(0.025)	(0.184)	(0.026)	(0.185)	(0.040)	(0.127)
Percentage of population above 65	0.0099	-0.0082	0.0058	-0.0144	0.0056	-0.0123
	(0.041)	(0.089)	(0.041)	(0.090)	(0.045)	(0.110)
Percentage of population between 5 and 17	0.0275	0.0244	0.0262	0.0246	0.0333*	0.0270
	(0.017)	(0.021)	(0.018)	(0.021)	(0.017)	(0.020)
Governor age	0.0023	0.0036	0.0020	0.0034	0.0020	0.0031
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
State Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
State-specific Linear Trends	No	Yes	No	Yes	No	Yes
Observations	1,423	1,423	1,423	1,423	1,333	1,333
R-squared	0.267	0.289	0.269	0.291	0.278	0.301

Notes: The dependent variable in all specifications is the number of policy experiments in a given state and year over the period 1978-2008. Columns (1) and (2) include the governor's vote margin in the last gubernatorial election (measured as the distance to the runner-up in percent). Columns (3) and (4) allow the effect of the vote margin to differ for governors who cannot be reelected ("lame ducks"). Columns (5) and (6) also control for lagged AFDC/TANF spending (measured in real US$ million) as a proxy for the potential savings from experimentation. All specifications control for year and state fixed effects. Additional controls include governor age, the size of the state population, its age structure, the state-level unemployment rate and per capita state income. Even columns also control for state-specific linear trends. Standard errors clustered at the state level are shown in parentheses. *** p<0.01, ** p<0.05 and *p<0.1.

Table 3: Electoral Incentives and Policy Experimentation (Robustness)

	Competitiveness of Election			Alt. Electoral Support		Alt. Definitions Lame Duck		Alt. Potential Savings	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Electoral Support Last Gubernatorial Election	-0.0068*	-0.0064*	-0.0065*	-0.0117*	-0.0923*	-0.0063*	-0.0074	-0.0067*	-0.0076**
	(0.003)	(0.003)	(0.004)	(0.006)	(0.052)	(0.003)	(0.005)	(0.003)	(0.003)
Lame Duck * Electoral Support	0.0091*	0.0087*	0.0091*	0.0062	0.1019	0.0090*	0.0100*	0.0086*	0.0088*
	(0.005)	(0.005)	(0.005)	(0.010)	(0.092)	(0.005)	(0.006)	(0.005)	(0.005)
Lame Duck	-0.0782	-0.0624	-0.0759	-0.0353	-0.0284	-0.0698	-0.0898	-0.0674	-0.0661
	(0.065)	(0.063)	(0.066)	(0.073)	(0.062)	(0.062)	(0.061)	(0.061)	(0.061)
Past AFDC/TANF Spending	0.0005***	0.0005***	0.0005***	0.0005***	0.0005***	0.0005***	0.0007***	0.0005***	
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	
Governor Change past Electoral Cycles	-0.0292		-0.0297						
	(0.031)		(0.031)						
Competitiveness of Elections		0.0070	0.0072						
		(0.005)	(0.005)						
Quadratic Term Electoral Support				0.0000					
				(0.000)					
Lameduck*Electoral Support (Quadratic)				0.0002					
				(0.000)					
% Population Black								0.0409	
								(0.058)	
% Immigrant Population								-0.0130	
								(0.014)	
% Unmarried Births								0.0067	
								(0.019)	
Income Inequality (90/10 ratio)								-0.0097	
								(0.021)	
Total Welfare Spending									0.0002*
									(0.000)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State-specific linear trends	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Governor age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,333	1,324	1,324	1,333	1,457	1,275	880	1,333	1,378
R-squared	0.301	0.301	0.301	0.301	0.289	0.300	0.374	0.301	0.293

Notes: The dependent variable in all specifications is the number of policy experiments in a given state and year over the period 1978-2008. The main variables are the same as in Table 2 (see notes for details). Columns (1) and (3) include an indicator whether a state has changed their governor over the past electoral cycles. Columns (2) and (3) include the Holbrook and Van Dunk index of competitiveness in state elections. Column (4) include a quadratic term for electoral support, while column (5) collapses electoral support into a dummy variable if the electoral support is above the median and zero otherwise. Column (6) drops states with a one-term limit for state governors; columns (7) keeps only states who have a term limit for state governors. Column (8) include the share of Black population, the share of immigrants, the share of unmarried births and the 90-10 ratio in income as controls for the demand for welfare. Column (9) uses lagged total welfare spending in a state (which includes spending on food stamps, for instance) as an alternative measure of potential gains. All specifications control for year and state fixed effects as well as state-specific linear trends. Additional controls include governor age, the size of the state population, its age structure, the state-level unemployment rate and per capita state income. Standard errors clustered at the state level are shown in parentheses. *** p<0.01, ** p<0.05 and *p<0.1.

Table 4: Governor Ideology and Policy Experimentation

	Governor Ideology		Baseline with Ideology		Split Sample	
	(1)	(2)	(3)	(4)	(5)	(6)
Republican Governor	0.0772**		0.0876**			
	(0.035)		(0.039)			
Ideology Governor		-0.0029**		-0.0026**		
		(0.001)		(0.001)		
Electoral Support Last Gubernatorial Election			-0.0064*	-0.0063	-0.0119*	-0.0062*
			(0.003)	(0.004)	(0.006)	(0.003)
Lame Duck * Electoral Support			0.0086*	0.0104*	0.0183**	0.0086*
			(0.005)	(0.006)	(0.008)	(0.005)
Lame Duck			-0.0778	-0.0769	-0.2050	-0.0789
			(0.062)	(0.069)	(0.128)	(0.064)
Past AFDC/TANF Spending			0.0005***	0.0006***	0.0008***	0.0005***
			(0.000)	(0.000)	(0.000)	(0.000)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
State fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
State-specific linear trends	Yes	Yes	Yes	Yes	Yes	Yes
State characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Governor age	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,333	1,333	1,333	1,333	643	669
R-squared	0.290	0.310	0.305	0.325	0.339	0.380

Notes: The dependent variable in all specifications is the number of policy experiments in a given state and year. All main variables are defined as before (see notes to Table 2 for details). The ideology measure for the government in columns (2) and (4) is calculated from Berry et al. (1998) and ranges from zero for most conservative to 100 for most liberal. The last two specifications estimate the baseline separately for Republican (column (5)) and Democratic governors (column (6)). All specifications control for year and state fixed effects, state-specific linear trends and the same state characteristics as in Table 2. Standard errors clustered at the state level are shown in parentheses. *** p<0.01, ** p<0.05 and *p<0.1.

Table 5: Voter Preferences, the Demand for Welfare and Policy Experimentation

	Voter Ideology				Preferences for Redistribution			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Citizen Ideology (Berry et al. 1998)	-0.0005		-0.0000					
	(0.004)		(0.004)					
Democratic Presidential Vote Share		-0.0042		-0.0020				
		(0.005)		(0.006)				
Preferences for Redistribution					0.0413	0.0517	0.0886*	0.1126*
					(0.033)	(0.041)	(0.050)	(0.060)
Electoral Support Last Gubernatorial Election			-0.0067*	-0.0067*		-0.0067*		-0.0063*
			(0.003)	(0.003)		(0.003)		(0.003)
Lame Duck * Electoral Support			0.0087*	0.0087*		0.0089*		0.0089*
			(0.005)	(0.005)		(0.005)		(0.005)
Lame Duck			-0.0651	-0.0662		-0.0714		-0.0674
			(0.062)	(0.062)		(0.064)		(0.063)
Past AFDC/TANF Spending			0.0005***	0.0005***		0.0006***		0.0005***
			(0.000)	(0.000)		(0.000)		(0.000)
Past Spending*Preferences						-0.0001*		-0.0000
						(0.000)		(0.000)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State-specific linear trends	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Governor age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,333	1,333	1,333	1,333	1,333	1,333	1,333	1,333
R-squared	0.286	0.301	0.287	0.301	0.291	0.302	0.292	0.303

Notes: The dependent variable in all specifications is the number of policy experiments in a given state and year between 1978 and 2008. All main variables are defined as before (see notes to Table 2 for details). The citizen ideology measure (in columns (1) and (3)) is taken from Berry et al. (1998); see the data appendix for further details. Democratic presidential vote share (in columns (2) and (4)) refers to the last presidential election. Preferences for redistribution in columns (5) and (6) is an indicator equal to one if respondents in a state have stronger preferences for redistribution than respondents in the median state (based on a General Social Survey question on a 5-point Likert scale). The preference measure in columns (7) and (8) is also an indicator if respondents in a state are more likely than the median state to say that their government spends too little on welfare (based on a 3-point scale "too little", "about right" or "too much"). For both measures, higher values indicate a preference for more redistribution. All specifications control for state and year fixed effects as well as state-specific linear trends. In addition, all specifications include unemployment rate, income per capita, state population, the share of the population above 65, the population share between age 5 and 17 and governor age. Standard errors clustered at the state level are shown in parentheses. *** p<0.01, ** p<0.05 and *p<0.1.

Table 6: Composition of Legislature and Policy Experimentation

	(1)	(2)	(3)	(4)
Electoral Support Last Gubernatorial Election	-0.0068*	-0.0091***	-0.0063*	-0.0088**
	(0.003)	(0.003)	(0.003)	(0.003)
Lame Duck * Electoral Support	0.0087*	0.0102*	0.0093*	0.0111**
	(0.005)	(0.005)	(0.005)	(0.005)
Lame Duck	-0.0653	-0.0518	-0.0820	-0.0729
	(0.063)	(0.067)	(0.064)	(0.069)
Past AFDC/TANF Spending	0.0005***	0.0006***	0.0005***	0.0005***
	(0.000)	(0.000)	(0.000)	(0.000)
Dem. Seat Share Upper House	-0.3900			-0.2287
	(0.290)			(0.315)
Dem. Seat Share Lower House	-0.1085			-0.1190
	(0.406)			(0.407)
Polarization Nominate Score House		-0.0941		-0.0593
		(0.259)		(0.267)
Polarization Nominate Score Senate		0.2011		0.3306
		(0.274)		(0.394)
Divided Government			0.0906**	0.1004**
			(0.034)	(0.040)
Year fixed effects	Yes	Yes	Yes	Yes
State fixed effects	Yes	Yes	Yes	Yes
State-specific linear trends	Yes	Yes	Yes	Yes
State unemployment rate	Yes	Yes	Yes	Yes
State income per capita	Yes	Yes	Yes	Yes
State population	Yes	Yes	Yes	Yes
Percentage of the population above 65	Yes	Yes	Yes	Yes
Percentage of the population between 5 and 17	Yes	Yes	Yes	Yes
Governor age	Yes	Yes	Yes	Yes
Observations	1,333	1,085	1,316	1,069
R-squared	0.302	0.294	0.305	0.299

Notes: The dependent variable in all specifications is the number of policy experiments in a given state and year. All main variables are defined as before (see notes to Table 2 for details). Divided Government is equal to one if the party of the governor is different from the party of the majority of legislators in either the state's lower or upper house. The polarization variables are calculated as absolute deviations of the democratic seat share from 50%. All specifications control for state and year fixed effects, state-specific linear trends and the same state characteristics as in Table 2. Standard errors clustered at the state level are shown in parentheses. *** $p<0.01$, ** $p<0.05$ and *$p<0.1$.

Table 7: Spillovers between States and Policy Experimentation

	(1)	(2)	(3)	(4)
Electoral Support Last Gubernatorial Election	-0.0058*	-0.0045*	-0.0045*	-0.0057*
	(0.003)	(0.003)	(0.003)	(0.003)
Lame Duck * Electoral Support	0.0085*	0.0088*	0.0088*	0.0084*
	(0.005)	(0.005)	(0.005)	(0.005)
Lame Duck	-0.0418	-0.0456	-0.0463	-0.0395
	(0.066)	(0.063)	(0.065)	(0.067)
Past AFDC/TANF Spending	0.0005***	0.0005***	0.0006***	0.0005***
	(0.000)	(0.000)	(0.000)	(0.000)
Geographic Neighbors' Experiments (t-2)	0.0841			0.0801
	(0.062)			(0.064)
Pop. Size Neighbors' Experiments (t-2)		0.0331		0.0325
		(0.075)		(0.075)
Ideological Neighbors' Experiments (t-2)			0.0295	0.0101
			(0.075)	(0.077)
Year fixed effects	Yes	Yes	Yes	Yes
State fixed effects	Yes	Yes	Yes	Yes
State-specific linear trends	Yes	Yes	Yes	Yes
State unemployment rate	Yes	Yes	Yes	Yes
State income per capita	Yes	Yes	Yes	Yes
State population	Yes	Yes	Yes	Yes
Percentage of the population above 65	Yes	Yes	Yes	Yes
Percentage of the population between 5 and 17	Yes	Yes	Yes	Yes
Governor age	Yes	Yes	Yes	Yes
Observations	1,411	1,411	1,411	1,411
R-squared	0.291	0.291	0.291	0.291

Notes: The dependent variable in all specifications is the number of policy experiments in a given state and year. All main variables are defined as before (see notes to Tables 2 and 3 for details). Experiments by geographic neighbors refers to the average number of experiments in adjacent states (sharing a boundary), while experiments by population size neighbors refers to the average number of experiments in states with similar population size (where states are divided into ten groups of similar population size). All specifications control for year and state fixed effects as well as governor age. Standard errors clustered at the state level are shown in parentheses. *** $p<0.01$, ** $p<0.05$ and * $p<0.1$.

Table 8: Policy Reversals in the TANF Period

	Dummy for Policy Reversal				Number of Policy Reversals			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Past TANF Spending	-0.0001**	-0.0001*	-0.0001*	-0.0001**	-0.0002**	-0.0002**	-0.0001**	-0.0002**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Lame Duck		0.0281	0.0268			0.0248	0.0191	
		(0.017)	(0.028)			(0.016)	(0.035)	
Republican Governor			-0.0211				-0.0197	
			(0.021)				(0.024)	
Electoral Support Last Gubernatorial Election				0.0008				0.0008
				(0.002)				(0.002)
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State-specific linear trends	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Governor Age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	751	749	733	729	751	749	733	729
R-squared	0.166	0.167	0.142	0.171	0.161	0.162	0.171	0.168

Notes: The dependent variable in columns (1)-(4) is a dummy whether a state has reversed a policy during the TANF period (1996-2008) and the number of policy reversals in columns (5)-(8). Columns (1) and (5) include include past state TANF spending (measured in US$ million) as a measure of the potential gains from experimentation. Columns (2) and (6) add an indicator whether the governor cannot be reelected ("lame duck"). Columns (3) and (7) include whether the governor is a Republican or Democrat. Finally, columns (4) and (8) include the electoral support in the past gubernatorial election. All specifications control for state and year fixed effects as well as state-specific linear trends. Standard errors clustered at the state level are shown in parentheses. *** p<0.01, ** p<0.05 and *p<0.1.

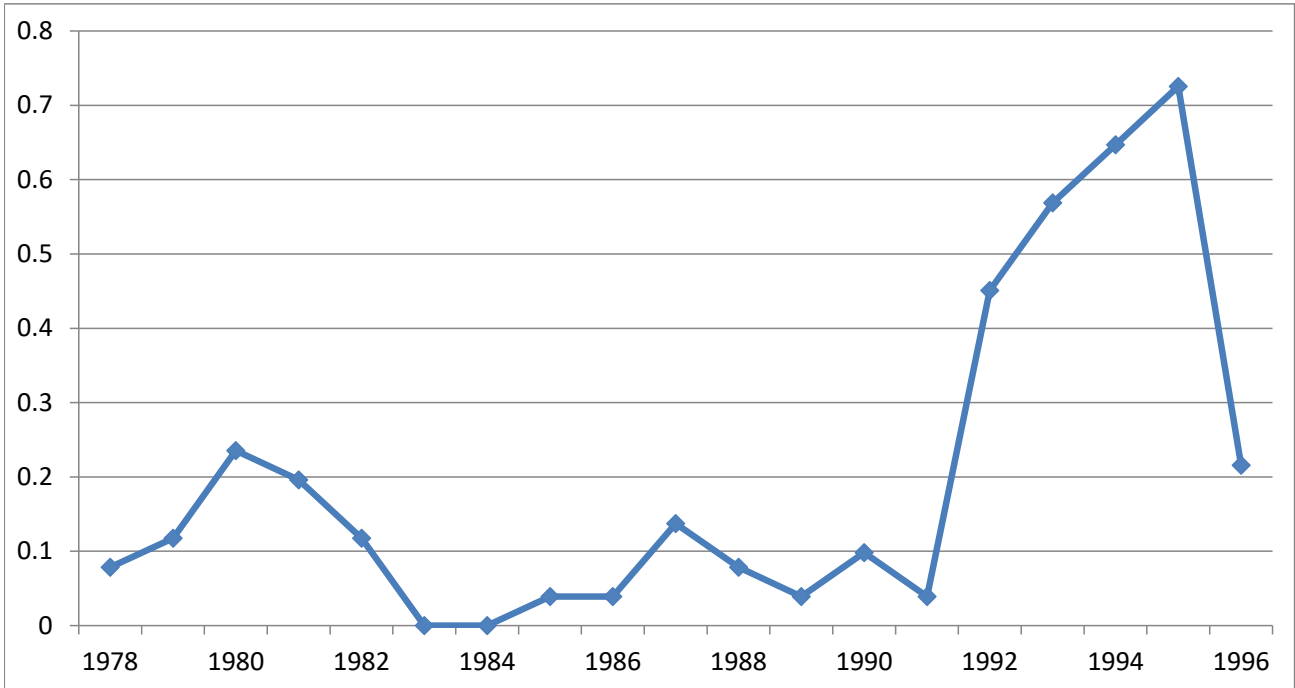**Figure 1: Evolution of Welfare Waivers, 1978-1996**

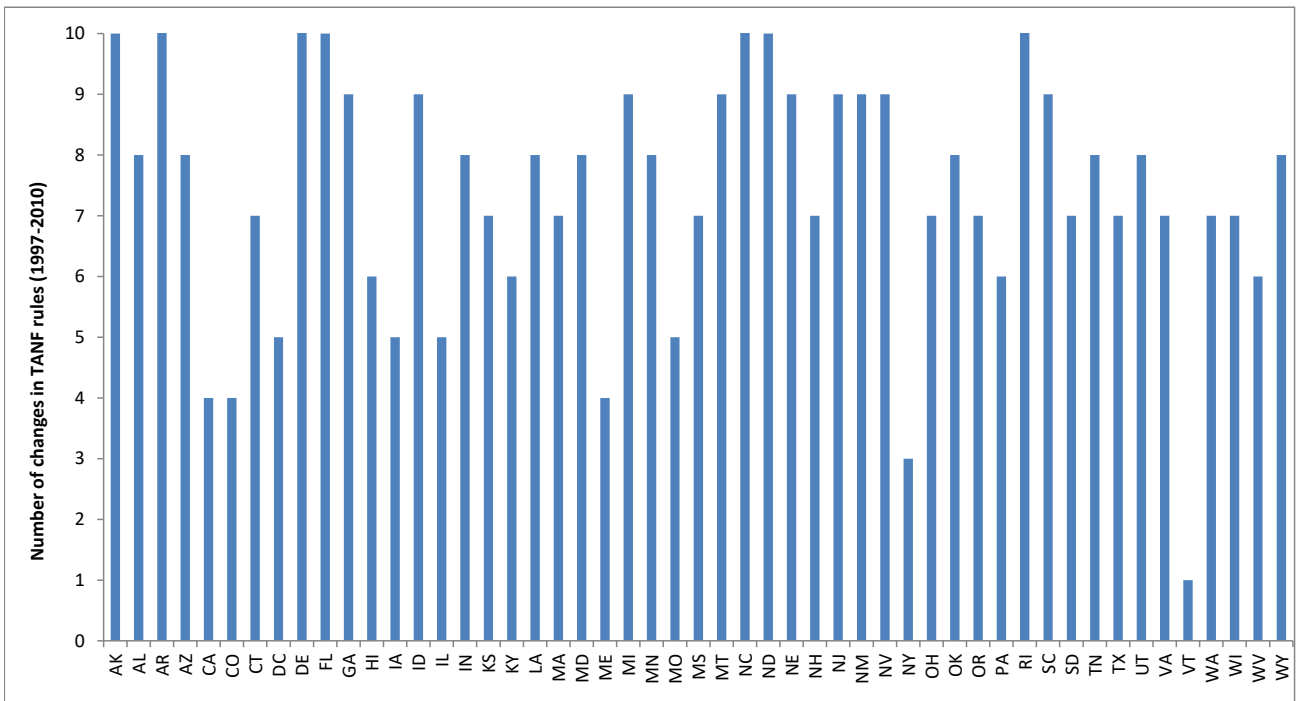**Figure 2: Distribution of Waiver Applications across States, 1978-1996**
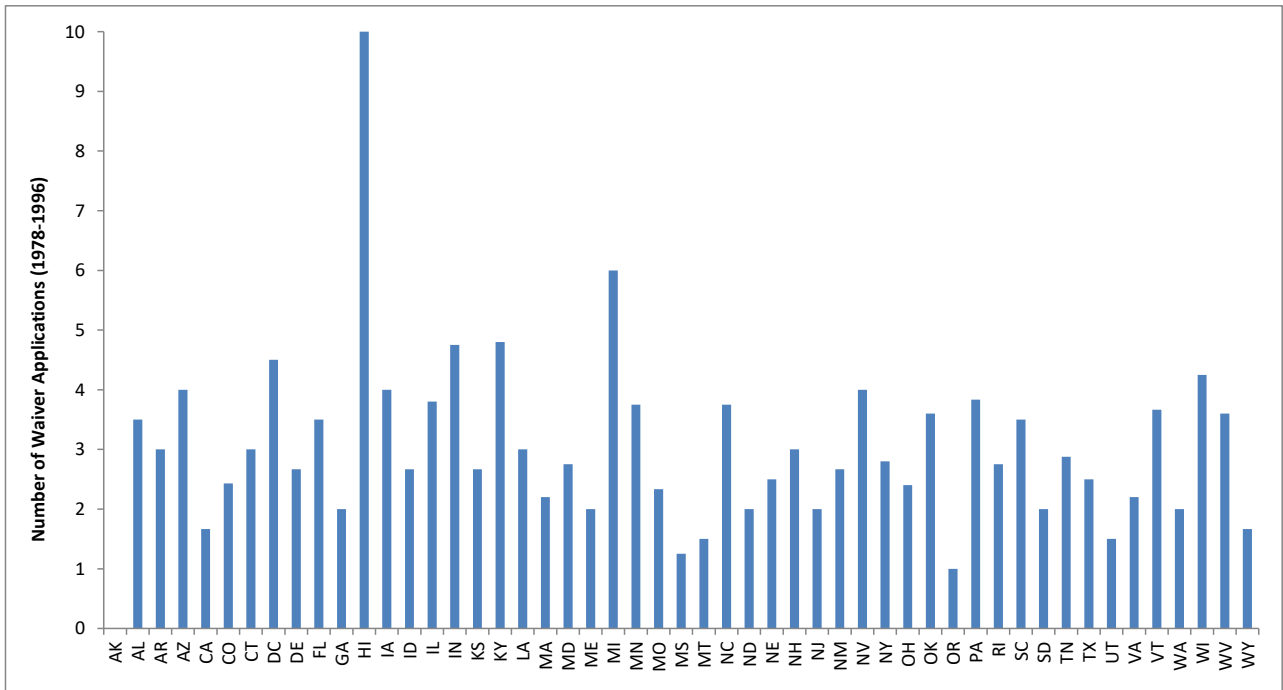
Figure 3: Evolution of Policy Experiments and Reversals, 1996-2010
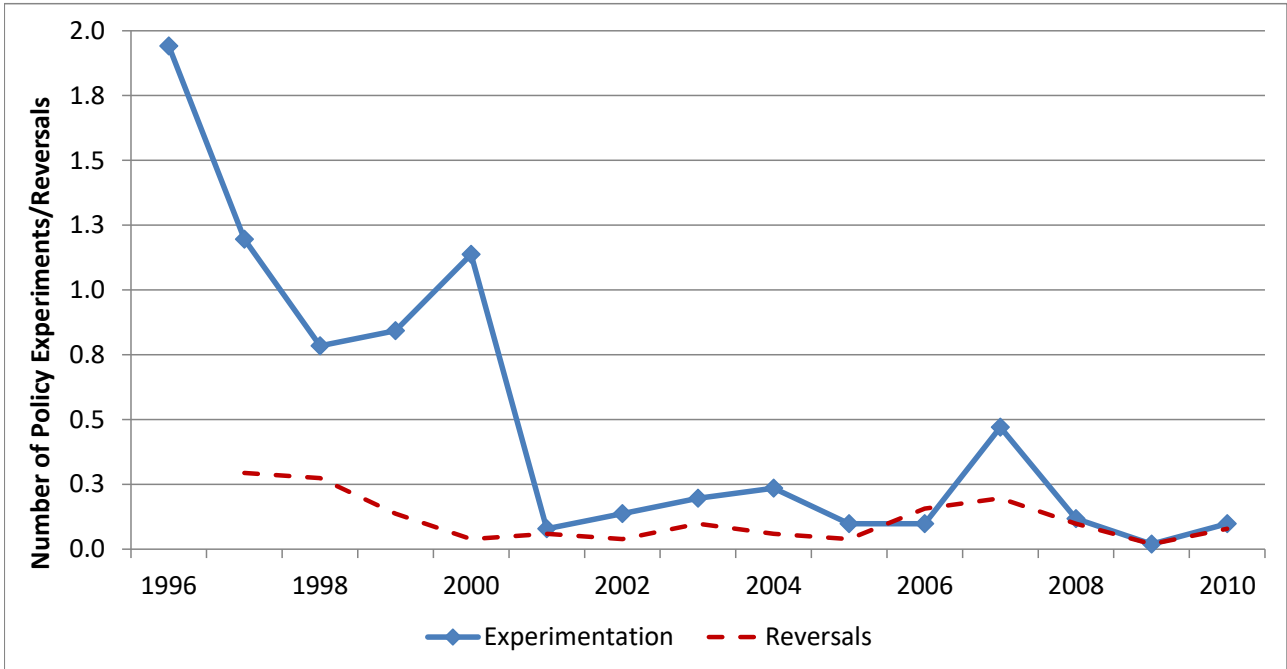
Figure 4: Distribution of Policy Experiments across States, 1996-2010

Table A1: Coding of Policy Experimentation and Reversals during the TANF Period (1996-2010)

Policy Rule	Description of Rule	Experimentation = 1	Reversal = 1
Family Cap (1)	Benefits do not increase if an additional child is born in family while receiving benefits.	State adopts family cap before or after 1996.	State abolishes family cap in 1997 or later.
Work Requirements (3)			
Hours Requirement	Minimum # of hours a recipient must participate in work-related activities.	Increase in minimum number of working hours required of the recipient.	State reduces or abolishes minimum number of working hours required of the recipient.
Work upon Enrollment	Work requirements apply at application, approval or upon benefit receipt (or only later).	State requires work requirement to be fulfilled at application, approval or upon benefit receipt.	State allows work requirement to be fulfilled later.
Time Limit to Work	Work at least 20 hours per week in an unsubsidized job after a certain period of benefit receipt.	State adopts a time limit to fulfill the requirement of working in a unsubsidized job for at least 20 hours.	State abolishes time limit to fulfill the requirement of working in an unsubsidized job for at least 20 hours.
Time Limits (3)			
Duration of Lifetime Limit	Maximum # of months an assistance unit can receive benefits over the lifetime.	Prior to 1997, if a time limit is adopted; after 1997, when a state adopts a time limit stricter than the 60 months federal time limit. In 1997, if a state adopts a time limit other than the federal time limit.	If a state abolishes the time limit or increases # of months.
Intermittent Time Limit	# months time an assistance units can receive benefits without interruption.	State reduces intermittent time limit.	State increases intermittent time limit.
Benefit Reduction after Intermittent Time Limit	How much benefits are reduced (adult portion or benefit unit) when assistance units hits the	State reduces benefits more after unit hits intermittent time limit.	State reduces benefits less after unit hits intermittent time limit.
Sanctions (4)			
Severity of Worst Sanction	How much benefits are reduced under worst sanction for non-compliance with work	State imposes a more severe penalty for non-compliance.	State reduces severity of worst sanction for non-compliance.
Duration of Worst Sanction	Length of most severe sanction for not complying with work requirements.	State increases duration of worst sanction.	State decreases duration of worst sanction.
Reapply	Whether the unit has to reapply after worst sanction for non-compliance has been imposed.	State adopts requirement that unit has to reapply after worst sanction has been imposed.	State abolishes requirement that unit has to reapply after worst sanction has been imposed.
Severity of Initial Sanction	Whether initial sanction for non-compliance is removal of full family benefit.	State adopts that full family benefit is removed as initial sanction.	State abolishes rule that full family benefit is removed as initial sanction.

Source: Welfare Rules Database (2015)

Table A2: Arellano-Bond Dynamic Panel Estimates of Policy Experimentation

	Policy Experimentation		
	(1)	(2)	(3)
Experimentation (t-1)	-0.1067**	-0.1125***	-0.1468***
	(0.044)	(0.041)	(0.038)
Vote Margin Last Gubernatorial Election		-0.0118**	-0.0091*
		(0.005)	(0.005)
Lame Duck * Gubernatorial Vote Margin		0.0164**	0.0194**
		(0.008)	(0.009)
Lame Duck		-0.1578	-0.2222**
		(0.106)	(0.111)
Past TANF Spending			0.0010***
			(0.000)
Observations	1,313	1,313	1,313
Wald test	4008.05	4045.51	3371.39
(df)	(37)	(39)	(39)
p-value of AR(2) residuals	0.282	0.425	0.998
p-value of AR(3) residuals	0.358	0.427	0.552
Sargan test overidentifying restrictions (p-value)	0.065	0.206	0.194

Notes: The table reports Arellano-Bond dynamic panel data estimates including a lagged dependent variable to account for unobserved heterogeneity. The estimator uses all included variables lagged 2 to 4 years as potential instruments. All specifications include year fixed effects and the same state characteristics than in Table 2. The p-value tests the null hypothesis that the residuals are serially correlated of the specific degree: AR(2) or AR(3). The bottom row reports the p-value of the Sargan test of overidentifying resitrictions. *** p<0.01, ** p<0.05 and * p<0.1.

Table A3: Governor Heterogeneity in Competence and Experimentation

	Little Prior Political Experience		Less than a Masters Degree	
	(1)	(2)	(3)	(4)
Governor with Low Education/Experience	-0.0230	-0.0630	-0.0223	-0.0591
	(0.065)	(0.074)	(0.054)	(0.083)
Lame Duck * Low Education / Experience		0.1456*		0.0619
		(0.086)		(0.068)
Electoral Support Last Gubernatorial Election	-0.0067*	-0.0068*	-0.0068*	-0.0068*
	(0.003)	(0.003)	(0.003)	(0.003)
Lame Duck * Electoral Support	0.0089*	0.0091*	0.0087*	0.0087*
	(0.005)	(0.005)	(0.005)	(0.005)
Lame Duck	-0.0649	-0.1308*	-0.0664	-0.0363
	(0.063)	(0.076)	(0.063)	(0.080)
Past AFDC/TANF Spending	0.0005***	0.0005***	0.0005***	0.0005***
	(0.000)	(0.000)	(0.000)	(0.000)
Year fixed effects	Yes	Yes	Yes	Yes
State fixed effects	Yes	Yes	Yes	Yes
State-specific linear trends	Yes	Yes	Yes	Yes
State characteristics	Yes	Yes	Yes	Yes
Governor age	Yes	Yes	Yes	Yes
Observations	1,330	1,330	1,330	1,330
R-squared	0.302	0.303	0.301	0.301

Notes: The dependent variable in all specifications is the number of policy experiments in a given state and year over the period 1978-2008. Columns (1) and (2) include an indicator whether the governor has little prior political experience (below the median). Columns (3) and (4) use an indicator whether a governor has less than a Masters degree. Columns (2) and (4) interact both measures of governor ability with an indicator whether the governor is a lame duck. The other independent variables are defined as in previous tables. All specifications control for year and state fixed effects as well as state-specific trends. Additional controls include governor age, the size of the state population, its age structure, the state-level unemployment rate and per capita state income. Standard errors clustered at the state level are shown in parentheses. *** p<0.01, ** p<0.05 and *p<0.1.

Table A4: Descriptives on Policy Reversals

Policy Rule	# Experiments (1)	# Reversals (2)	% Same Elec. Cycle (3)	Mean # Years (4)	% Same Party (5)	% Same Governor (6)
Family Cap	22	5	0%	10.2	20%	0%
Work Requirements						
Hours Requirement	126	14	30%	5.4	50%	30%
Work upon Enrollment	43	2	0%	6.5	100%	50%
Time Limit to Work	3	2	0%	6	100%	50%
Time Limits						
Duration of Lifetime Limit	15	6	17%	3.7	100%	80%
Intermittent Time Limit	20	10	30%	5.4	50%	40%
Benefit Reduction after Intermittent Time Limit	53	3	33%	7.7	100%	33%
Sanctions						
Severity of Worst Sanction	55	1	0%	2	100%	100%
Duration of Worst Sanction	15	38	42%	4.3	78%	50%
Reapply	20	6	0%	3.5	67%	100%
Severity of Initial Sanction	22	0				

Notes : The table reports for the 11 policy rules (and 87 policy reversals in total) the percentage of reversals that occur within the same electoral cycle as the experiment; the mean number of years between experiment and reversal; the share of reversals that are implemented by the same party (Republican or Democrat) as the initial experiment; and the percentage of reversals that are implemented by the same governor as the experiment. The number of experiments of the duration of the worst sanction is lower than the number of reversals because many states initially applied the federal guidelines when adopting their state TANF policies. Later on, many states deviated from these rules which accounts for the relatively high number of reversals over our sample period.